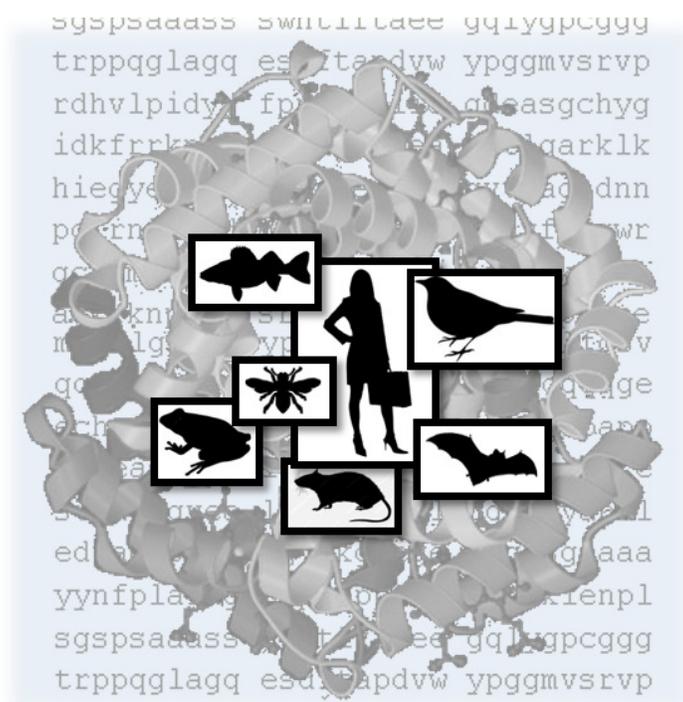


EPA/600/R-18/062

# Sequence Alignment to Predict Across Species Susceptibility

(SeqAPASS)

VERSION 3.0



## User Guide

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

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

**Quick Notes:** Use [Mozilla Firefox or Chrome](#) 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.

### Table of Contents

Background	page 2
Accessing SeqAPASS	page 3-4
<i>Returning Users</i> (page 3)	
<i>First Time Users</i> (page 4)	
Messages from the SeqAPASS Development Team	page 5
SeqAPASS Home Tab	page 6
Request SeqAPASS Run Tab	page 7-12
<i>Query “By Species”</i> (page 8)	
<i>Query “By Accession”</i> (page 11)	
SeqAPASS Run Status	page 13-14
View SeqAPASS Reports	page 15-21
<i>View Report</i> (page 16)	
<i>Save Report(s)</i> (page 17)	
Level 1: Primary Amino Acid Sequence Alignment	page 22-26
<i>Primary Report Settings</i> (page 24)	
Susceptibility Cutoff Box for Level 1	page 27-31
<i>No Ortholog Candidate</i> (page 30)	
Level 2: Functional Domains Alignment	page 32-34
View Level 2 Data Page	page 36-41
<i>Primary Report Settings</i> (page 37)	
Susceptibility Cutoff Box for Level 2	page 40-43
<i>No Ortholog Candidate</i> (page 42)	
Level 1 and Level 2 Data Visualization	page 44-55
<i>Level 1 Information Page</i> (page 46)	
<i>Level 2 Information Page</i> (page 47)	
<i>Level 1 and 2 BoxPlot Page – Controls</i> (page 48)	
Level 3: Individual Amino Acid Residue Alignment	page 56-61
View Level 3 Individual Amino Acid Query and Data Page	page 62-67
<i>Level 3 Data – Primary Report</i> (page 65)	
<i>Level 3 Data – Primary Report</i> (page 66)	
Moving Between Level 1, Level 2, and Level 3 Data Pages	page 68
Search, View, and Download Data Tables	page 69-70
Log out	page 71
Pop-up Messages	page 72-74
Detailed Documentation	page 75-84

Note: Screen-shots displayed throughout User Guide may depict a different version than the current version if functionality did not change in SeqAPASS Version 3.0.

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

Updated 3/06/18; 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 pathway 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).

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

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

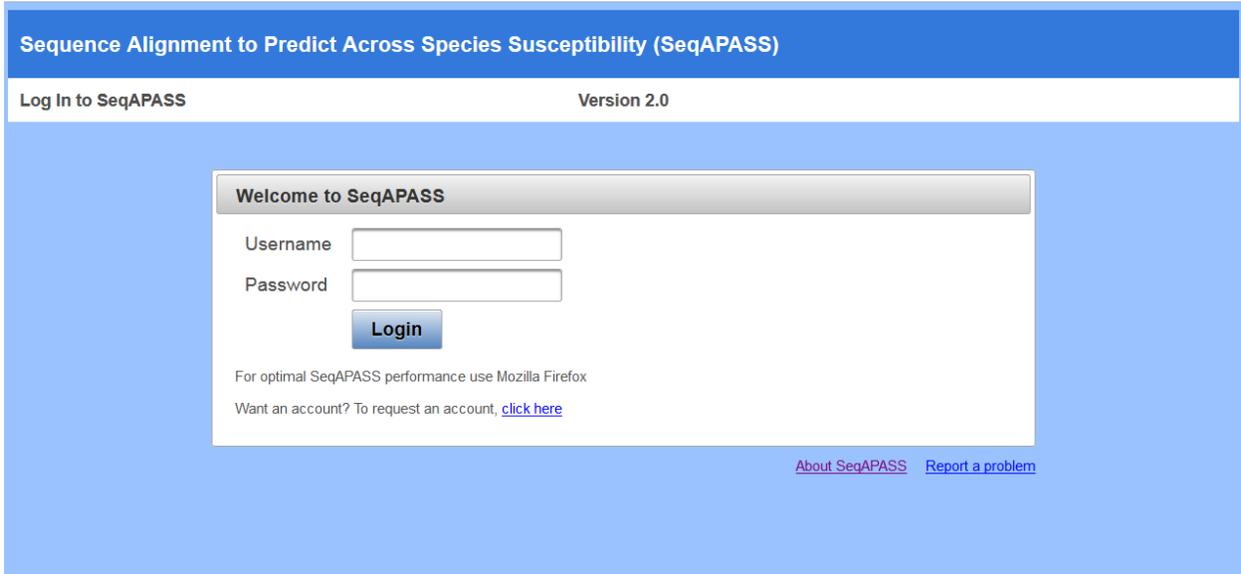
## Accessing SeqAPASS

For optimal SeqAPASS performance use **Chrome** or **Mozilla Firefox**

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

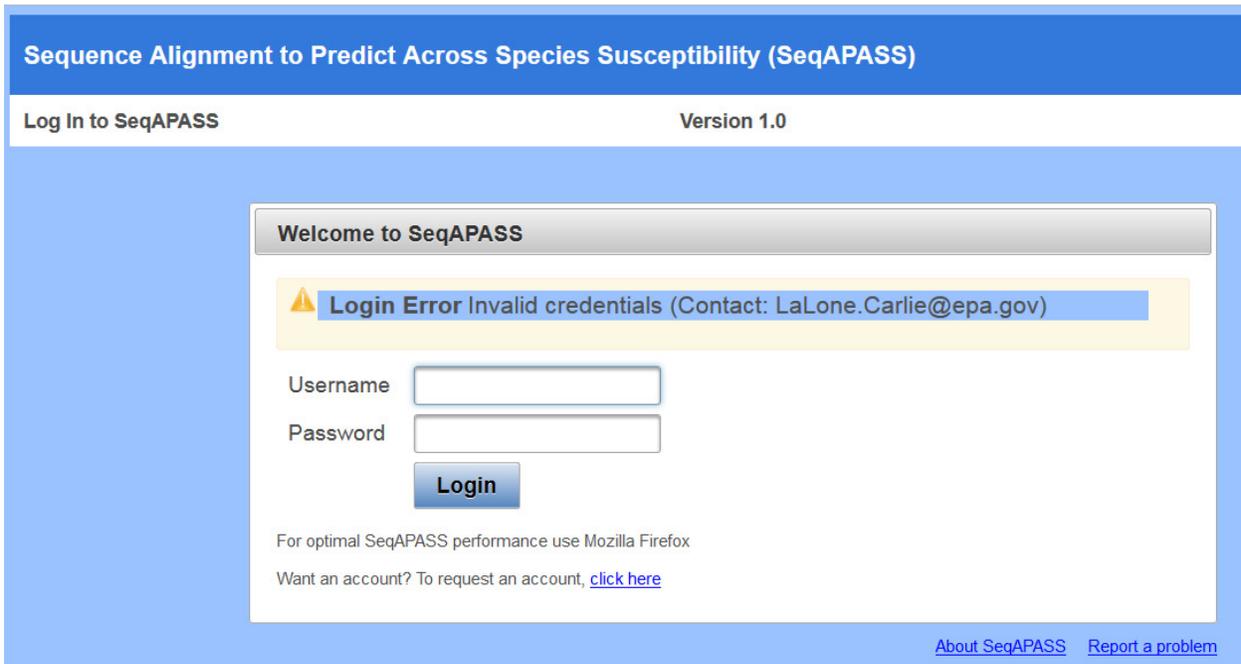
### Returning Users

Enter Username and Password



The screenshot shows the SeqAPASS login interface for Version 2.0. 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 2.0" 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



The screenshot shows the SeqAPASS login interface for Version 1.0. 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". A yellow error banner with a warning icon and the text "Login Error Invalid credentials (Contact: LaLone.Carlie@epa.gov)" is displayed above the login fields. Below the banner, 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".

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

Updated 3/06/18; Contact Carlie LaLone with Questions: [LaLone.Carlie@epa.gov](mailto: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](mailto:LaLone.Carlie@epa.gov) requesting a password. A reply email will be delivered to you with your temporary password. The login does not limit access to the tool. Everyone that requests an account will be given one in a timely manner. Individual account allow users to store all previous SeqAPASS runs.

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 2.0", 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 "Re-Enter New Password" field.

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

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Messages from the SeqAPASS development team

Look for messages about planned version releases, data updates, and/or fixes to the SeqAPASS tool. These will occasionally be displayed below the SeqAPASS banner when the development team has information to share with SeqAPASS users.

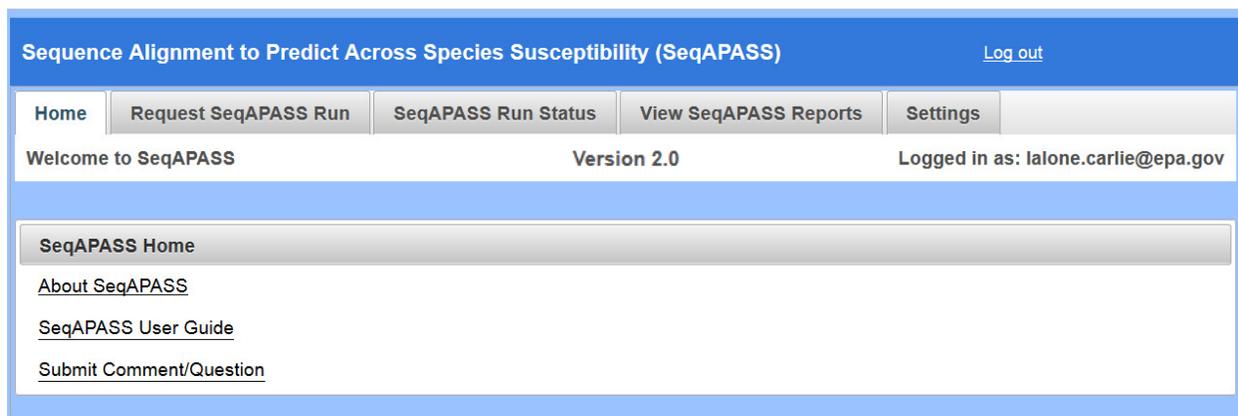
The screenshot displays the SeqAPASS web interface. At the top, a blue banner reads "Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)". Below this, a light blue box contains an information icon and a heading "New to SeqAPASS Version 2 (See [user guide](#) for more details)". A bulleted list follows, detailing updates such as new data from NCBI, report customization options, and support for additional sequences. Below the message box, the navigation bar shows "Log In to SeqAPASS" on the left and "Version 2.0" on the right. The main content area features a "Welcome to SeqAPASS" login form with fields for "Username" and "Password", a "Login" button, and a note recommending Mozilla Firefox. At the bottom right of the page, there are links for "About SeqAPASS" and "Report a problem".

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

Updated 3/06/18; 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 shows "Welcome to SeqAPASS" on the left, "Version 2.0" in the center, and "Logged in as: lalone.carlie@epa.gov" on the right. The main content area features a list of links: "SeqAPASS Home" (highlighted), "About SeqAPASS", "SeqAPASS User Guide", and "Submit Comment/Question".

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

Updated 3/06/18; 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 SeqAPASS web application interface. At the top, there is a blue header with the title "Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)" and a "Log out" link. Below the header is a navigation menu with buttons for "Home", "Request SeqAPASS Run", "SeqAPASS Run Status", "View SeqAPASS Reports", and "Settings". The "Request SeqAPASS Run" button is highlighted. Below the navigation menu, there is a status bar showing "Request Level 1 SeqAPASS Run", "Version 2.0", and "Logged in as: LaLone.Carlie@epa.gov". The main content area is divided into four sections: "Compare Primary Amino Acid Sequences" with radio buttons for "By Species" and "By Accession"; "Query Species Selection" with a search box and "Add Query Species" button; "Query Protein Selection" with a search box and "Filter Protein" button; and "SeqAPASS Submission" with a "Final Query Protein(s)" text area 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.

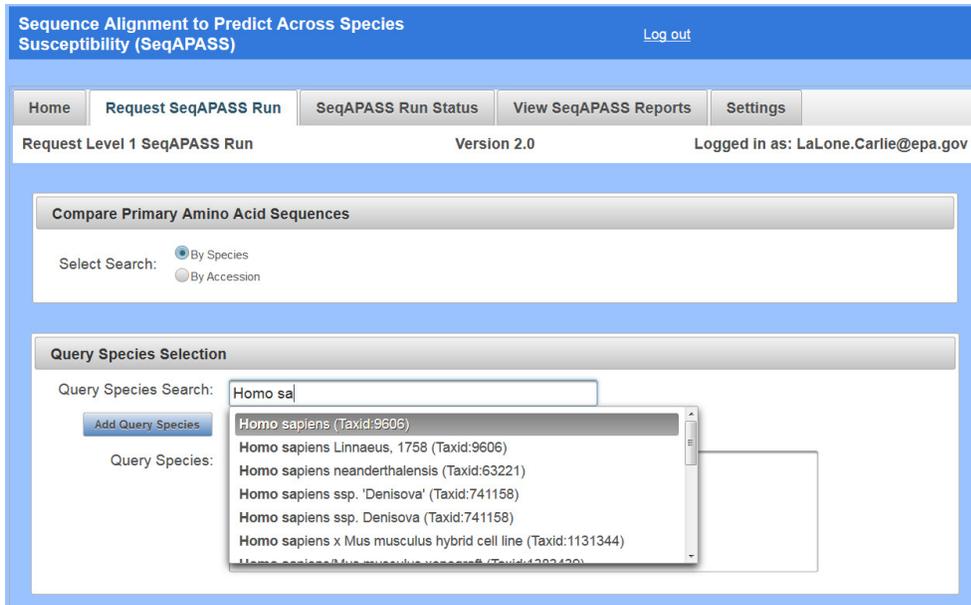
The screenshot shows the SeqAPASS web application interface. At the top, there is a blue header with the title "Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)" and a "Log out" link. Below the header is a navigation menu with buttons for "Home", "Request SeqAPASS Run", "SeqAPASS Run Status", "View SeqAPASS Reports", and "Settings". The "Request SeqAPASS Run" button is highlighted. Below the navigation menu, there is a status bar showing "Request Level 1 SeqAPASS Run", "Version 2.0", and "Logged in as: LaLone.Carlie@epa.gov". The main content area is divided into four sections: "Compare Primary Amino Acid Sequences" with radio buttons for "By Species" and "By Accession"; "Query Species Selection" with a search box and "Add Query Species" button; "Query Protein Selection" with a search box and "Filter Protein" button; and "SeqAPASS Submission" with a "Final Query Protein(s)" text area and "Request Run" and "Clear" buttons.

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

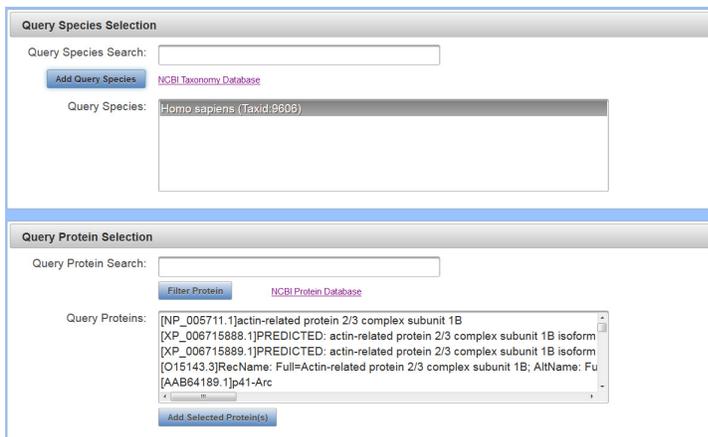
## Query “By Species”

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 auto-complete function and will generate a list of species with corresponding Taxid. When text is typed into the search bar, the auto-complete function queries the database in the order of “starts with” then “contains.” If an integer is typed in the search bar the autocomplete function queries the database in the order of “Taxid”, “starts with”, then “contains.”



**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 initial 200 proteins/species). The protein list includes the protein NCBI accession, protein name, and species scientific name.



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

Updated 3/06/18; 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 proteins of interest).

Query Protein Selection

Query Protein Search:

[NCBI Protein Database](#)

Query Proteins: [A8MWY0.2]RecName: Full=UPF0577 protein KIAA1324-like; AltName: Full=Estrogen  
[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:

[NCBI Protein Database](#)

Query Proteins: [NP\_001258805.1]estrogen receptor beta isoform 5  
[NP\_001258806.1]estrogen receptor beta isoform 6  
[NP\_001278159.1]estrogen receptor isoform 2  
[NP\_001278170.1]estrogen receptor isoform 3  
[NP\_001278641.1]estrogen receptor beta isoform 2

---

SeqAPASS Submission

Final Query Protein(s) [NP\_001258805.1]estrogen receptor beta isoform 5  
[NP\_001278159.1]estrogen receptor isoform 2  
[NP\_001278641.1]estrogen receptor beta isoform 2

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

Updated 3/06/18; 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.

The screenshot shows the SeqAPASS web interface. At the top, there are navigation tabs: Home, Request SeqAPASS Run (selected), SeqAPASS Run Status, View SeqAPASS Reports, and a Success notification. Below the tabs, the page title is "Request Level 1 SeqAPASS Run" and "Version 2.0". A success message in the top right corner reads: "Success Submitted NP\_001278159.1: submitted". Below this, there is a section titled "Compare Primary Amino Acid Sequences" with a "Select Search:" dropdown set to "By Species". The "Query Species Selection" section contains a "Query Species Search:" input field, an "Add Query Species" button, and a list of query species: "Homo sapiens (Taxid:9606)" and "Bos taurus (Taxid:9913)".

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 populated in the “Query Proteins” box will always be associated with the species highlighted in the “Query Species” box.

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

The screenshot shows the "Query Protein Selection" section of the SeqAPASS web interface. It features a "Query Protein Search:" input field, a "Filter Protein" button, and a list of query proteins: "[NP\_776588.1]ras-related C3 botulinum toxin substrate 1 precursor", "[P62998.1]RecName: Full=Ras-related C3 botulinum toxin substrate 1; AltName: Fu", "[AAF00714.1]GTPase", "[AAI03062.1]Ras-related C3 botulinum toxin substrate 1 (rho family, small GTP bind", and "[NP\_001015639.1]26S protease regulatory subunit 7". Below the list is an "Add Selected Protein(s)" button.

**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 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

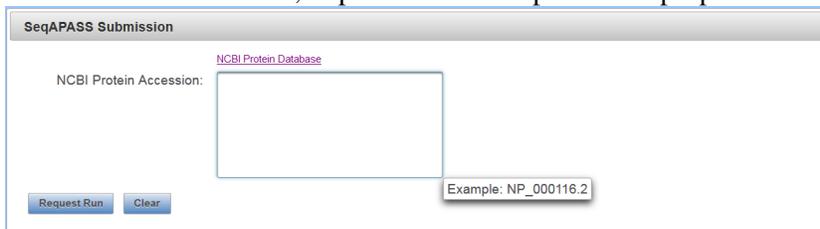
### Query “By Accession”

Users familiar with the NCBI database can utilize NCBI protein accessions (e.g., NP\_000116.2) 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.

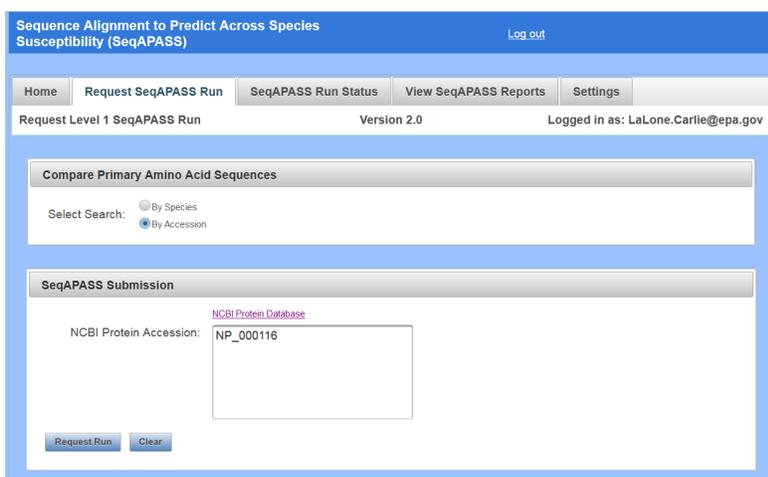


Upon selecting the “By Accession” radio button, a new query page will be displayed. Type the NCBI protein accession (e.g., NP\_000116.2) for the protein of interest (this Accession comes from the NCBI protein database; See “SeqAPASS Documentation” for details) in the “NCBI Protein Accession” box. If 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.

Upon clicking 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.



**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.**



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

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

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 “NCBI Protein Accession” 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 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 bar with buttons for "Home", "Request SeqAPASS Run", "SeqAPASS Run Status", "View SeqAPASS Reports", and "Settings". The main content area is divided into two sections. The first section, "Compare Primary Amino Acid Sequences", includes a "Select Search:" label with two radio buttons: "By Species" (unselected) and "By Accession" (selected). The second section, "SeqAPASS Submission", features a "NCBI Protein Accession:" label, a text input field, and a "Request Run" button. A "Clear" button is also present. A success message overlay in the top right corner reads: "Success NP\_001315029: submitted: NP\_001315029,1 Logged in as: LaLone.Carlie@epa.gov".

Note: All NCBI Accessions can include the version number (one digit after the decimal place, e.g., NP\_000116.2). Otherwise, if the version is not included, the most recent version of the accession will be queried automatically.

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

Updated 3/06/18; 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  $\leq 0.01$ ) 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., 2<sup>nd</sup> 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 % completed for RPS-BLAST or reciprocal best hit BLAST, respectively, will be displayed. Please see example below:

SeqAPASS Run Status

Version 2.0

Logged in as: LaLone.Carlie@epa.gov

Level 1 Status (selected)

Level 2 Status

Level 3 Status

Refresh Data

SeqAPASS Level 1 Run Status

Search: Enter keyword

SeqAPASS Run Id	Data Version	User	Level 1 Query Accession	BLASTp	Common Domains	Ortholog Candidate	Start Date	Date Completed	SeqAPASS Run Duration
1567	2	lalone.carlie@epa.gov	NP_001496.1	complete	100%	100%	2017 05 09 11:40:40	2017 05 09 11:45:01	4 minute(s) 21 second(s)
1567	2	lalone.carlie@epa.gov	NP_001315029.1	complete	100%	100%	2017 05 09 11:40:40	2017 05 09 11:43:27	2 minute(s) 47 second(s)
1566	2	lalone.carlie@epa.gov	NP_001496.1	complete	100%	100%	2017 05 09 11:38:52	2017 05 09 11:45:01	6 minute(s) 9 second(s)
1565	1	lalone.carlie@epa.gov	NP_001116214.1	complete	100%	100%	2017 05 09 11:38:20	2017 05 09 11:38:20	1 seconds
1564	1	lalone.carlie@epa.gov	NP_001116213.1	complete	100%	100%	2017 05 09 11:37:40	2017 05 09 11:37:40	1 seconds
1563	1	lalone.carlie@epa.gov	NP_000116.2	complete	100%	100%	2017 05 09 11:36:12	2017 05 09 11:36:12	1 seconds
1562	2	lalone.carlie@epa.gov	NP_001258905.1	complete	100%	100%	2017 05 09 11:22:07	2017 05 09 11:29:25	7 minute(s) 18 second(s)
1562	2	lalone.carlie@epa.gov	NP_001278169.1	complete	100%	100%	2017 05 09 11:22:07	2017 05 09 11:40:06	17 minute(s) 59 second(s)
1562	2	lalone.carlie@epa.gov	NP_001278641.1	complete	100%	100%	2017 05 09 11:22:07	2017 05 09 11:34:42	12 minute(s) 35 second(s)
1561	2	goldAdmin	ABI98425.1	complete	100%	100%	2017 05 08 11:16:19	2017 05 08 11:28:08	11 minute(s) 49 second(s)

(1 of 116) Download Table

Top of Page

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 in the “Request SeqAPASS Run” tab. 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 “Data Version” column indicates which version of NCBI data is being used (See “About” page for details on Data Versions)

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

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. Please see below:

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:

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 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

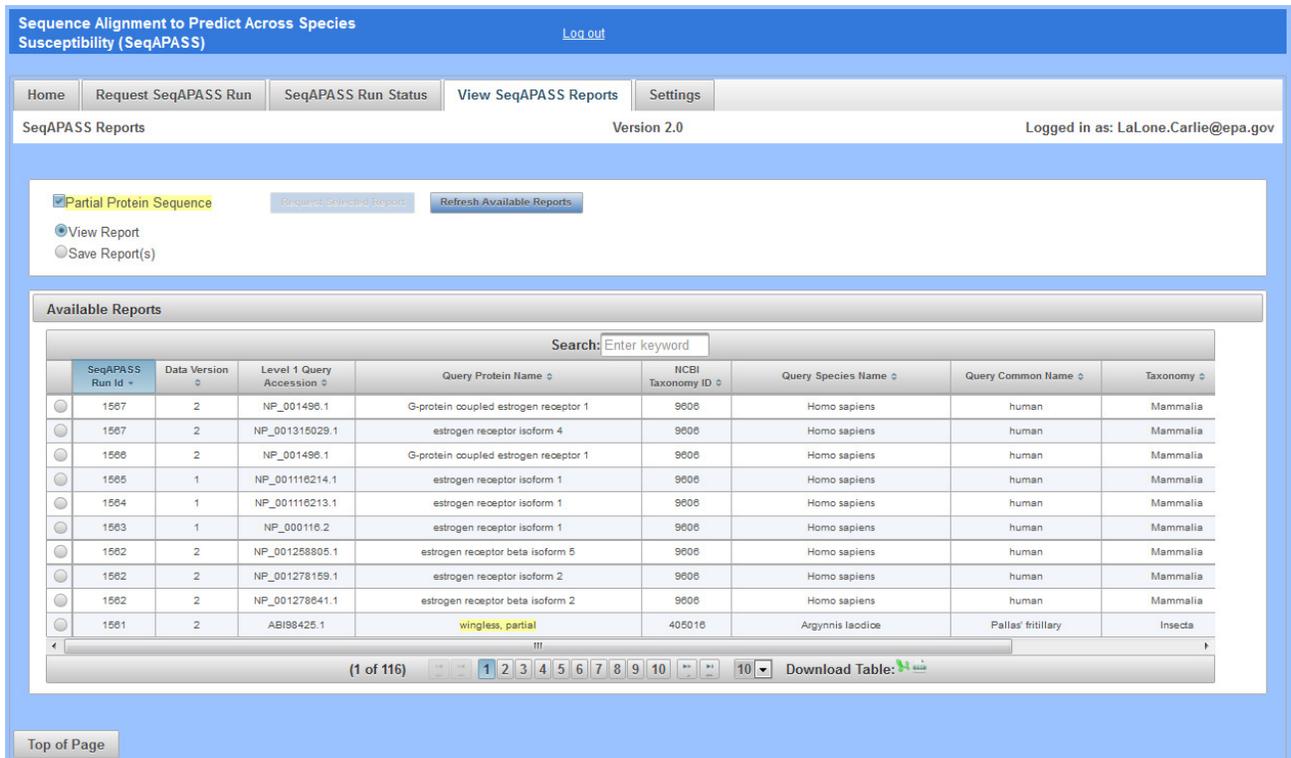
## View SeqAPASS Reports Tab

The “View SeqAPASS Reports” tab provides a table of completed SeqAPASS runs. From this page the user can choose to either “View Report” or “Save Report(s).”



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), Data Version, Ortholog Count (number of orthologs detected from the aligned hit sequences in Level 1; see Detailed Documentation page 79), NCBI Accession, Query Protein Name, taxonomy information for the query species, and the date/time of run completion.

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).



**Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide**  
 Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

*View Report*

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.

The screenshot shows the SeqAPASS web application interface. At the top, there is a navigation bar with 'Home', 'Request SeqAPASS Run', 'SeqAPASS Run Status', 'View SeqAPASS Reports', and 'Settings'. Below this is a header area with 'SeqAPASS Reports', 'Version 2.0', and 'Logged in as: LaLone.Carlie@epa.gov'. The main content area features a 'Partial Protein Sequence' checkbox, 'Request Selected Report', and 'Refresh Available Reports' buttons. Below these are radio buttons for 'View Report' and 'Save Report(s)'. The 'Available Reports' section contains a table with a search bar and the following data:

SeqAPASS Run ID	Data Version	Level 1 Query Accession	Query Protein Name	NCBI Taxonomy ID	Query Species Name	Query Common Name	Taxonomy
1567	2	NP_001498.1	G-protein coupled estrogen receptor 1	9606	Homo sapiens	human	Mammalia
1567	2	NP_001315029.1	estrogen receptor isoform 4	9606	Homo sapiens	human	Mammalia
1566	2	NP_001498.1	G-protein coupled estrogen receptor 1	9606	Homo sapiens	human	Mammalia
1565	1	NP_001118214.1	estrogen receptor isoform 1	9606	Homo sapiens	human	Mammalia
1564	1	NP_001118213.1	estrogen receptor isoform 1	9606	Homo sapiens	human	Mammalia
1563	1	NP_000116.2	estrogen receptor isoform 1	9606	Homo sapiens	human	Mammalia
1562	2	NP_001258805.1	estrogen receptor beta isoform 5	9606	Homo sapiens	human	Mammalia
1562	2	NP_001278159.1	estrogen receptor isoform 2	9606	Homo sapiens	human	Mammalia
1562	2	NP_001278641.1	estrogen receptor beta isoform 2	9606	Homo sapiens	human	Mammalia
1561	2	ABI98425.1	wingless, partial	405016	Argynnis laodice	Pallas' fritillary	Insecta

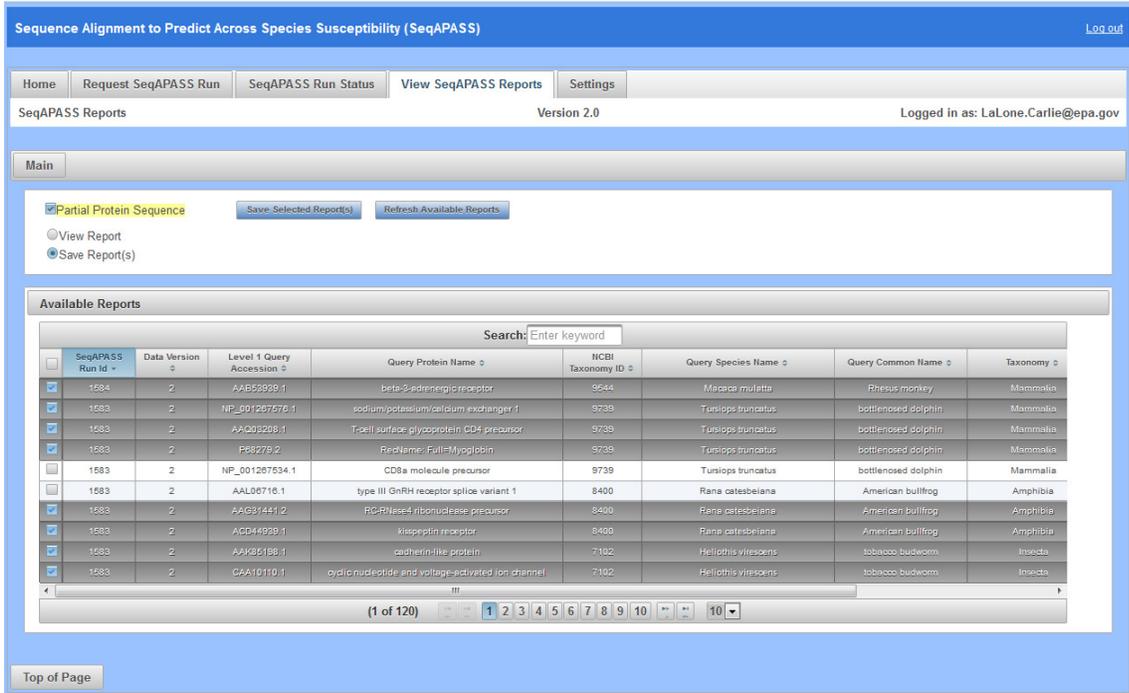
At the bottom of the table, there is a pagination bar showing '(1 of 116)' and a 'Download Table' button.

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

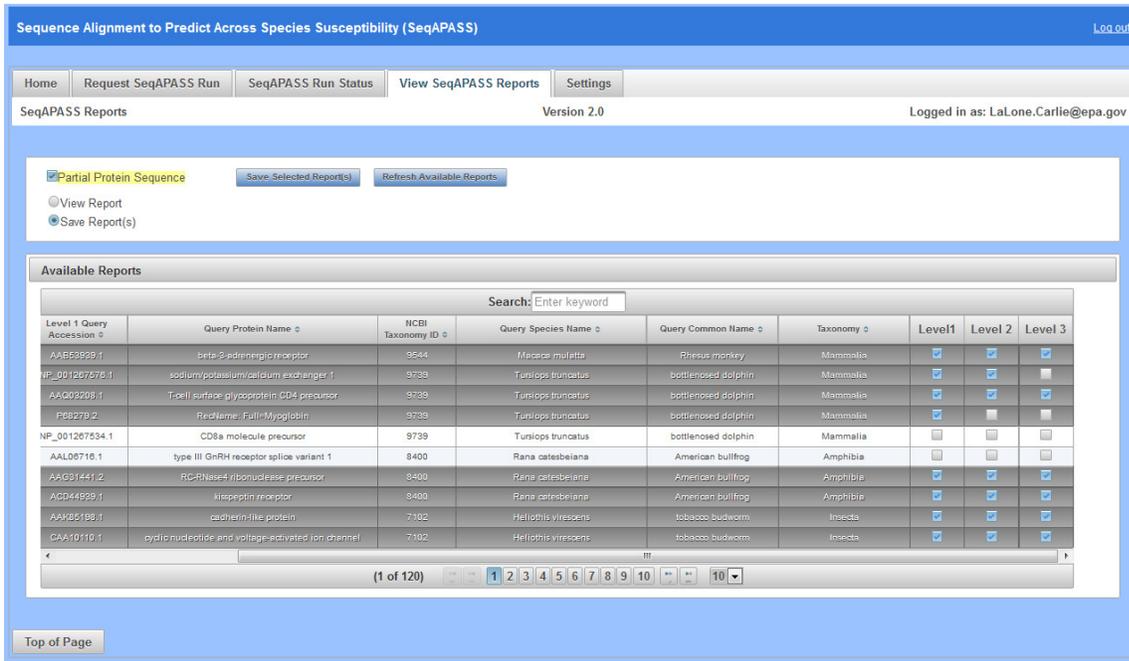
Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Save Report(s)

To download completed Level 1, 2, and/or 3 data, select the “Save Report(s)” radio button. Upon doing so the user can select which accession(s) to download by clicking the checkbox in the first column of the table associated with desired accession and click “Save Selected Report(s).”



The user can also deselect data that is not wanted in the download by scrolling to the far right of the table and deselecting the checkboxes for the different levels of the SeqAPASS analysis. By default, all available data for the selected accession will be downloaded in a zip file.



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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

A zip file will be created for all of the selected Reports. Ensure in the popup you select WinZip to open the file and click OK.

The screenshot shows the SeqAPASS web application interface. At the top, there are navigation tabs: Home, Request SeqAPASS Run, SeqAPASS Run Status, View SeqAPASS Reports, and Settings. Below the navigation is a section for 'SeqAPASS Reports' with a 'Version 2.0' indicator. There are buttons for 'Save Selected Report(s)' and 'Refresh Available Reports'. A 'Partial Protein Sequence' checkbox is checked, and there are radio buttons for 'View Report' and 'Save Report(s)'. Below this is a table of 'Available Reports' with a search bar. The table has columns for SeqAPASS Run Id, Data Version, Level 1 Query Accession, Query Protein Name, NCBI Taxonomy ID, and Query Species. A 'Top of Page' button is at the bottom left. A WinZip dialog box is open on the right, titled 'Opening seqapass.zip'. It shows the file 'seqapass.zip' which is a WinZip File from 'https://seqapassstage.rtpnc.epa.gov'. The dialog asks 'What should Firefox do with this file?' and has options for 'Open with WinZip (default)', 'Save File', and 'Do this automatically for files like this from now on'. 'OK' and 'Cancel' buttons are at the bottom right of the dialog.

SeqAPASS Run Id	Data Version	Level 1 Query Accession	Query Protein Name	NCBI Taxonomy ID	Query Species
1584	2	AAB53939.1	beta-3-adrenergic receptor	9544	Macaca m...
1583	2	NP_001267576.1	sodium/potassium/calcium exchanger 1	9739	Tursiops tru...
1583	2	AAQ03208.1	T-cell surface glycoprotein CD4 precursor	9739	Tursiops truncatus
1583	2	P68279.2	RecName: Full=Myoglobin	9739	Tursiops truncatus
1583	2	NP_001267534.1	CD8a molecule precursor	9739	Tursiops truncatus
1583	2	AAL06710.1	type III GnRH receptor splice variant 1	8400	Rana catesbeiana
1583	2	AAG31441.2	RC-RNase4 ribonuclease precursor	8400	Rana catesbeiana
1583	2	ACD44939.1	kisspeptin receptor	8400	Rana catesbeiana
1583	2	AAK85198.1	cadherin-like protein	7102	Heliothis virescens
1583	2	CAA10110.1	cyclic nucleotide and voltage-activated ion channel	7102	Heliothis virescens

A pop-up seqapass.zip file should appear with data files for each selected report. The naming convention is the NCBI Protein Accession and the Data Version (e.g., AAG31441.2\_v2).

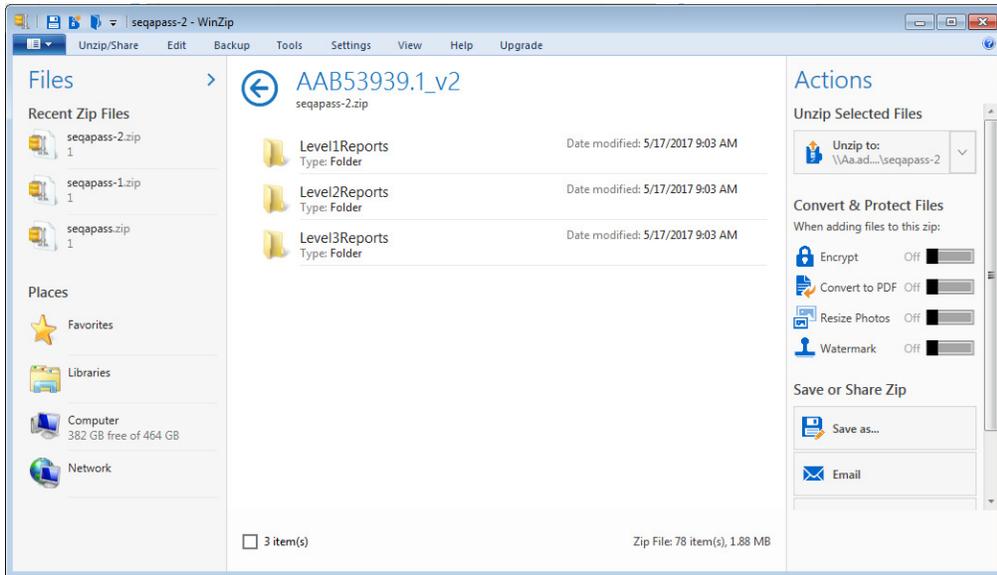
The screenshot shows a Windows File Explorer window titled 'seqapass - WinZip'. The address bar shows 'Unzip/Share' and the file name 'seqapass.zip'. The left sidebar shows 'Recent Zip Files' with 'seqapass.zip' listed. The main area shows the contents of the zip file, which are folders named with NCBI Protein Accession and Data Version: AAG31441.2\_v2, AAK85198.1\_v2, AAQ03208.1\_v2, ACD44939.1\_v2, CAA10110.1\_v2, NP\_001267576.1\_v2, and P68279.2\_v2. The right sidebar shows 'Actions' with options for 'Unzip All Files', 'Convert & Protect Files', and 'Save or Share Zip'.

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

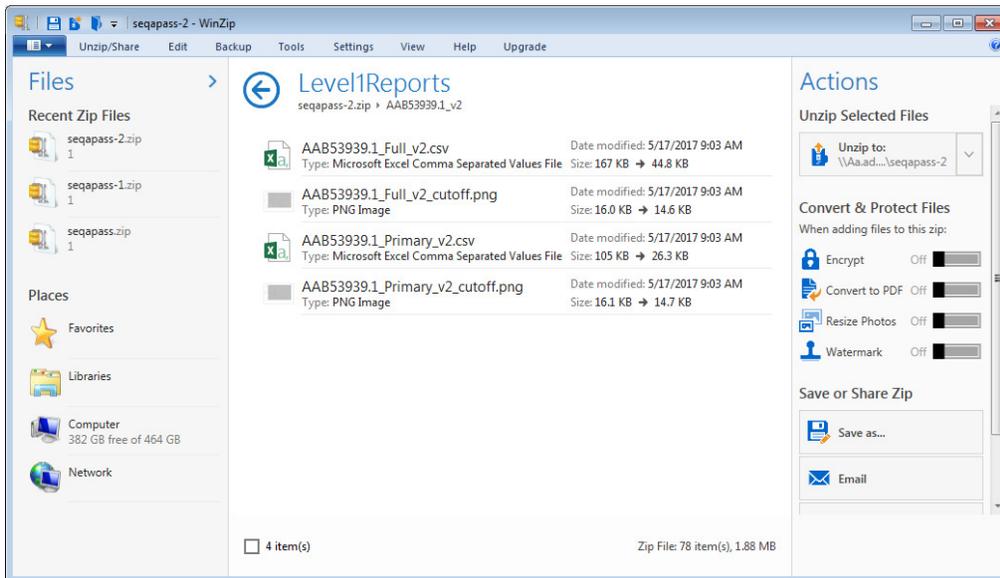
Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

By clicking on one of the Reports for a Protein Accession\_version, all available files for each Level of the SeqAPASS evaluation are available.

**Note:** This download includes default settings only. If susceptibility cut-off or any defaults were manipulated on Level 1 or 2 pages they will **NOT** be downloaded here and can **ONLY** be downloaded directly from the Level 1 or Level 2 page where the setting was manipulated by the user. Also, data visualizations can **ONLY** be downloaded from the Level 1 and 2 pages. They **DO NOT** populate in the zip file folders.



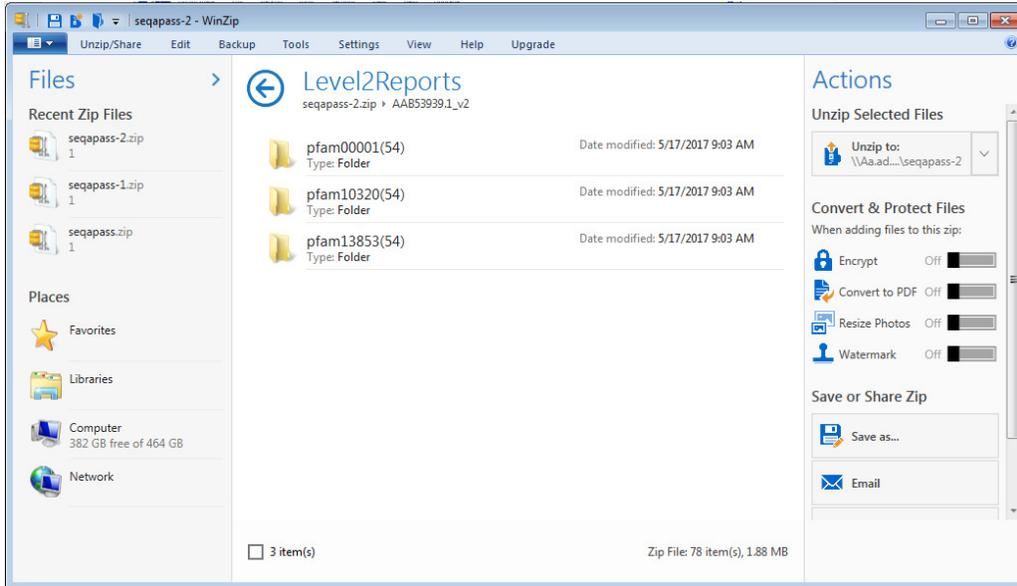
By selecting Level1Reports, both full and primary reports are available as csv files as well as a graphic of the density plot for determining the susceptibility cut-off.



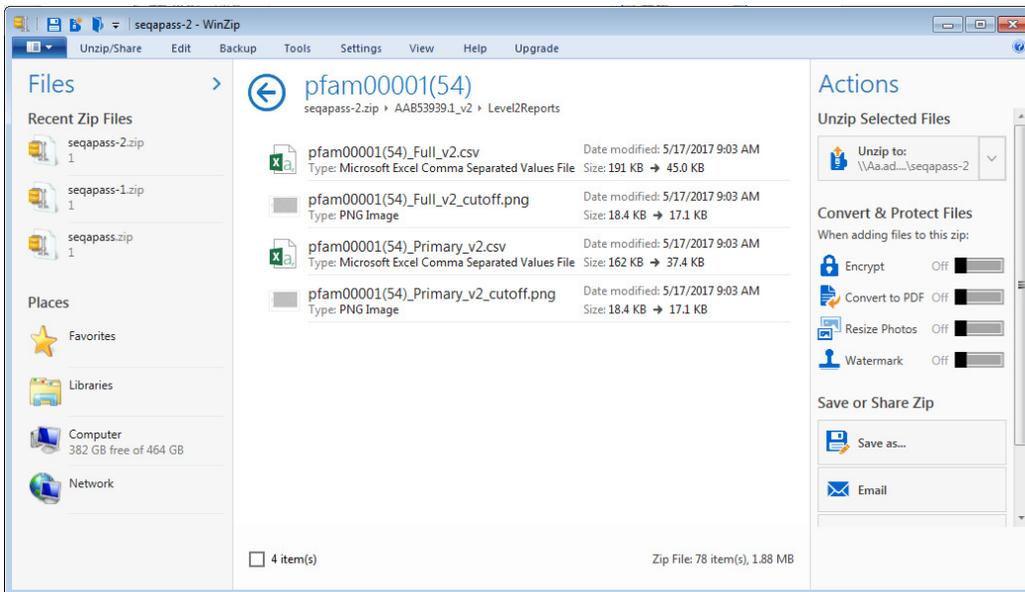
# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

By selecting Level2Reports, all completed domain comparisons will be available and named by NCBI domain accession with the starting amino acid residue position for the domain (e.g., pfam00001(54)).



Upon selecting a domain file to view, both full and primary reports are available as csv files as well as a graphic of the density plot for determining the susceptibility cut-off.

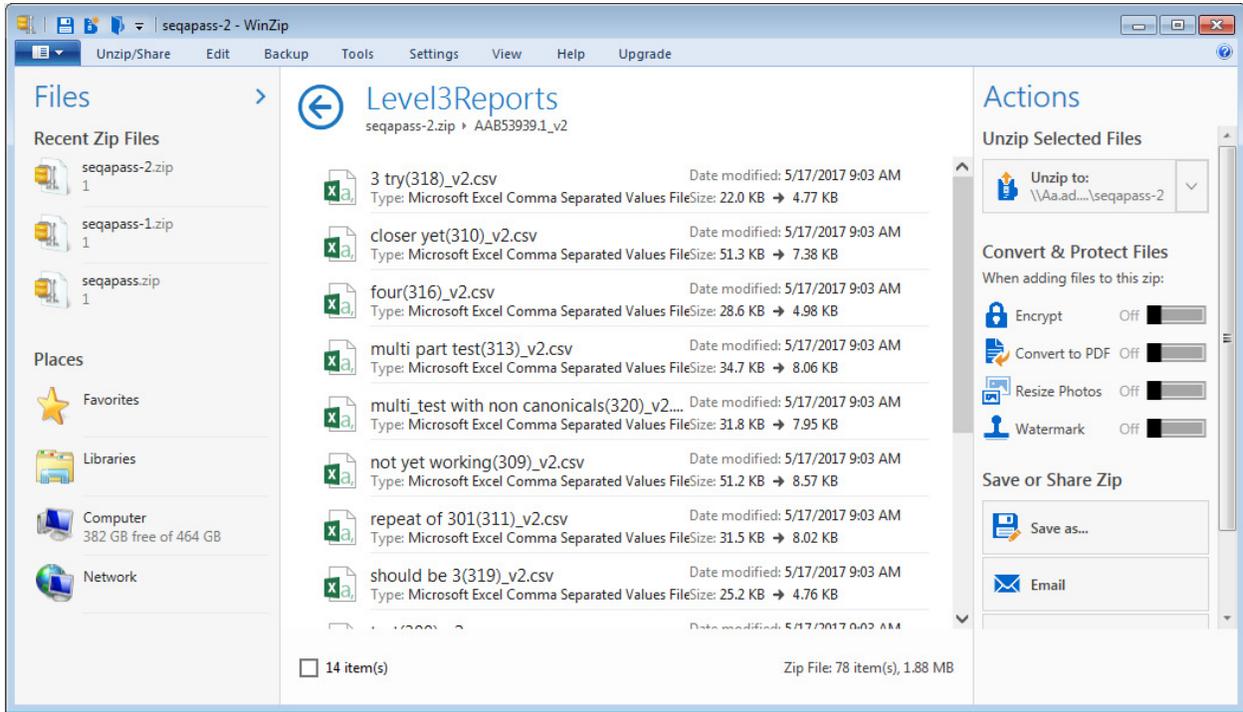


# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

By selecting Level3Reports, all user defined Level 3 alignments are available as csv.

**Note:** These csv files show the alignments across the entire sequence, not just those amino acid residues selected by the user.



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

Updated 3/06/18; 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 updates (“Protein and Taxonomy Data:” displays the date that NCBI databases were downloaded and incorporated into the SeqAPASS database; BLAST Version: and Software Version: displays the version being used by the SeqAPASS tool for the selected data), Query Species, and Query Protein. Other information in this box will be described below.



The screenshot shows the SeqAPASS web interface. At the top, there is a blue header with the text "Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)" and a "Log out" link. Below the header is a navigation bar with buttons for "Home", "Request SeqAPASS Run", "SeqAPASS Run Status", "View SeqAPASS Reports", and "Settings". The "View SeqAPASS Reports" button is selected. Below the navigation bar, there is a sub-header for "SeqAPASS Reports" with "Version 2.0" and "Logged in as: LaLone.Carlie@epa.gov". Below this is another navigation bar with "Main" and "Level 1" buttons. The "Level 1" button is selected. Below the navigation bar is a section titled "Level 1 Query Protein Information". This section contains the following text: "Hit proteins are identified for the following query protein. Use the main button to go back to the SeqAPASS Reports list." Below this text are several key-value pairs: "SeqAPASS ID: 1563", "Query Accession: [NP\\_000118.2](#)", "Ortholog Count: 281", "Protein and Taxonomy Data: 11/09/2016", "Query Species: Homo sapiens", "BLAST Version: 2.3.0", and "Query Protein: estrogen receptor isoform 1", "Software Version: 1.0".

The default table displayed at the bottom of the page is the “Primary Report”, which includes query protein information in the first row below the column titles, 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, Data version, NCBI Accession and species information is provided including the “Protein Count” which indicates the number of protein records per species in the NCBI protein database, taxonomic information (See “Primary Report Settings” section below in user guide for more detail on “Taxonomic Group” versus “Filtered Taxonomic Group” columns), 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 ([hit bitscore/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. 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. Additionally, there is a column that identifies if the species is a Eukaryote noted with a “Y” for yes or alternatively an “N” for no. Links out to the NCBI Protein Database and NCBI Taxonomy Database (specific to the data row) are embedded in the Level 1 data table 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).

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

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 or check the full report). Please see “Susceptibility Cutoff Box for Level 1” section of user guide for details when no orthologs are detected. Additionally, the default setting for the report shows only eukaryote data if a eukaryote is selected as the query protein, excluding prokaryote data from the table with the “Show Only Eukaryotes” checkbox checked. To view prokaryote data, deselect this checkbox. If a prokaryote is selected as the query protein, the default setting will include both eukaryote and prokaryote data and the “Show Only Eukaryotes” checkbox will not be selected. To limit the data to eukaryotes only, the user would check the “Show Only Eukaryotes” checkbox.

Columns in left side of table:

Level 1 Data - Primary

Search: Enter keyword

Data Version	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Filtered Taxonomic Group	Scientific Name	Common Name	Protein Name
1	<a href="#">NP_000116.2</a>	1058918	<a href="#">9606</a>	Mammalia	Mammalia	<a href="#">Homo sapiens</a>	human	<a href="#">estrogen receptor isoform 1</a>
1	<a href="#">ABY64717.1</a>	2004	<a href="#">9593</a>	Mammalia	Mammalia	<a href="#">Gorilla gorilla</a>	western gorilla	<a href="#">estrogen receptor alpha</a>
1	<a href="#">XP_003311596.1</a>	113984	<a href="#">9598</a>	Mammalia	Mammalia	<a href="#">Pan troglodytes</a>	chimpanzee	<a href="#">PREDICTED: estrogen receptor isoform X2</a>
1	<a href="#">XP_018884801.1</a>	68319	<a href="#">9595</a>	Mammalia	Mammalia	<a href="#">Gorilla gorilla gorilla</a>	western lowland gorilla	<a href="#">PREDICTED: estrogen receptor isoform X2</a>
1	<a href="#">XP_003811544.1</a>	49145	<a href="#">9597</a>	Mammalia	Mammalia	<a href="#">Pan paniscus</a>	pygmy chimpanzee	<a href="#">PREDICTED: estrogen receptor isoform X2</a>
1	<a href="#">ABY64718.1</a>	1703	<a href="#">9600</a>	Mammalia	Mammalia	<a href="#">Pongo pygmaeus</a>	Bornean orangutan	<a href="#">estrogen receptor alpha</a>
1	<a href="#">XP_002817538.1</a>	44332	<a href="#">9601</a>	Mammalia	Mammalia	<a href="#">Pongo abelii</a>	Sumatran orangutan	<a href="#">PREDICTED: estrogen receptor isoform X2</a>
1	<a href="#">XP_011922091.1</a>	66739	<a href="#">9531</a>	Mammalia	Mammalia	<a href="#">Cercocebus atys</a>	sooty mangabey	<a href="#">PREDICTED: estrogen receptor isoform X2</a>
1	<a href="#">XP_014982596.1</a>	86552	<a href="#">9544</a>	Mammalia	Mammalia	<a href="#">Macaca mulatta</a>	Rhesus monkey	<a href="#">PREDICTED: estrogen receptor isoform X2</a>
1	<a href="#">XP_005552209.1</a>	121048	<a href="#">9541</a>	Mammalia	Mammalia	<a href="#">Macaca fascicularis</a>	crab-eating macaque	<a href="#">PREDICTED: estrogen receptor isoform X1</a>

(1 of 82) Download Table:

Top of Page

Columns in right side of table:

Level 1 Data - Primary

Search: Enter keyword

Common Name	Protein Name	BLASTp Bitscore	Ortholog Candidate	Ortholog Count	Cut-off	Percent Similarity	Susceptibility Prediction	Analysis Completed	Eukaryote
human	<a href="#">estrogen receptor isoform 1</a>	1241.9	Y	281	33.9	100.0	Y	2017 03 02 16:17:23	Y
western gorilla	<a href="#">estrogen receptor alpha</a>	1229.5	Y	281	33.9	99.0	Y	2017 03 02 16:17:23	Y
chimpanzee	<a href="#">PREDICTED: estrogen receptor isoform X2</a>	1229.5	Y	281	33.9	99.0	Y	2017 03 02 16:17:23	Y
western lowland gorilla	<a href="#">PREDICTED: estrogen receptor isoform X2</a>	1228.8	Y	281	33.9	98.9	Y	2017 03 02 16:17:23	Y
pygmy chimpanzee	<a href="#">PREDICTED: estrogen receptor isoform X2</a>	1228.0	Y	281	33.9	98.9	Y	2017 03 02 16:17:23	Y
Bornean orangutan	<a href="#">estrogen receptor alpha</a>	1227.6	Y	281	33.9	98.9	Y	2017 03 02 16:17:23	Y
Sumatran orangutan	<a href="#">PREDICTED: estrogen receptor isoform X2</a>	1227.6	Y	281	33.9	98.9	Y	2017 03 02 16:17:23	Y
sooty mangabey	<a href="#">PREDICTED: estrogen receptor isoform X2</a>	1227.2	Y	281	33.9	98.8	Y	2017 03 02 16:17:23	Y
Rhesus monkey	<a href="#">PREDICTED: estrogen receptor isoform X2</a>	1227.2	Y	281	33.9	98.8	Y	2017 03 02 16:17:23	Y
crab-eating macaque	<a href="#">PREDICTED: estrogen receptor isoform X1</a>	1227.2	Y	281	33.9	98.8	Y	2017 03 02 16:17:23	Y

(1 of 82) Download Table:

Top of Page

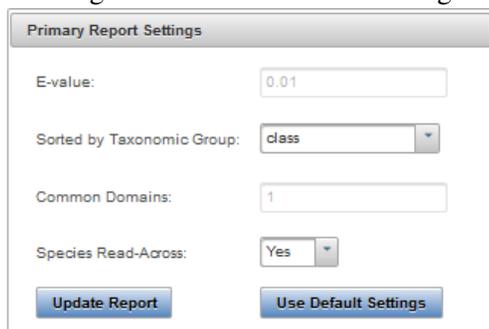
## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### Level 1: Primary Report Settings

#### Default settings:

The “Primary Report Settings” box allows the user to view default settings on the table below and manipulate certain settings. The “Primary Report Settings box is only available on the “Primary Report” display, not the “Full Report.” The default settings show data for hits whose E-value are  $\leq 0.01$  and have been identified to have  $\geq 1$  domain in common with the query sequence. The default setting for the “Sorted by Taxonomic Group” is “class,” therefore the “Filtered Taxonomic Group” column in the table is set to identify and report the taxonomic lineage of “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 subclass, to superorder, to order, to suborder, to superfamily, to family, to subfamily, 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.



Primary Report Settings

E-value:

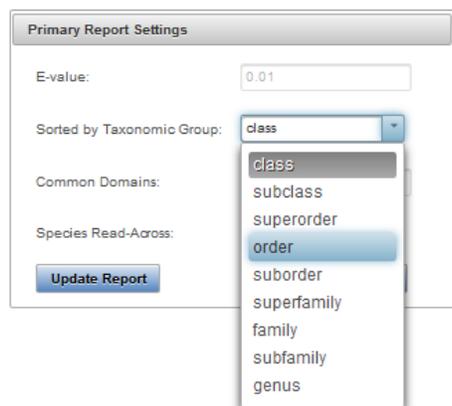
Sorted by Taxonomic Group:

Common Domains:

Species Read-Across:

#### Changing Default Settings:

The “E-value” and “Common Domains” settings can be manipulated by the user by entering the desired E-value or number of Common Domains in the respective text boxes and clicking “Update Report.” The table and data visualization will automatically be updated after a few seconds. The user may choose to change the level of the taxonomic hierarchy that is used for the susceptibility prediction. From the Sorted by Taxonomic Group” dropdown the user may choose to display a different taxonomic group in the “Filtered Taxonomic Group” column of the data table.



Primary Report Settings

E-value:

Sorted by Taxonomic Group:

Common Domains:

Species Read-Across:

- class
- subclass
- superorder
- order
- suborder
- superfamily
- family
- subfamily
- genus

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

If the user chooses “order” for example, the “Filtered Taxonomic Group” column in the data table will report the taxonomic lineage of “order” from the NCBI Taxonomy Database and all species read-across for the susceptibility prediction will be based on order instead of class. The data visualization will also update. As described previously, if order 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 suborder, to superfamily, to family, to subfamily, to genus. Upon selecting the Taxonomic Group from the dropdown and clicking “Update Report,” the Level 1 Data for the Primary report will update to the selected taxonomic level.

The screenshot displays the SeqAPASS web interface. On the left, the "Primary Report Settings" panel includes fields for E-value (0.01), Sorted by Taxonomic Group (order), Common Domains (1), and Species Read-Across (Yes). In the center, there are dropdowns for "Choose Domain to View" and "Choose Query to View", both set to "-Select Completed Domain -". On the right, there are sections for "NCBI COBALT" and "NCBI Taxonomy Database" with a "Request Residue Run" button. Below these is a "Level 1 Data - Primary" table with a search bar and a table of results. The table has columns for Data Version, NCBI Accession, Protein Count, Species Tax ID, Taxonomic Group, Filtered Taxonomic Group, Scientific Name, and Common Name. The table shows 10 rows of data for various primate species. At the bottom, there is a pagination bar showing "(1 of 82)" and a "Download Table" button.

Data Version	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Filtered Taxonomic Group	Scientific Name	Common Name
2	<a href="#">NP_000116.2</a>	1058918	<a href="#">9506</a>	Mammalia	Primates	<a href="#">Homo sapiens</a>	human
2	<a href="#">ABY64717.1</a>	2004	<a href="#">9593</a>	Mammalia	Primates	<a href="#">Gorilla gorilla</a>	western gorilla
2	<a href="#">XP_003311596.1</a>	113964	<a href="#">9598</a>	Mammalia	Primates	<a href="#">Pan troglodytes</a>	chimpanzee
2	<a href="#">XP_018884801.1</a>	68319	<a href="#">9595</a>	Mammalia	Primates	<a href="#">Gorilla gorilla gorilla</a>	western lowland gorilla
2	<a href="#">XP_003811544.1</a>	49145	<a href="#">9597</a>	Mammalia	Primates	<a href="#">Pan panisus</a>	pygmy chimpanzee
2	<a href="#">ABY64718.1</a>	1703	<a href="#">9500</a>	Mammalia	Primates	<a href="#">Pongo pygmaeus</a>	Bornean orangutan
2	<a href="#">XP_002817538.1</a>	44332	<a href="#">9501</a>	Mammalia	Primates	<a href="#">Pongo abelii</a>	Sumatran orangutan
2	<a href="#">XP_011922091.1</a>	66739	<a href="#">9531</a>	Mammalia	Primates	<a href="#">Cercocebus atys</a>	sooty mangabey
2	<a href="#">XP_014992596.1</a>	86552	<a href="#">9544</a>	Mammalia	Primates	<a href="#">Macaca mulatta</a>	Rhesus monkey
2	<a href="#">XP_005562209.1</a>	121048	<a href="#">9541</a>	Mammalia	Primates	<a href="#">Macaca fascicularis</a>	crab-eating macaque

The user may also choose to turn species read-across off, by using the “Species Read-Across” drop-down and selecting “No” and clicking “Update Report.” When “No” is selected, the susceptibility predictions will only be “Y” in the table below if Percent Similarity is above the Cut-off or if the hit is identified as an Ortholog Candidate, yes or “Y.” Any hit below the cut-off will yield a susceptibility prediction of no or “N.”

This screenshot shows the "Primary Report Settings" panel with the "Species Read-Across" dropdown menu open. The menu options are "Yes" and "No". The "Update Report" button is visible at the bottom left of the panel.

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

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).

Level 1 Data - Full

Search: Enter keyword

Accession	Hit Length	Identity	Positives	E-value	BLASTp Bitscore	Ortholog Candidate	Ortholog Count	Cut-off	Common Domain Count	Percent Similarity	Susceptibility Prediction	Analysis Completed	Eukaryote
<a href="#">isoform_1</a>	595	595	595	0.000E0	1241.9	Y	281	33.9	78	100.0	Y	2017 03 02 16:17:23	Y
<a href="#">r_alpha</a>	595	590	592	0.000E0	1229.5	Y	281	33.9	75	99.0	Y	2017 03 02 16:17:23	Y
<a href="#">ceptor isoform X2</a>	595	590	592	0.000E0	1229.5	Y	281	33.9	75	99.0	Y	2017 03 02 16:17:23	Y
<a href="#">ceptor isoform X2</a>	595	589	592	0.000E0	1228.8	Y	281	33.9	75	98.9	Y	2017 03 02 16:17:23	Y
<a href="#">ceptor isoform X2</a>	595	589	592	0.000E0	1228.0	Y	281	33.9	75	98.9	Y	2017 03 02 16:17:23	Y
<a href="#">r_alpha</a>	595	589	591	0.000E0	1227.6	Y	281	33.9	75	98.9	Y	2017 03 02 16:17:23	Y
<a href="#">ceptor isoform X2</a>	595	589	591	0.000E0	1227.6	Y	281	33.9	75	98.9	Y	2017 03 02 16:17:23	Y
<a href="#">ceptor isoform X2</a>	595	588	592	0.000E0	1227.2	Y	281	33.9	75	98.8	Y	2017 03 02 16:17:23	Y
<a href="#">ceptor isoform X2</a>	595	588	592	0.000E0	1227.2	Y	281	33.9	75	98.8	Y	2017 03 02 16:17:23	Y
<a href="#">ceptor isoform X1</a>	595	588	592	0.000E0	1227.2	Y	281	33.9	75	98.8	Y	2017 03 02 16:17:23	Y

(1 of 85) 1 2 3 4 5 6 7 8 9 10 10 Download Table:

Top of Page

Note: SeqAPASS v2.0 and newer parse the BLASTP query and hit accessions to identify all the species/accessions from identical proteins. Therefore, if a hit sequence represents multiple species, all species with the identical sequence will be found in the data tables for Level 1. To determine which sequence/species was identified from BLASTP as a hit and which sequence/species was parsed from the identical sequence, view the “Full Report” for Level, column “Identical Protein,” where “N” is indicative of the original hit sequence and “Y” is the parsed sequence.

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

Updated 3/06/18; 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, which will open a new tab in the web browser. 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 2<sup>nd</sup> local minimum, where the 2<sup>nd</sup> 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, which again opens a new tab in the web browser. Alternatively, the user can view and closely examine the density plot and manipulate the cut-off by clicking the “View Cutoff” button.



Upon clicking “View Cutoff” button, a new tab opens in the web browser with a page that displays 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 close the cutoff tab and return to the Level 1 page, click “Update Cut-off” button.

Note: The user should have a justification for changing the susceptibility cut-off, either based on evaluation of Ortholog cutoffs in the data visualization or from empirical evidence.

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

Updated 3/06/18; 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.



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 3/06/18; 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 checks the full report for Ortholog candidates or identifies 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.

**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: 1570    Query Accession: [NP\\_051217544.1](#)    Ortholog Count: 0    Protein and Taxonomy Date: 01/04/2017  
 Query Species: Homo sapiens    BLAST Version: 2.5.0  
 Query Protein: perisikome proliferator-activated receptor gamma isoform 3    Software Version: 2.0

**Susceptibility Cut-off**

Default  
 Second Local Minimum  
 User Defined

**View Cut-off**

**Primary Report Settings**

E-value:

Sorted by Taxonomic Group:

Common Domains:

Species Read-Access:

**Update Report**    **Use Default Settings**

**Level 2**

Level 2 Query Domain

[NCBI Conserved Domain Database](#)

Functional Domains

Select Domain:

**Request Domain Run**

View Level 2 Data

Choose Domain to View

Select Completed Domain:

**View Level 2 Data**

**Level 3**

Level 3 Query Amino Acid Residue(s)

[NCBI Protein Database](#)

Select Template Sequence

Additional Comparisons (optional)

[NCBI COBALT](#)

Enter Level 3 Run Name

[NCBI Taxonomy Database](#)

Choose Taxonomic Group(s)

All Groups

Use table below to select sequences

0 species selected

**Request Residue Run**

View Level 3 Data

Choose Query to View

Select Level 3 Run Name:

**View Level 3 Data**

Primary Report     Partial Hit Protein Sequences

Full Report     Exclude Orthologs with 0 hits

Show Only Eukaryotes     Exclude Species with Missing Columns

**Level 1 Data - Primary**

Search:

Data Version	NCBI Accession ID	Protein Count	Species Tax ID	Taxonomic Group	Filled Taxonomic Group	Scientific Name	Common Name	Protein Name	BLAST Ebitscore	Ortholog Candidate	Ortholog Count	Cut-off	Percent Similarity
2	<a href="#">XP_01152145.1</a>	1055910	9558	Mammalia	Mammalia	<a href="#">Homo sapiens</a>	Human	<a href="#">PREDICTED: perisikome proliferator-activated receptor gamma</a>	527.2	Y	0	47.5	100.0
2	<a href="#">XP_013890300.1</a>	82319	9558	Mammalia	Mammalia	<a href="#">Bovis gorilla gorilla</a>	western lowland gorilla	<a href="#">PREDICTED: perisikome proliferator-activated receptor gamma</a>	521.9	N	0	47.0	99.9
2	<a href="#">XP_003849362.1</a>	49145	9557	Mammalia	Mammalia	<a href="#">Pan paniscus</a>	pygmy chimpanzee	<a href="#">PREDICTED: perisikome proliferator-activated receptor gamma</a>	520.9	N	0	47.0	99.9
2	<a href="#">XP_002280027.1</a>	38938	81883	Mammalia	Mammalia	<a href="#">Nomascus leucogenus</a>	northern white-cheeked gibbon	<a href="#">PREDICTED: perisikome proliferator-activated receptor gamma</a>	520.4	N	0	47.0	99.7
2	<a href="#">XP_015799731.1</a>	113964	9559	Mammalia	Mammalia	<a href="#">Pan troglodytes</a>	chimpanzee	<a href="#">PREDICTED: perisikome proliferator-activated receptor gamma</a>	520.0	N	0	47.0	99.6
2	<a href="#">XP_011977719.1</a>	33679	239933	Mammalia	Mammalia	<a href="#">Colobus polycomis polycomis</a>	Angolan colobus	<a href="#">PREDICTED: perisikome proliferator-activated receptor gamma</a>	519.9	N	0	47.0	99.4
2	<a href="#">NP_05124666.1</a>	44332	9601	Mammalia	Mammalia	<a href="#">Pongo abelii</a>	Burman orangutan	<a href="#">perisikome proliferator-activated receptor gamma</a>	519.1	N	0	47.0	99.2
2	<a href="#">XP_01172698.1</a>	8508	9548	Mammalia	Mammalia	<a href="#">Macaca nemestrina</a>	pig-tailed macaque	<a href="#">PREDICTED: perisikome proliferator-activated receptor gamma</a>	516.9	N	0	47.0	99.0

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## 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.

**Note:** De-select the “Show Only Eukaryotes” checkbox to see if prokaryotes were identified as orthologs.

### 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

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

Query Species: Bubalus bubalis

Query Protein: insulin receptor

Select Cut-off:  Enter Cut-off:

#### Density Plot

Cut-off Based on Ortholog Candidates

Cut-off #	Susceptibility Cut-off
1	49.0

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

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.

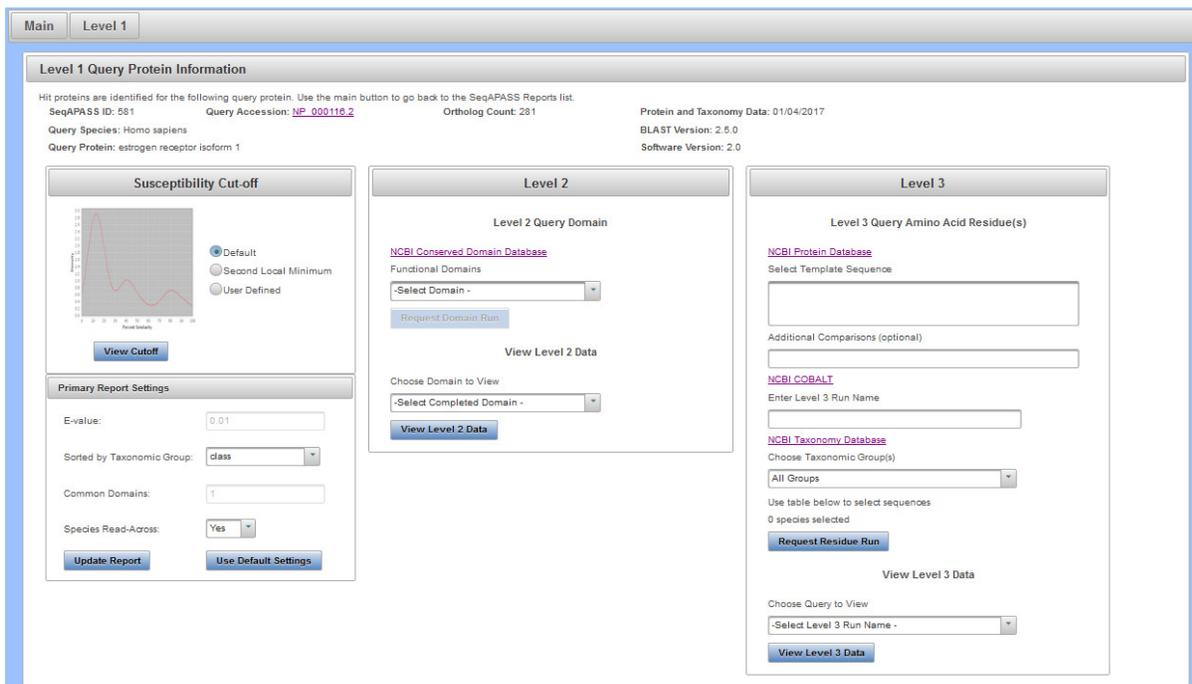
The screenshot displays the SeqAPASS web application interface. At the top, a blue header contains the title "Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)" and a "Log out" link. Below the header is a navigation bar with buttons for "Home", "Request SeqAPASS Run", "SeqAPASS Run Status", "View SeqAPASS Reports", and "Settings". The main content area is titled "Level 1 Query Protein Information" and includes a "Main" button and a "Level 1" tab. The page displays query details: SeqAPASS ID: 1570, Query Accession: [XP\\_021517544.1](#), Ortholog Count: 0, Protein and Taxonomy Data: 01/04/2017, BLAST Version: 2.5.0, and Software Version: 2.0. The query species is Homo sapiens and the query protein is peroxisome proliferator-activated receptor gamma isoform 3. Three main panels are visible: "Susceptibility Cut-off" with a graph and "View Cut-off" button; "Level 2 Query Domain" with a "Functional Domains" dropdown and "View Level 2 Data" button; and "Level 3 Query Amino Acid Residue(s)" with a "Request Residue Run" button. A "Primary Report Settings" panel is also present with fields for E-value, taxonomic group, common domains, and species read-across.

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

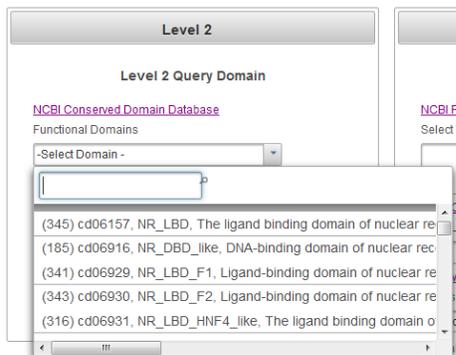
## Level 2: Functional Domain(s) Alignment

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 3/06/18; 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 hovering over the NR\_LBD\_ER domain with the computer mouse.

The screenshot displays the NCBI Conserved Domains Database interface. At the top, the NCBI logo and navigation menu are visible. The main content area shows a search for "estrogen receptor isoform 1 [Homo sapiens]" with a graphical summary of the protein sequence (residues 1-595). Specific hits are highlighted, including NR\_LBD\_ER. A detailed view of the NR\_LBD\_ER domain (cd06949) is shown, including its description and a sequence alignment with the query sequence. The description states: "[Specific hit, evalue = 1.46e-146]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 is as a transcription factor by binding to the estrogen response element of target genes upon activation by estrogen and".

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-

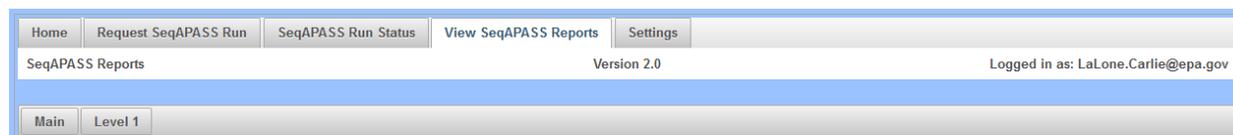
## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 3/06/18; 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.

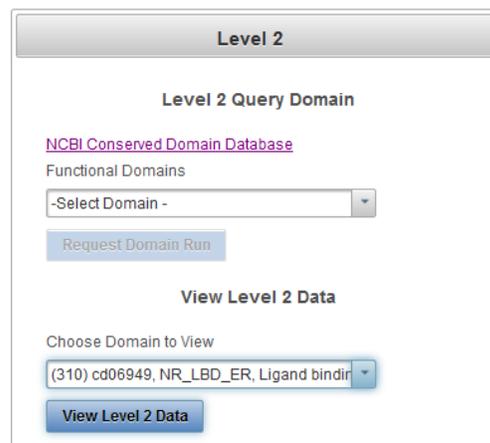
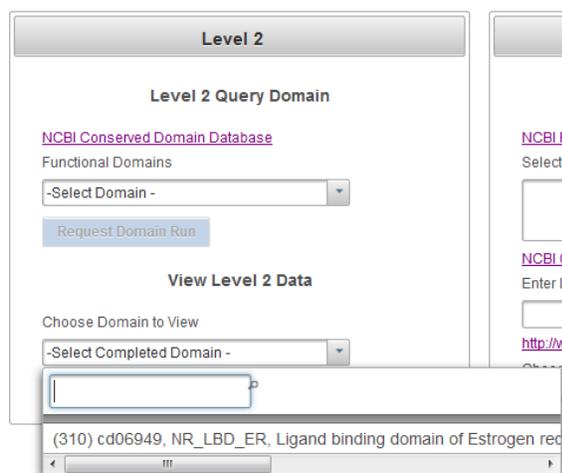


When sequence comparisons have completed for the selected functional domain, the domain will be present in the drop-down in the View Level 2 Data area. The **drop-down is not automatically populated** with the completed domain run. The **user must click on the “Level 1” button to update the page** for 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 3/06/18; 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 updates (“Protein and Taxonomy Data:” and “CDD Data:” display the dates that NCBI databases were downloaded and incorporated into the SeqAPASS database; “BLAST version:” and “Software Version:” displays the version being used by the SeqAPASS tool for the selected data), Query Species, Query Domain (with link out to NCBI domain page), Query Protein name.

The screenshot displays the SeqAPASS Reports interface. At the top, there are navigation tabs: Home, Request SeqAPASS Run, SeqAPASS Run Status, View SeqAPASS Reports (selected), and Settings. Below the tabs, the page title is "SeqAPASS Reports" and the version is "Version 2.0". The user is logged in as "LaLone.Carlie@epa.gov".

The main content area is titled "Level 2 Query Domain Information". 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: 581
- Query Accession: [NP\\_000116.2](#)
- Ortholog Count: 281
- Protein and Taxonomy Data: 01/04/2017
- Query Species: Homo sapiens
- BLAST Version: 2.5.0
- Query Domain: (310) [p006849](#), NR\_LBD\_ER , Ligand binding domain of Estrogen receptor, which are activated by the hormone 17beta-estradiol (estrogen)
- CDD Data: 04/25/2015
- Query Protein: estrogen receptor isoform 1
- Software Version: 2.0

There are two interactive panels:

- Susceptibility Cut-off:** A graph showing a curve with a peak and a valley. Three radio buttons are present: "Default" (selected), "Second Local Minimum", and "User Defined". A "View Cutoff" button is at the bottom.
- Primary Report Settings:** A form with the following fields:
  - E-value: 10.0
  - Sorted by Taxonomic Group: class
  - Species Read-Across: YesButtons for "Update Report" and "Use Default Settings" are at the bottom.

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, Data Version, NCBI Accession and 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 ([hit bitscore/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). Additionally, there is a column that identifies if the species is a eukaryote, noted with a “Y” for yes or alternatively a “N” for no if the hit is a prokaryote.

**Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide**  
 Updated 3/06/18; 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). Additionally, the default setting for the report shows only eukaryote data, excluding prokaryote data from the table with the “Shows Only Eukaryotes” checkbox checked. To view prokaryote data deselect this checkbox.

The screenshot shows the SeqAPASS web interface. At the top left, there are search filters:
 

- Partial Hit Protein Sequence
- Percent Similarity > 100%
- Susceptible = Y, Ortholog Count = 1
- Show Only Eukaryotes

 Below the filters is a section titled "Level 2 Data - Primary" with a search bar containing "Enter keyword". The main part of the interface is a table with the following columns:
 

Data Version	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Filtered Taxonomic Group	Scientific Name	Common Name	Protein Name
2	NP_000116.2	1058918	9506	Mammalia	Mammalia	<a href="#">Homo sapiens</a>	human	<a href="#">estrogen receptor isoform 1</a>
2	XP_005552209.1	121048	9541	Mammalia	Mammalia	<a href="#">Macaca fascicularis</a>	crab-eating macaque	<a href="#">PREDICTED_estrogen receptor isoform X1</a>
2	ABY64721.1	909	9534	Mammalia	Mammalia	<a href="#">Chlorocebus aethiops</a>	grivet	<a href="#">estrogen receptor alpha</a>
2	XP_011781932.1	85086	9545	Mammalia	Mammalia	<a href="#">Macaca nemestrina</a>	pig-tailed macaque	<a href="#">PREDICTED_estrogen receptor isoform X2</a>
2	NP_001158059.1	77385	9555	Mammalia	Mammalia	<a href="#">Papio anubis</a>	olive baboon	<a href="#">estrogen receptor</a>
2	XP_011922091.1	86739	9531	Mammalia	Mammalia	<a href="#">Cercopithecus atys</a>	sooty mangabey	<a href="#">PREDICTED_estrogen receptor isoform X2</a>
2	ABY64717.1	2004	9593	Mammalia	Mammalia	<a href="#">Gorilla gorilla</a>	western gorilla	<a href="#">estrogen receptor alpha</a>
2	XP_014892586.1	86552	9544	Mammalia	Mammalia	<a href="#">Macaca mulatta</a>	Rhesus monkey	<a href="#">PREDICTED_estrogen receptor isoform X2</a>
2	XP_002255939.1	38938	61853	Mammalia	Mammalia	<a href="#">Nomascus leucogenys</a>	northern white-cheeked gibbon	<a href="#">PREDICTED_estrogen receptor isoform X3</a>
2	XP_008005788.1	82274	60711	Mammalia	Mammalia	<a href="#">Chlorocebus sabaeus</a>	green monkey	<a href="#">PREDICTED_estrogen receptor isoform X2</a>

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

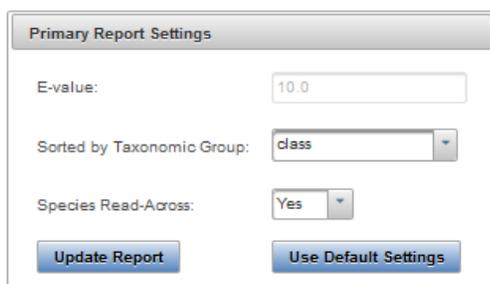
## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### Level 2: Primary Report Settings

#### *Default settings:*

The “Primary Report Settings” box allows the user to view default settings on the table below and manipulate certain settings. The “Primary Report Settings box is only available on the “Primary Report” display. The default settings show data for hits whose E-value are  $\leq 10$ . The default setting for the “Sorted by Taxonomic Group” is “class,” therefore the “Filtered Taxonomic Group” column in the table is set to identify and report the taxonomic lineage of “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 subclass, to superorder, to order, to suborder, to superfamily, to family, to subfamily, 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.



Primary Report Settings

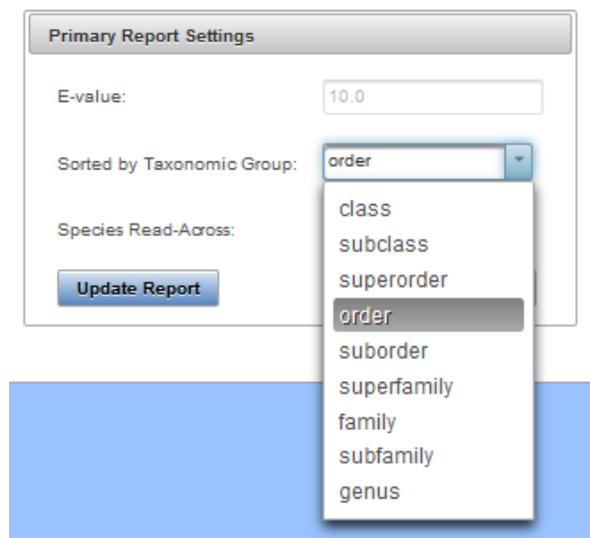
E-value:

Sorted by Taxonomic Group:

Species Read-Across:

#### *Changing Default Settings:*

The user may choose to change the level of the taxonomic hierarchy that is used for the susceptibility prediction. From the “Sorted by Taxonomic Group” dropdown the user may choose to display a different taxonomic group in the “Filtered Taxonomic Group” column of the data table.



Primary Report Settings

E-value:

Sorted by Taxonomic Group:

Species Read-Across:

- class
- subclass
- superorder
- order**
- suborder
- superfamily
- family
- subfamily
- genus

**Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide**  
 Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

If the user chooses “order” for example, the “Filtered Taxonomic Group” column in the data table will report the taxonomic lineage of “order” from the NCBI Taxonomy Database and all species read-across for the susceptibility prediction will be based on order instead of class. As described previously, if order 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 suborder, to superfamily, to family, to subfamily, to genus. Upon selecting the Taxonomic Group from the dropdown and clicking “Update Report,” the Level 1 Data for the Primary report will update to the selected taxonomic level.

### 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: 581      Query Accession: [NP\\_000116.2](#)      Ortholog Count: 281

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

[View Cutoff](#)

#### Primary Report Settings

E-value:

Sorted by Taxonomic Group:

Species Read-Across:

Primary Report  
 Full Report

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

### Level 2 Data - Primary

Search:

Data Version	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Filtered Taxonomic Group	Scientific Name	
2	<a href="#">NP_000116.2</a>	1058918	<a href="#">9606</a>	Mammalia	Primates	<a href="#">Homo sapiens</a>	
2	<a href="#">XP_005552209.1</a>	121048	<a href="#">9541</a>	Mammalia	Primates	<a href="#">Macaca fascicularis</a>	
2	<a href="#">ABY64721.1</a>	909	<a href="#">9534</a>	Mammalia	Primates	<a href="#">Chlorocebus aethiops</a>	
2	<a href="#">XP_011751932.1</a>	65086	<a href="#">9545</a>	Mammalia	Primates	<a href="#">Macaca nemestrina</a>	
2	<a href="#">NP_001158059.1</a>	77385	<a href="#">9555</a>	Mammalia	Primates	<a href="#">Papio anubis</a>	
2	<a href="#">XP_011922091.1</a>	66739	<a href="#">9531</a>	Mammalia	Primates	<a href="#">Cercopithecus atys</a>	
2	<a href="#">ABY64717.1</a>	2004	<a href="#">9593</a>	Mammalia	Primates	<a href="#">Gorilla gorilla</a>	
2	<a href="#">XP_014992596.1</a>	86552	<a href="#">9544</a>	Mammalia	Primates	<a href="#">Macaca mulatta</a>	
2	<a href="#">XP_003255939.1</a>	38938	<a href="#">61853</a>	Mammalia	Primates	<a href="#">Nomascus leucogenys</a>	no
2	<a href="#">XP_008005788.1</a>	62274	<a href="#">60711</a>	Mammalia	Primates	<a href="#">Chlorocebus sabaeus</a>	

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

The user may also choose to turn species read across off, by using the “Species Read-Across” drop-down and selecting “No” and clicking “Update Report.” When “No” is selected, the susceptibility predictions will only be “Y” in the table below if Percent Similarity is above the Cut-off or if the hit is identified as an Ortholog Candidate, yes or “Y.” Any hit below the cut-off will yield a susceptibility prediction of no or “N.”

Primary Report Settings

E-value:

Sorted by Taxonomic Group:

Species Read-Across:

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).

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	Susc Predi
<a href="#">orm_1</a>	132747	cd06949	NR_LBD_ER	238	238	238	1.078E-177	487.3	Y	42.8	100.0	
<a href="#">n</a>	132747	cd06949	NR_LBD_ER	238	238	238	1.078E-177	487.3	N	42.8	100.0	
<a href="#">_1</a>	132747	cd06949	NR_LBD_ER	238	238	238	1.078E-177	487.3	Y	42.8	100.0	
<a href="#">tor isoform X3</a>	132747	cd06949	NR_LBD_ER	238	237	238	6.557E-177	485.3	Y	42.8	99.6	
<a href="#">tor isoform X2</a>	132747	cd06949	NR_LBD_ER	238	237	238	6.557E-177	485.3	Y	42.8	99.6	
<a href="#">tor isoform X1</a>	132747	cd06949	NR_LBD_ER	238	237	238	6.557E-177	485.3	Y	42.8	99.6	
<a href="#">tor isoform X2</a>	132747	cd06949	NR_LBD_ER	238	237	238	6.557E-177	485.3	Y	42.8	99.6	
<a href="#">if</a>	132747	cd06949	NR_LBD_ER	238	237	238	6.557E-177	485.3	Y	42.8	99.6	
<a href="#">tor isoform X2</a>	132747	cd06949	NR_LBD_ER	238	237	238	6.557E-177	485.3	Y	42.8	99.6	
<a href="#">tor isoform X1</a>	132747	cd06949	NR_LBD_ER	238	237	238	6.557E-177	485.3	N	42.8	99.6	

(1 of 75)

**Note:** SeqAPASS v2.0 and newer parse the BLASTP query and hit accessions to identify all the species/accessions from identical proteins. Therefore, if a hit sequence represents multiple species, all species with the identical sequence will be found in the data tables for Level 2. To determine which sequence/species was identified from BLASTP as a hit and which sequence/species was parsed from the identical sequence, view the “Full Report” for Level, column “Identical Protein,” where “N” is indicative of the original hit sequence and “Y” is the parsed sequence.

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

Updated 3/06/18; 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 2<sup>nd</sup> local minimum, where the 2<sup>nd</sup> 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.

The screenshot displays the 'Level 2 Query Domain Information' interface. At the top, it provides query details: SeqAPASS ID: 581, Query Accession: NP\_000116.2, Ortholog Count: 281, Protein and Taxonomy Data: 01/04/2017, BLAST Version: 2.5.0, CDD Data: 04/25/2015, and Software Version: 2.0. The main content area is divided into two panels. The left panel, titled 'Susceptibility Cut-off', features a density plot of Percent Similarity (0-100) on the x-axis and Percent Frequency (0-20) on the y-axis. Three radio buttons are present: 'Default' (selected), 'Second Local Minimum', and 'User Defined'. A 'View Cutoff' button is located below the plot. The right panel, titled 'Primary Report Settings', includes an 'E-value' input field set to 10.0, a 'Sorted by Taxonomic Group' dropdown menu set to 'class', and a 'Species Read-Across' dropdown menu set to 'Yes'. 'Update Report' and 'Use Default Settings' buttons are at the bottom of this panel.

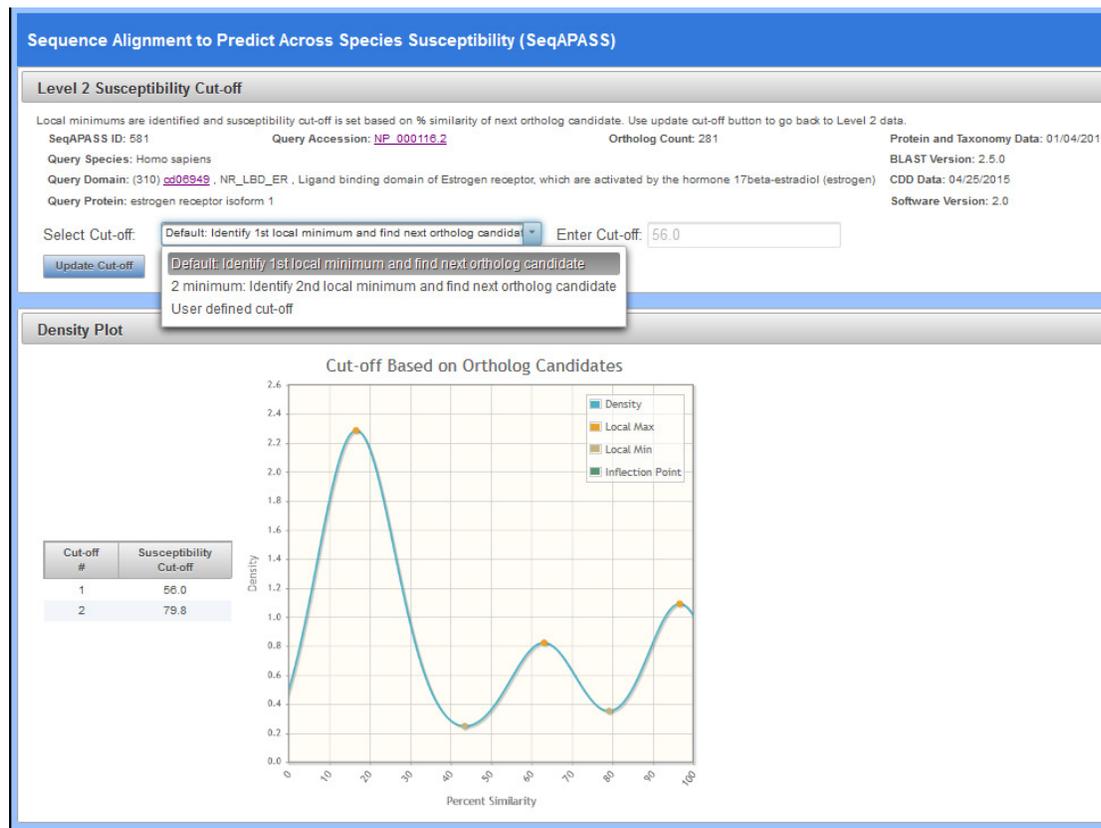
Upon clicking “View Cutoff” button, a new page is displayed with a drop-down that allows the user to set the susceptibility cut-off using the first local minimum and the identified ortholog candidate, the second local minimum and the identified ortholog candidate, or by the “User defined cut-off” (where the user selects the cutoff). To update the cut-off in the Level 2 data report and/or return to the Level 2 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 3/06/18; 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.



**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 3/06/18; 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 are navigation tabs: Home, Request SeqAPASS Run, SeqAPASS Run Status, View SeqAPASS Reports, and Settings. The user is logged in as LaLone.Carlie@epa.gov. Below the navigation is a header for 'SeqAPASS Reports' with 'Version 2.0' and 'Logged in as: LaLone.Carlie@epa.gov'. The main content area is titled 'Level 2 Query Domain Information' and includes the following details:

- SeqAPASS ID: 1570
- Query Accession: NP\_001317544.1
- Ortholog Count: 0
- Query Species: Homo sapiens
- Query Domain: (138) o08965, NR\_DBD\_Ppar, DNA-binding domain of peroxisome proliferator-activated receptors (PPAR) is composed of two C4-type zinc fingers
- Query Protein: peroxisome proliferator-activated receptor gamma isoform 3
- Protein and Taxonomy Data: 01/04/2017
- BLAST Version: 2.5.0
- CDD Data: 02/05/2016
- Software Version: 2.0

Below the information is a 'Susceptibility Cut-off' plot showing a density curve with a peak and a local minimum. The 'Primary Report Settings' panel includes:

- E-value: 10.0
- Sorted by Taxonomic Group: class
- Species Read-Across: Yes
- Buttons: Update Report, Use Default Settings

At the bottom, there are filter options: Partial Hit Protein Sequence, Percent Similarity > 100%, Susceptible = Y, Ortholog Count = 0, and Show Only Eukaryotes. The 'Level 2 Data - Primary' table is shown below, with a search bar and columns for Data Version, NCBI Accession, Protein Count, Species Tax ID, Taxonomic Group, Filtered Taxonomic Group, Scientific Name, Common Name, and Protein Name. The table contains several rows of data, with the first few rows highlighted in dark pink, indicating 0 orthologs detected.

Data Version	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Filtered Taxonomic Group	Scientific Name	Common Name	Protein Name
2	XP_011532145.1	1058915	9500	Mammalia	Mammalia	Homo sapiens	human	PREDICTED: peroxisome proliferator-activated receptor gamma isoform 3
2	XP_015496805.1	41101	9157	Aves	Aves	Falco major	Great Tit	PREDICTED: peroxisome proliferator-activated receptor gamma isoform 3
2	XP_017356934.1	53175	173749	Mammalia	Mammalia	Cebus capucinus imitator	white-faced sapsajou	PREDICTED: peroxisome proliferator-activated receptor gamma isoform 3
2	KF022300.1	28923	97421	Aves	Aves	Merops nubiola	damiana bee-eater	PREDICTED: peroxisome proliferator-activated receptor gamma isoform 3
2	YP_005658868.1	29026	75827	Echinodermata	Mollusca	Meropis caudata	caudata	PREDICTED: peroxisome proliferator-activated receptor gamma isoform 3

## 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.

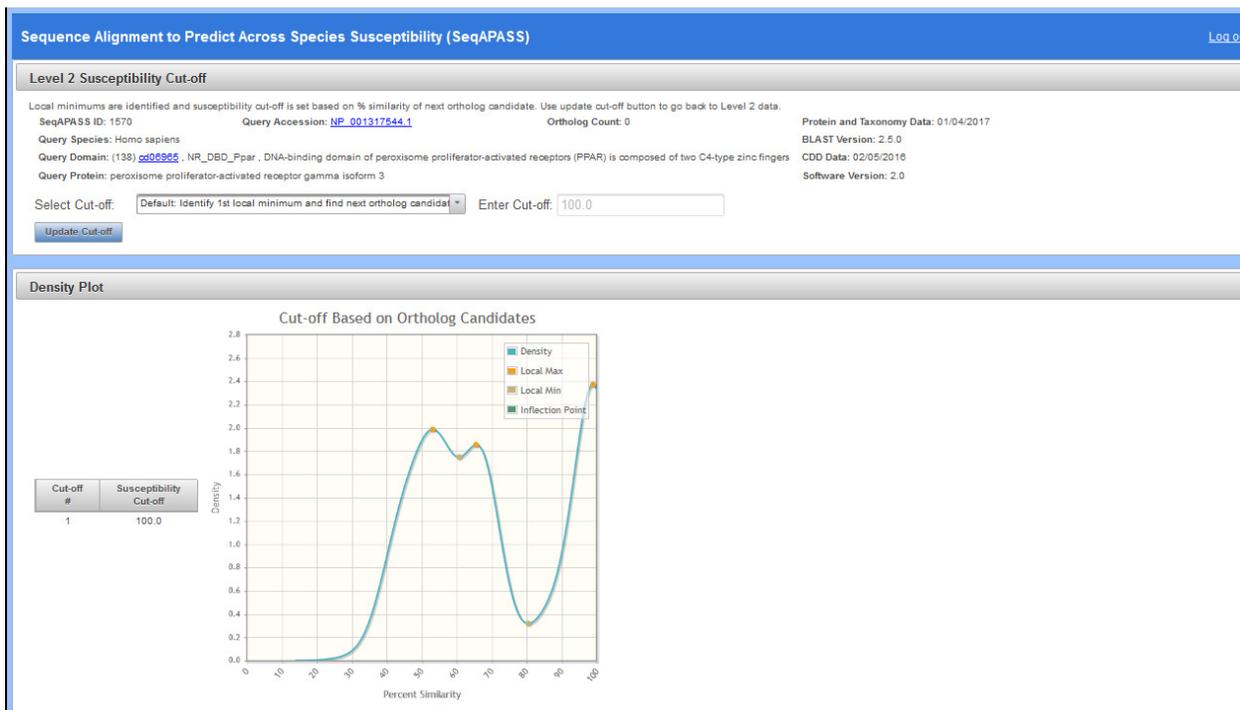
This screenshot is similar to the one above, showing the SeqAPASS web interface. The 'Level 2 Query Domain Information' section is identical. The 'Level 2 Data - Primary' table is also shown, with the same search bar and columns. The table contains several rows of data, with the first few rows highlighted in dark pink, indicating 0 orthologs detected.

Data Version	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Filtered Taxonomic Group	Scientific Name	Common Name	Protein Name
2	XP_011532145.1	1058915	9500	Mammalia	Mammalia	Homo sapiens	human	PREDICTED: peroxisome proliferator-activated receptor gamma isoform 3
2	XP_015496805.1	41101	9157	Aves	Aves	Falco major	Great Tit	PREDICTED: peroxisome proliferator-activated receptor gamma isoform 3
2	XP_017356934.1	53175	173749	Mammalia	Mammalia	Cebus capucinus imitator	white-faced sapsajou	PREDICTED: peroxisome proliferator-activated receptor gamma isoform 3
2	KF022300.1	28923	97421	Aves	Aves	Merops nubiola	damiana bee-eater	PREDICTED: peroxisome proliferator-activated receptor gamma isoform 3
2	YP_005658868.1	29026	75827	Echinodermata	Mollusca	Meropis caudata	caudata	PREDICTED: peroxisome proliferator-activated receptor gamma isoform 3

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

Updated 3/06/18; 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.



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 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Level 1 and Level 2: Data Visualization

From the Level 1 or Level 2 results page SeqAPASS users can access an interactive data visualization for both the “Primary Report” or “Full Report” by clicking on the “Visualize Data” button.

Example of Level 1 page:

SeqAPASS Reports Version TBD 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: 588 Query Accession: [NP\\_600116.2](#) Ortholog Count: 305 Protein and Taxonomy Data: 10/25/2017  
Query Species: Homo sapiens BLAST Version: 2.6.0  
Query Protein: estrogen receptor isoform 1 Software Version: TBD

#### Susceptibility Cut-off

Default  
 Second Local Minimum  
 User Defined

**View Cutoff**

This will open in a separate tab.

#### Primary Report Settings

E-value:

Sorted by Taxonomic Group:

Common Domains:

Species Read-Across:

**Update Report** **Use Default Settings**

#### Visualization

**Visualize Data** This will open in a separate tab.

#### Level 2

##### Level 2 Query Domain

[NCBI Conserved Domain Database](#)  
Functional Domains  
-Select Domain -  
**Request Domain Run**

##### View Level 2 Data

Choose Domain to View  
-Select Completed Domain -  
**View Level 2 Data**

#### Level 3

##### Level 3 Query Amino Acid Residue(s)

[NCBI Protein Database](#)  
Select Template Sequence

Additional Comparisons (optional)

[NCBI COBALT](#)  
Enter Level 3 Run Name

[NCBI Taxonomy Database](#)  
Choose Taxonomic Group(s)  
All Groups

Use table below to select sequences  
0 species selected  
**Request Residue Run**

##### View Level 3 Data

Choose Query to View  
-Select Level 3 Run Name -  
**View Level 3 Data**

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Example of Level 2 page:

The screenshot displays the SeqAPASS Level 2 Query Domain Information page. 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 TBD\_3'. The user is logged in as 'LaLone.Carlie@epa.gov'. The main content area is titled 'Level 2 Query Domain Information' and includes 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: 1042
- Query Accession: [NP\\_000116.2](#)
- Ortholog Count: 305
- Protein and Taxonomy Data: 10/25/2017
- Query Species: Homo sapiens
- BLAST Version: 2.6.0
- Query Domain: (310) [g08948](#): NR\_LBD\_ER . Ligand binding domain of Estrogen receptor, which are activated by the hormone 17beta-estradiol (estrogen)
- CDD Data: 12/08/2016
- Query Protein: estrogen receptor isoform 1
- Software Version: TBD\_3

The page contains three main sections:

- Susceptibility Cut-off:** A graph showing the susceptibility cut-off for the query domain. The y-axis is labeled 'Susceptibility' and the x-axis is labeled 'Percent Identity'. There are three radio buttons: 'Default' (selected), 'Second Local Minimum', and 'User Defined'. A 'View Cutoff' button is located below the graph.
- Primary Report Settings:** A form with the following fields:
  - E-value: 10.0
  - Sorted by Taxonomic Group: class
  - Species Read-Across: YesButtons for 'Update Report' and 'Use Default Settings' are located below the form.
- Visualization:** A section with a 'Visualize Data' button and the text 'This will open in a separate tab.'

The data visualization will then open in a new web browser tab, one for Level 1 and a different one for Level 2. The visualization will display for the report selected by the user on the Level 1 or Level 2 report page and be identified as “Level One Visualization – Primary Report” or “Level One Visualization – Full Report” and “Level Two Visualization – Primary Report” or “Level Two Visualization – Full Report.”

**Note:** One report type at a time, either “Primary Report” or “Full Report,” can be displayed in the visualization tab for Level 1 and Level 2. Therefore, if the user is viewing the “Level One Visualization – Primary Report” page and returns to the Level 1 results page and clicks the radio button for “Full Report,” the data visualization tab will update to “Level One Visualization – Full Report.”

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Level 1 Information Page

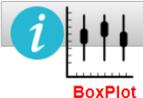
The initial page that opens upon clicking the “Visualize Data” button provides Level 1 query protein information, including SeqAPASS ID, query protein, query species, ortholog count, and query accession information. A link out to the NCBI protein database page corresponding to the queried accession is available by clicking the query accession. Information on the visualization is provided in the “Visualization Info” text box. To view the data visualization boxplots click the BoxPlot icon. The BoxPlot will then generate below the Visualization Info box.

### Level 1 Query Protein Information

SeqAPASS ID: 758      Query Accession: [NP\\_000116.2](#)  
Query Protein: estrogen receptor isoform 1  
Query Species: Homo sapiens  
Ortholog Count: 306

---

Select to Open Information or Data Visualization



BoxPlot

---

### Info

#### Visualization Info

The following data visualization is available for Level 1 and Level 2 data:

- **BoxPlot** - Boxplots depicting SeqAPASS data illustrating the percent similarity across species compared to the query species examining the primary amino acid sequences (Level 1 Visualization) or functional domain (Level 2 Visualization).
  - The open circle, o, represents the query species and closed circles, ●, represent the species with the highest percent similarity within the specified taxonomic group.
  - The top and bottom of each box represent the 75th and 25th percentiles, respectively. The top and bottom whiskers extend to 1.5 times the interquartile range.
  - The mean and median values for each taxonomic group are represented by horizontal thick and thin black lines on the box, respectively.
  - The dashed line indicates the cut-off for susceptibility predictions (based on ortholog analysis).

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Level 2 Information Page

The initial page that opens upon clicking the “Visualize Data” button provides Level 2 query protein information, including SeqAPASS ID, Query Species, Ortholog Count, Query Domain, and Query information. A link out to the NCBI protein database page corresponding to the queried accession and domain are available by clicking the Query Accession and Query Domain links, respectively. Information on the visualization is provided in the “Visualization Info” text box. To view the data visualization boxplots click the BoxPlot icon. The BoxPlot will then generate below the Visualization Info box.

### Level Two Visualization - Primary Report

---

**Level 1 Query Protein Information**

SeqAPASS ID: 1042 Query Accession: [NP\\_000119.2](#)  
Query Species: Homo sapiens  
Ortholog Count: 305  
Query Domain: (310) [cd05049](#), NR\_LBD\_ER, Ligand binding domain of Estrogen receptor, which are activated by the hormone 17beta-estradiol (estrogen)

---

Select to Open Information or Data Visualization



BoxPlot

---

**Info**

**Visualization Info**

The following data visualization is available for Level 1 and Level 2 data:

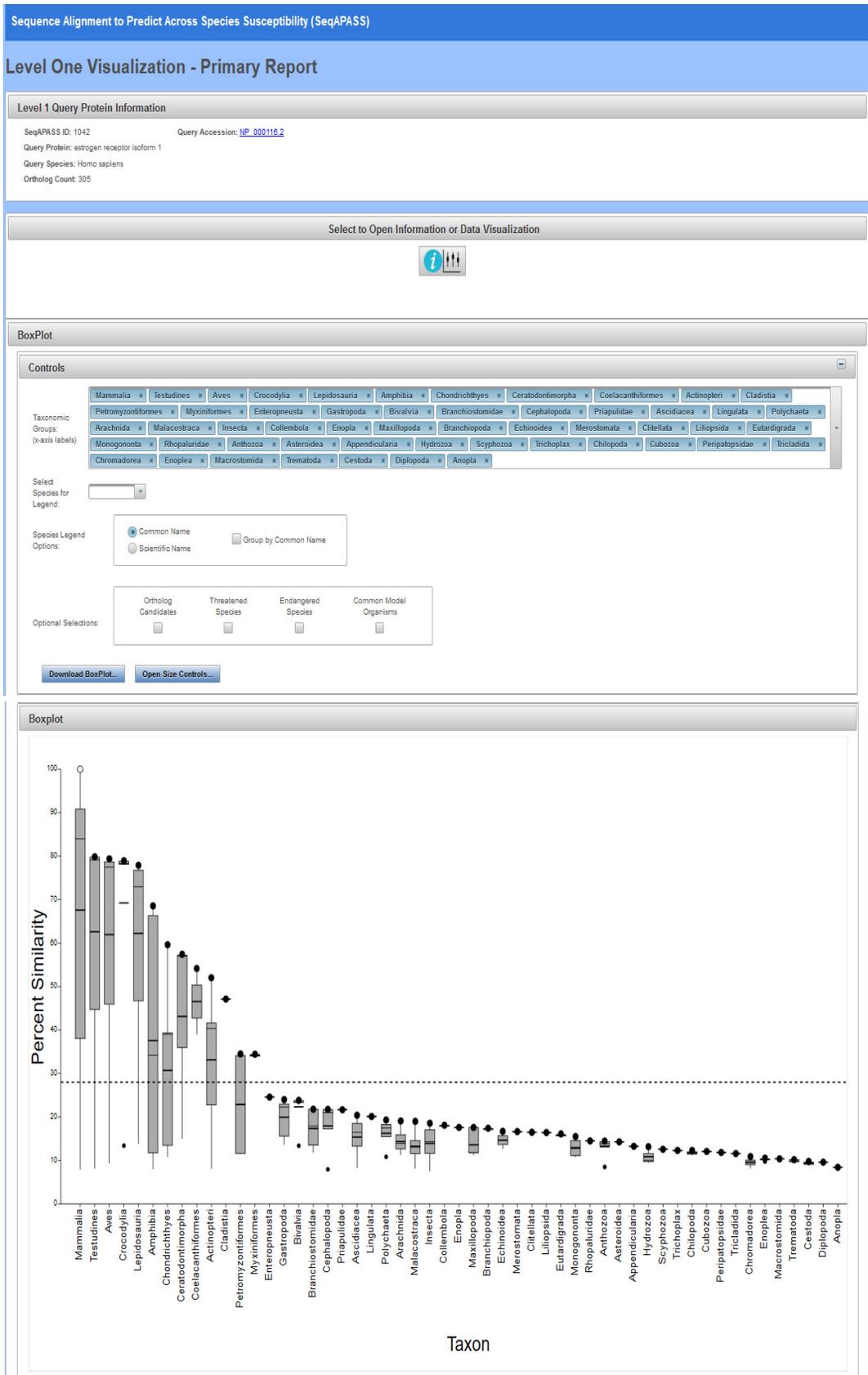
- **BoxPlot** - Boxplots depicting SeqAPASS data illustrating the percent similarity across species compared to the query species examining the primary amino acid sequences (Level 1 Visualization) or functional domain (Level 2 Visualization).
  - The open circle, o, represents the query species and closed circles, ●, represent the species with the highest percent similarity within the specified taxonomic group.
  - The top and bottom of each box represent the 75th and 25th percentiles, respectively. The top and bottom whiskers extend to 1.5 times the interquartile range.
  - The mean and median values for each taxonomic group are represented by horizontal thick and thin black lines on the box, respectively.
  - The dashed line indicates the cut-off for susceptibility predictions (based on ortholog analysis).

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Level 1 and 2 BoxPlot Page – Controls

Upon clicking the “BoxPlot” icon on either Level 1 or Level 2 Visualization Information pages, a box for the boxplot “Controls” and a box for the interactive boxplot will open, respectively.

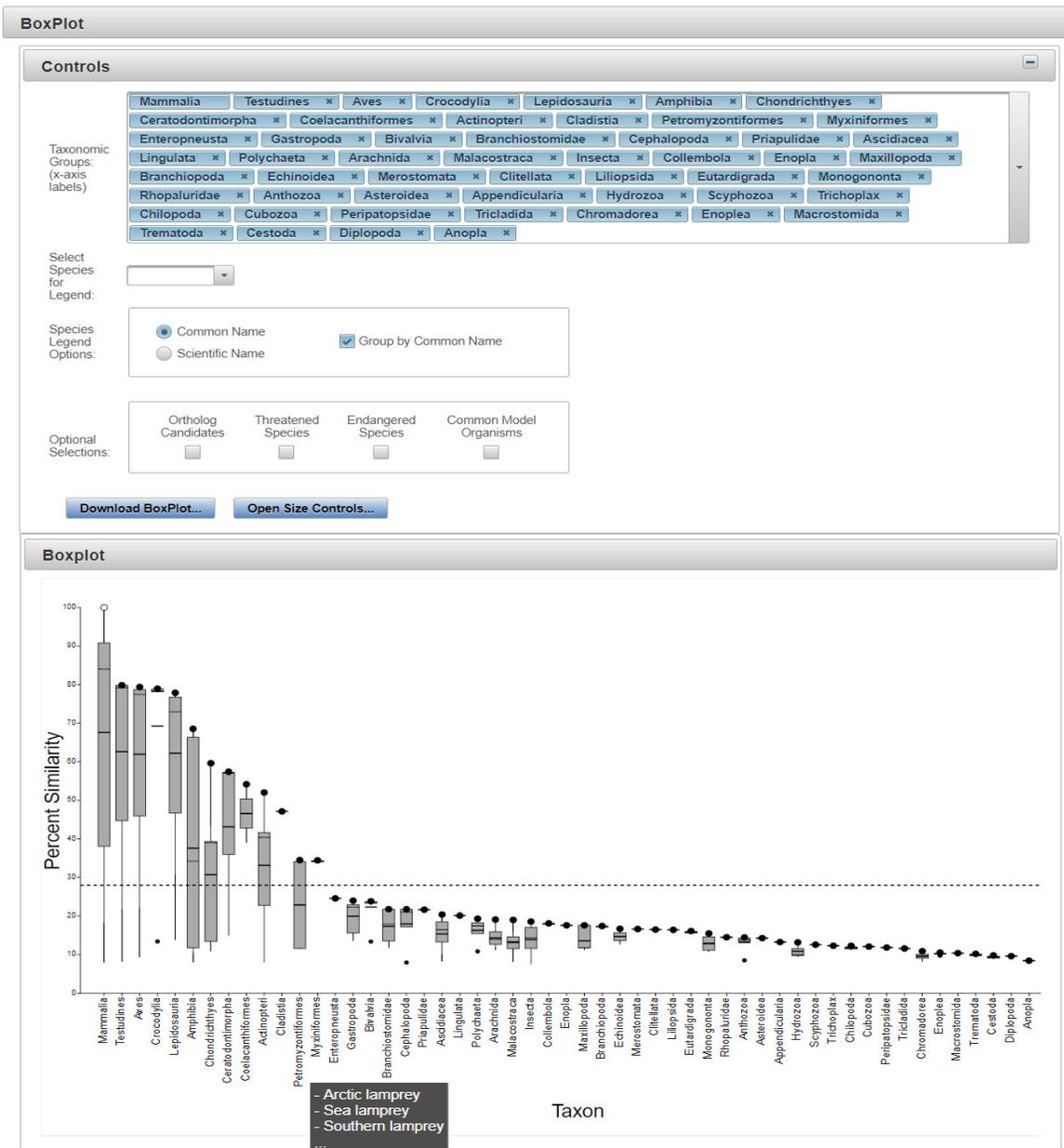


# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Manipulating Taxonomic Groups on x-axis

The controls allow the user to edit the taxonomic groups that are displayed on the x-axis by clicking on the “X” for the Taxonomic Group name (e.g., Aves). This action removes the selected group from the x-axis. To the right of the “Taxonomic Groups” controls box is a drop-down that allows the user to remove or add back taxonomic groups to the x-axis of the boxplot graphic, by deselecting or selecting checkboxes in the dropdown. Similarly, unwanted taxonomic groups may be removed directly from the BoxPlot by hovering the cursor over the taxonomic groups listed along the x-axis. The user will notice that the selection arrow changes to a black arrow with a red ‘x’ next to it; clicking the taxonomic group will then remove it from the BoxPlot and the “Taxonomic Groups” controls box. Additionally, that taxonomic group will have the checkbox deselected in the “Taxonomic Groups” controls box drop-down list.



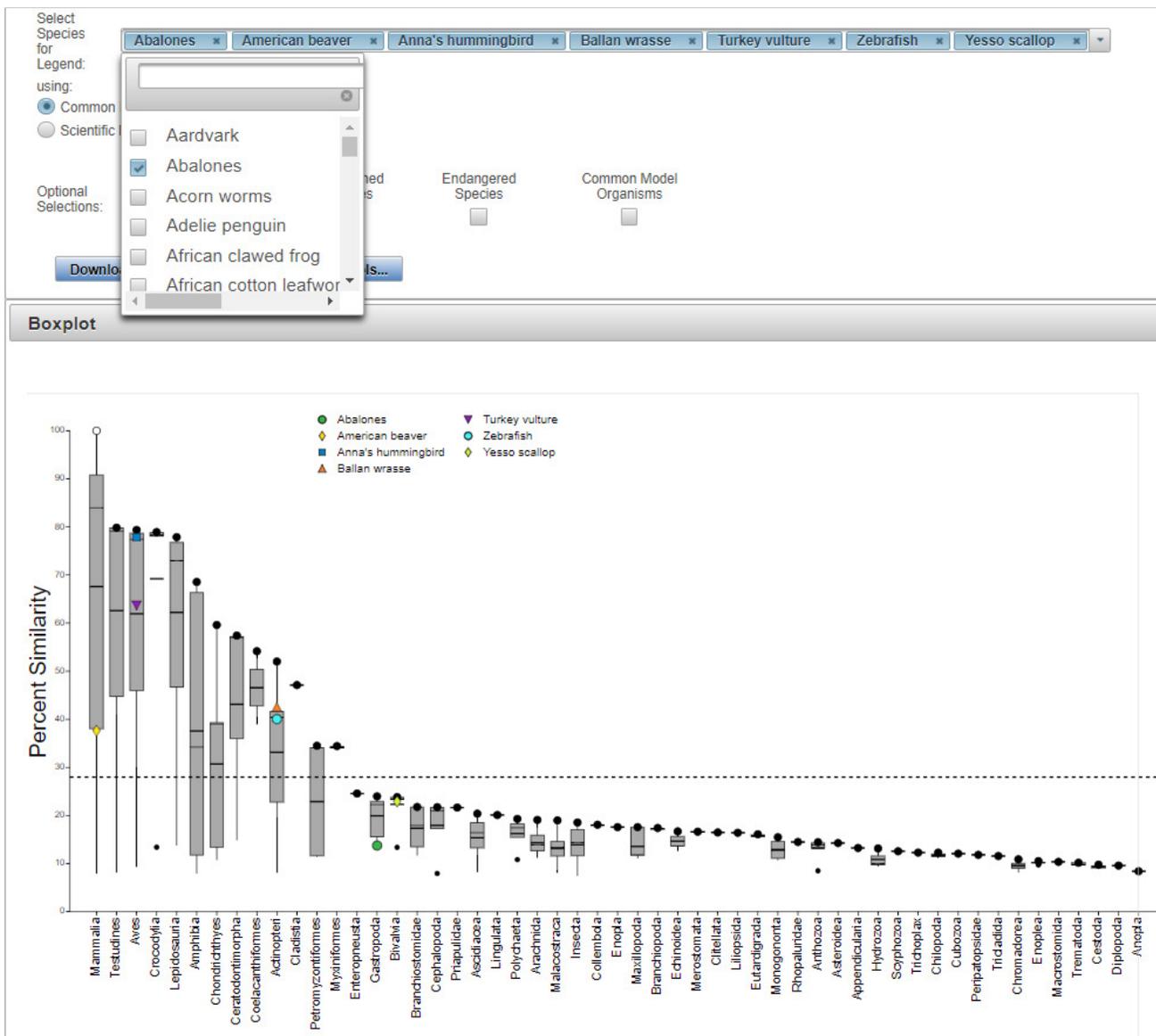
# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Customize BoxPlot Legend

The user may customize the BoxPlot by adding a legend that will pinpoint species of interest on the BoxPlot. Upon clicking the drop-down for “Select Species for Legend” in the controls box the user may search in the text box for specific species to display in the boxplot legend. Upon identifying a species from the drop-down menu and selecting the checkbox the species name will be placed in the boxplot legend and a corresponding data point will be produced on the graph. The default settings display the species common name both in the “Select Species for Legend” dropdown and on the boxplot. However, if the species scientific name is desired, the user can select the radio button for “Scientific Name” in the controls box for “Species Legend Options.” This action will change the drop-down menu and species in the legend to display the species scientific name.

**Note:** The database will take a brief moment to update the list upon changing between “Common Name” and “Scientific Name.”



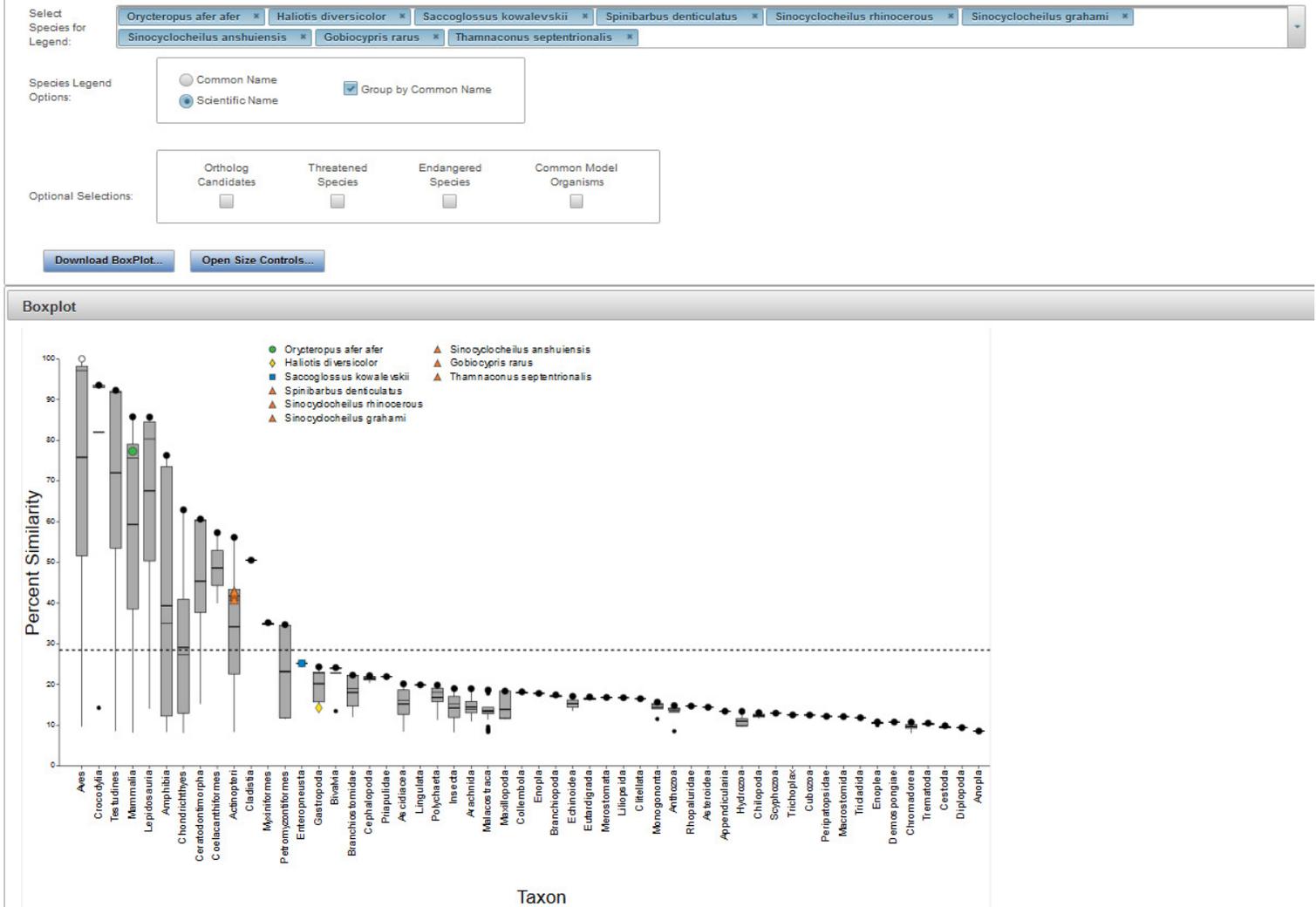
# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Change Species Display on Plot

Multiple scientific names can be represented by only one common name (e.g., Common name: Teleost fishes; corresponding scientific names: *Spinibarbus denticulatus*, *Sinocyclocheilus rhinoceros*, *Sinocyclocheilus grahami*, *Sinocyclocheilus anshuiensis*, *Gobiocypris rarus*, *Thamnaconus septentrionalis*). Therefore, if a species common name that represents multiple species was used to create the legend, and the user decides to instead select “Scientific Name,” by default the boxplot legend will change to display multiple scientific names that representing the individual common name and each scientific name will be represented by a unique color/shape point on the plot. However, if the user selects the checkbox “Group by Common Name” in the “Species Legend Options” control box, then the scientific names that are represented by one common name will all display the same color/shape point on the plot.

The user has the option of removing selected species from the legend either by removing them directly from the “Select Species for Legend” drop-down box or by hovering the mouse directly over the species name in the legend. The mouse will change to a black arrow with a red ‘x’ next to it. Clicking the name while this arrow is displayed will remove the species from the legend and from the control box.



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

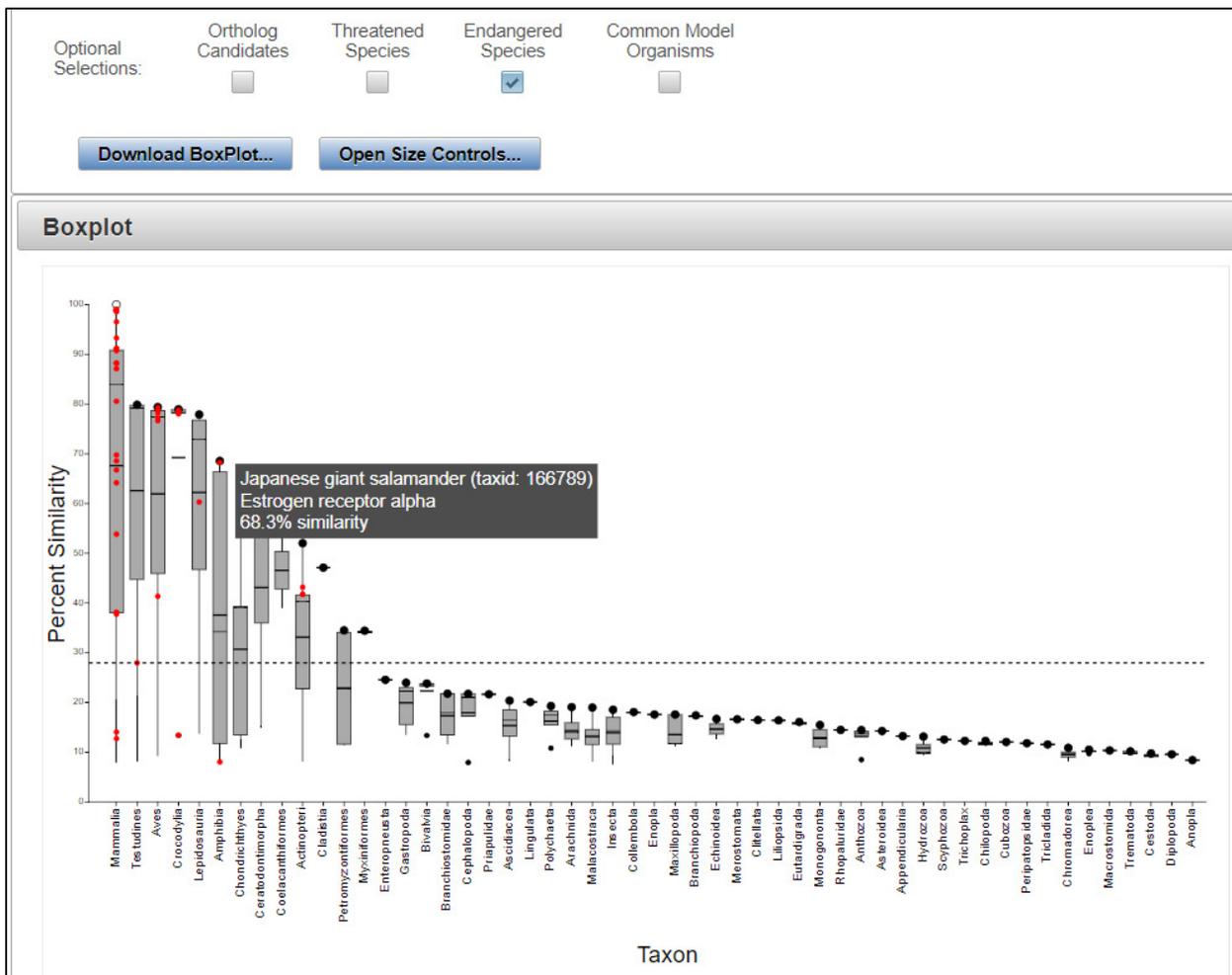
Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Customize the Legend to Display Species Groups of Interest

In the “Optional Selections” controls box, the user has the option of displaying “Ortholog Candidates,” “Threatened Species,” “Endangered Species,” or “Common Model Organisms.” Upon selecting one of the checkboxes, red data points corresponding to species will be displayed on the boxplot. By hovering the mouse over a single red point, a pop-up box will appear with the corresponding species name, taxonomic ID, query protein, and percent similarity.

**Note:** The user can select to display either species common name or scientific name in the hover over information box by selecting from the “Species Legend Options.”

If the user selects either “Threatened Species” or “Endangered Species,” clicking on an individual red dot will open a new web browser tab and link to the corresponding species page on the US Fish and Wildlife Service’s Environmental Conservation Online System (USFWS, ECOS; e.g., <https://ecos.fws.gov/ecp0/profile/speciesProfile?sId=1506>).



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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### *BoxPlot Controls Widget for Bar Width, Zoom and Pan*

By clicking the “Open Size Controls” button, a “BoxPlot Controls” widget opens that allows the user to adjust the size of the bars on the boxplot by increasing or decreasing the “Bar Width” using the up and down arrows. The minimum and maximum size for bars are 6 and 60, respectively. To reset the bar width on the boxplot to default size, click the “Reset” button to the right of the “Bar Width” adjustment box in the “BoxPlot Controls” box. The user can also Zoom and Pan the boxplot by toggling the on /off button under the “Zoom” heading. The user can then zoom in or out by clicking the up or down arrows or entering a number in the text box and clicking enter. To reset the zoom on the boxplot to default size, click the “Reset” button to the right of the “Zoom” adjustment box in the “BoxPlot Controls” widget. The pan option is available when the “Zoom and Pan” option is toggled to the “on” position, which allows the user to click on the boxplot and drag the plot around the screen to reposition. To reset all BoxPlot Controls to default settings click the “Reset All” button.

Note: Upon exiting out of the BoxPlot Controls widget, the Zoom and Pan options are automatically turned off.

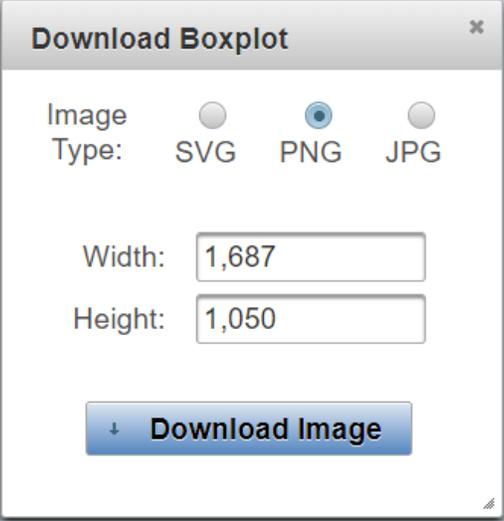


## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### Download BoxPlot Widget

To download the boxplot, click “Download BoxPlot” button in the controls box. A “Download Boxplot” Widget will pop up. It will be necessary to specify which type of file (SVG, PNG, or JPG,) to downloaded by clicking on the desired radio button for “Image Type.” The user may customize the resolution of the boxplot for PNG and JPG files prior to download by altering the “Width” and “Height” of the BoxPlot. To change “Width” or “Height,” enter the desired number in the text boxes. Click “Download Image” button to download the file. To close the “Download Boxplot” widget, click the “x” on the top right of the widget.



**Download Boxplot** ×

Image Type:  SVG  PNG  JPG

Width:

Height:

↓ **Download Image**

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### *Hover-over Features in the BoxPlot*

By hovering over a taxonomic group name on the x-axis of the box plot, an information box will pop-up listing the top three species in order by highest percent similarity. If only one or two species are represented in the taxonomic group, then only those species will be displayed. Hovering the mouse over any of the species in the legend will generate a pop-up box with the corresponding species name, taxonomic ID, query protein, and percent similarity. The susceptibility cut-off is displayed in a pop-up text box upon hovering over the dashed horizontal cut-off line.

### *Summary Table for Species in a Specific Taxonomic Group*

By clicking on a box representing a taxonomic group in the boxplot a table will pop-up providing summary information for that particular group. The table header will provide summary statistics (i.e., mean and median percent similarity), including the Taxonomic Group name, number of species represented in the box, the overall susceptibility prediction for the selected taxonomic group. Data table includes protein and species information along with metrics for evaluated protein similarity and predicting susceptibility. Also included in the table are columns indicating if a species belongs to a certain group of interest (e.g., Threatened Species; Endangered Species, Model Organism). Table can be downloaded by clicking on the icon for excel or csv file.

### *Interactive Visualization with Level 1 Data Page and Level 2 Data Page*

The data visualization is programmed to update with changes made to the Level 1 Data page and Level 1 Data page, respectively. Therefore, if the user updates the Susceptibility Cut-off (See user guide section **Susceptibility Cutoff Box for Level 1** and **Susceptibility Cutoff Box for Level 2**) to the “Second Local Minimum” or “User Defined Cut-off,” the previously opened data visualization boxplot tab will update the cut-off accordingly. Similarly, the user modifies the Primary Report Settings (See user guide section **Level 1: Primary Report Settings** and **Level 2: Primary Report Settings**), the data visualization will update accordingly.

Note: If the user updates the “Primary Report Settings” for “Sorted by Taxonomic Group” the boxplot will update to display the new taxonomic group selection that is present in the “Filtered Taxonomic Group” column in the data table. The user should be aware that manipulating the “Sorted by Taxonomic Group” to a different level in the taxonomic lineage (e.g., from class to order; from class to genus) adds a larger number of taxonomic groups to the x-axis. Therefore, the plot may require greater user manipulation using the BoxPlot Controls to view the data.

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

Updated 3/06/18; 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 choice of template sequence and residues that are selected to align will be derived from the published literature in most cases. Publications evaluating homology models, protein crystal structures, 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 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 the tabs, it says "SeqAPASS Reports" and "Version 2.0". On the right, it says "Logged in as: LaLone.Carlie@epa.gov".

The main content area is titled "Level 1 Query Protein Information". It contains 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: 581
- Query Accession: [NP\\_000116.2](#)
- Ortholog Count: 281
- Protein and Taxonomy Data: 01/04/2017
- Query Species: Homo sapiens
- BLAST Version: 2.5.0
- Query Protein: estrogen receptor isoform 1
- Software Version: 2.0

The interface is divided into three main sections:

- Susceptibility Cut-off:** Includes a graph showing susceptibility vs. amino acids, radio buttons for Default, Second Local Minimum, and User Defined, and a "View Cutoff" button.
- Level 2:** Includes a "Level 2 Query Domain" section with a link to "NCBI Conserved Domain Database", a "Functional Domains" dropdown menu, a "Request Domain Run" button, and a "View Level 2 Data" button.
- Level 3:** Includes a "Level 3 Query Amino Acid Residue(s)" section with a link to "NCBI Protein Database", a "Select Template Sequence" text box, an "Additional Comparisons (optional)" text box, a link to "NCBI COBALT", an "Enter Level 3 Run Name" text box, a link to "NCBI Taxonomy Database", a "Choose Taxonomic Group(s)" dropdown menu, a "Use table below to select sequences" section with "0 species selected", a "Request Residue Run" button, and a "View Level 3 Data" button.

At the bottom of the Level 3 section, there is a "Choose Query to View" dropdown menu with the option "-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., < >gil62821794|ref|NP\_000116.2| estrogen receptor isoform 1 [Homo sapiens]

```
MTMTLHTKASGMALLHQIQGNELEPLNRPQLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNA
AAAANA
QVYGGQTGLPYGPGSEAAAFSGNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLENE
PSGYT
VREAGPPAFYRPNSDNRRQGGRRERLASTNDKGSMAKESAKETRYCAVCNDYASGYHYGVWSC
EGCKAFFK
RSIQGHNDYMCATNQCTIDKNRRKSCQACRLRKCYEVGMMKGGIRKDRRGGRMLKHKRQRD
DGEGRGEV
GSAGDMRAANLWPSPLMIKRSKKNLSLALSLTADQMVSALLDAEPPILYSEYDPTRPFSEASMMG
LLTNLA
DRELVHMINWAKRVPGFVDLTLHDQV).
```

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

Updated 3/06/18; 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. A link out to the NCBI Protein Database is available for the user and found above the template entry text box.

The screenshot shows two main panels: Level 2 and Level 3. The Level 2 panel is titled "Level 2 Query Domain" and includes a link to "NCBI Conserved Domain Database", a "Functional Domains" dropdown menu, and a "Request Domain Run" button. The Level 3 panel is titled "Level 3 Query Amino Acid Residue(s)" and includes a link to "NCBI Protein Database", a "Select Template Sequence" text box, an "Additional Comparisons (optional)" text box, a link to "NCBI COBALT", an "Enter Level 3 Run Name" text box, a link to "NCBI Taxonomy Database", a "Choose Taxonomic Group(s)" dropdown menu, a "Request Residue Run" button, and a "View Level 3 Data" button. A pop-up window is overlaid on the Level 2 panel, containing the following text: "-Enter NCBI Protein Accession OR FASTA Sequence- Examples: NP\_000116.2 OR >Sequence description in first line MTMTLHTKASGMALLHQIQGNELEPLNRPQLKIPLERPLGEVYLDSSKPAVY".

Additional sequences can (this is an optional field the user can choose to fill in) also be incorporated into the Level 3 alignment using the “Additional Comparisons (optional)” text box. Upon clicking on the “Additional Comparisons (optional)” text box, a pop-up message will appear to provide examples for the proper format of Accessions or FASTA files to be entered.

**Note:** In the “Additional Comparisons (optional)” text box, zero or more NCBI Protein Accession must be entered prior to FASTA sequence(s) if they are to be included in the Level 3 alignment.

The screenshot shows the same Level 2 and Level 3 panels as above. A pop-up window is overlaid on the Level 3 panel, containing the following text: "-Enter 0 or more NCBI Protein Accession(s) followed by 0 or more FASTA Sequence(s)- Examples: NP\_000116.2 1JLY\_A >Sequence description of first FASTA MTMTLHTKASGMALLHQIQGNELEPLNRPQLKIPLERPLGEVYLDSSKPAVY >Sequence description of second FASTA XAGLPVIMCLKSNHQQYLRYSQSDNIQQYGLLQFSADKILDPLAQFEVPSKTYDGLV".

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Below the text box where the user can choose to add additional sequences for comparison, 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 algorithms utilize 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.

### Level 3

**Level 3 Query Amino Acid Residue(s)**

[NCBI Protein Database](#)  
Select Template Sequence

Additional Comparisons (optional)

[NCBI COBALT](#)  
Enter Level 3 Run Name

[NCBI Taxonomy Database](#)  
Choose Taxonomic Group(s)

Use table below to select sequences  
0 species selected

**View Level 3 Data**

Choose Query to View

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

Updated 3/06/18; 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.

**Note:** The “Choose Taxonomic Group(s):” drop-down will display the level of the taxonomic hierarchy being displayed in the “Filtered Taxonomic Group” column of the Level 1 Data table. For example, if the user changes the default option from “class” to “order,” then “order will be displayed in the dropdown.

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

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.

**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.

The screenshot displays the SeqAPASS web interface. At the top, there are three main configuration panels: "Susceptibility Cut-off", "Level 2", and "Level 3".

- Susceptibility Cut-off:** Includes a line graph showing a peak at approximately 0.01. Below the graph are radio buttons for "Default", "Second Local Minimum", and "User Defined". A "View Cutoff" button is present. Under "Primary Report Settings", there are input fields for "E-value" (0.01), "Sorted by Taxonomic Group" (class), "Common Domains" (1), and "Species Read-Across" (Yes). "Update Report" and "Use Default Settings" buttons are at the bottom.
- Level 2:** Titled "Level 2 Query Domain", it features a "Functional Domains" dropdown menu set to "-Select Domain -" and a "Request Domain Run" button. Below, "View Level 2 Data" is shown with a "Choose Domain to View" dropdown set to "-Select Completed Domain -" and a "View Level 2 Data" button.
- Level 3:** Titled "Level 3 Query Amino Acid Residue(s)", it includes a "Select Template Sequence" input field with "NP\_00018.2". Below is an "Additional Comparisons (optional)" field. The "NCBI COBALT" section has an "Enter Level 3 Run Name" field set to "Actinopteri". The "NCBI Taxonomy Database" section has a "Choose Taxonomic Group(s)" dropdown set to "Actinopteri". Below this, it says "Use table below to select sequences" and "4 species selected". A "Request Residue Run" button is present. At the bottom, "View Level 3 Data" is shown with a "Choose Query to View" dropdown set to "-Select Level 3 Run Name -" and a "View Level 3 Data" button.

Below the configuration panels is a search filter section with a search box containing "Actinopteri". To the left of the search box are several checkboxes: "Primary Report", "Full Report", "Partial Hit Protein Sequence" (checked), "Percent Similarity > 100%" (checked), "Susceptible = Y, Chromolog Count = 1" (checked), and "Show Only Eukaryotes" (checked).

The "Level 1 Data - Primary" table is displayed below the search filter. The search box above the table contains "Actinopteri". The table has columns for "Data Version", "NCBI Accession", "Protein Count", "Species Tax ID", "Taxonomic Group", "Filtered Taxonomic Group", "Scientific Name", "Common Name", and "Protein Name". Three rows are visible, each with a checked checkbox in the first column:

Data Version	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Filtered Taxonomic Group	Scientific Name	Common Name	Protein Name	
<input checked="" type="checkbox"/>	2	BA982853.1	115	512342	Actinopteri	Actinopteri	<i>Atractosteus tropicus</i>	tropical gar	astrocan_receptor_alpha
<input checked="" type="checkbox"/>	2	XP_005625909.1	58003	7918	Actinopteri	Actinopteri	<i>Lepidosteus oculatus</i>	spotted gar	PREDICTED_astrocan_receptor
<input checked="" type="checkbox"/>	2	BA982850.1	268	111304	Actinopteri	Actinopteri	<i>Acipenser schranckii</i>	Amur sturgeon	astrocan_receptor_alpha.1

(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.”

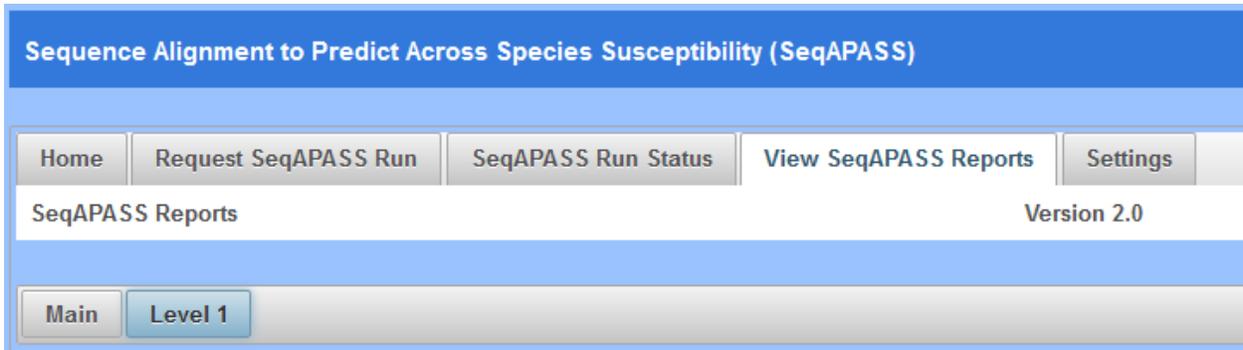
## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

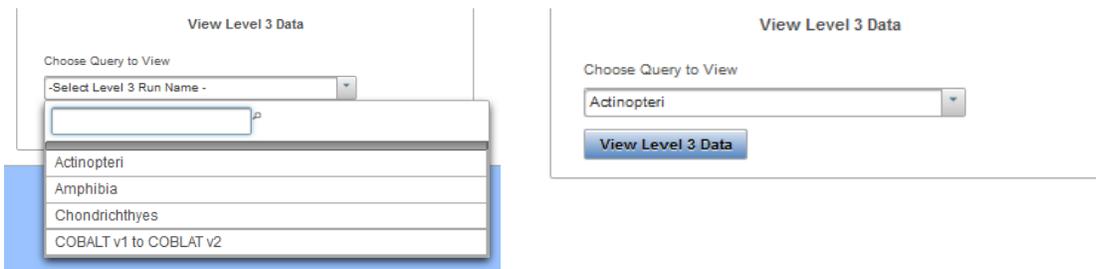
When the user has selected all sequences they would like to align, then click the “Request Residue Run” button. Upon successful submission of a Level 3 query the user will see the following pop-up message. If submission is unsuccessful, a message will appear describing the reason for the unsuccessful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



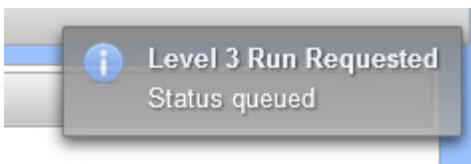
To update the View Level 3 Data, “Choose Query to View” drop-down menu with the completed Level 3 alignments, the user can click on the top left “Level 1” button.



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 3/06/18; 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 and executables were downloaded and incorporated into SeqAPASS), 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).

**Level 3 Template Protein Information**

Individual amino acid residue(s) aligned with template sequence. Use the main button to go back to the SeqAPASS Reports list.

SeqAPASS ID: 1042    Query Accession: [NP\\_000116.2](#)    Ortholog Count: 305    Protein and Taxonomy Data: 10/25/2017

Level 3 Run Name: Actinopteri    BLAST Version: 2.6.0

Template Species: Homo sapiens    Cobalt Data: 07/09/2010

Template Protein: [NP\_000116.2] estrogen receptor isoform 1    Cobalt Version: 2.1.0

Query Residues: No Residues Selected    Software Version: TBD\_3

[Show Amino Acid Info...](#)

**Select Amino Acid Residues**

1M  
2T  
3M  
4T  
5L  
6H  
7T  
8K

Enter Amino Acid Residue Positions

[Copy to Residue List](#)

[Update Report](#)

Primary Report  
 Full Report

**Level 3 Data - Primary**

Search:

Data Version	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Scientific Name	Common Name	Protein Name
5	NP_000116.2	1151619	<a href="#">9506</a>	Mammalia	<a href="#">Homo sapiens</a>	Human	estrogen receptor isoform 1
5	AAU87498.1	449	<a href="#">90888</a>	Actinopteri	<a href="#">Pimephales promelas</a>	Fathead minnow	estrogen receptor alpha
5	XP_014061037.1	112155	<a href="#">8030</a>	Actinopteri	<a href="#">Salmo salar</a>	Atlantic salmon	PREDICTED: estrogen receptor isoform X2
5	XP_020570162.1	42591	<a href="#">8090</a>	Actinopteri	<a href="#">Oryzias latipes</a>	Japanese medaka	estrogen receptor
5	XP_021454037.1	124278	<a href="#">8022</a>	Actinopteri	<a href="#">Oncorhynchus mykiss</a>	Rainbow trout	estrogen receptor isoform X3
5	AAI62486.1	87287	<a href="#">7955</a>	Actinopteri	<a href="#">Danio rerio</a>	Zebrafish	Estrogen receptor 1

(1 of 1)   Download Table:

The user can view the “Level 3” data page, which includes the Data Version, 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 template 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).

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

For additional information on Amino Acid Residues, including definition of the acronym, the amino acid residue name, the classification for the amino acid side chain and the size of the amino acid residue based on molecular weight, the user can click the “Show Amino Acid Info...” button. A pop-up table, “Amino Acid info,” will be displayed providing this information.

Main Level 1 Level 3

### Level 3 Template Protein Information

Individual amino acid residue(s) aligned with template sequence. Use the main button to go back to the SeqAPASS Reports list.

SeqAPASS ID: 1042      Query Accession: [NP\\_000116.2](#)      Ortholog Count: 305      Protein and Taxonomy Data: 10/25/2017

Level 3 Run Name: Actinopteri      BLAST Version: 2.6.0

Template Species: Homo sapiens      Cobalt Data: 07/09/2010

Template Protein: [NP\_000116.2] estrogen receptor isoform 1      Cobalt Version: 2.1.0

Query Residues: 351D, 353E, 362K, 364V, 394R, 524H      Software Version: TBD\_3

[Show Amino Acid Info...](#)

#### Select Amino Acid Residues

1M  
2T  
3M  
4T  
5L  
6H  
7T  
8K

→
←

[Update Report](#)

Primary Report  
 Full Report

### Level 3 Data - Primary

Data Version	NCBI Accession	Protein Count	Species Tax ID	Species	Protein Name
5	NP_000116.2	1151619	<a href="#">9606</a>	Actinopteri	estrogen receptor isoform 1
5	AAU87498.1	449	<a href="#">90988</a>	<a href="#">Primephates promelas</a>	estrogen receptor alpha
5	XP_014061037.1	112155	<a href="#">8030</a>	Actinopteri	PREDICTED: estrogen receptor isoform X2
5	XP_020570152.1	42591	<a href="#">8090</a>	Actinopteri	estrogen receptor
5	XP_021454037.1	124276	<a href="#">8022</a>	Actinopteri	estrogen receptor isoform X3
5	AAI62466.1	87287	<a href="#">7955</a>	Actinopteri	Estrogen receptor 1

#### Amino Acid info

ID	Name	Side Chain	Size
A	Alanine	Aliphatic	89.094
C	Cysteine	Sulfur-Containing	121.154
D	Aspartic Acid	Acidic	133.104
E	Glutamic Acid	Acidic	147.131
F	Phenylalanine	Aromatic	165.192
G	Glycine	Aliphatic	75.067
H	Histidine	Basic	155.156
I	Isoleucine	Aliphatic	131.175
K	Lysine	Basic	146.189
L	Leucine	Aliphatic	131.175
M	Methionine	Sulfur-Containing	149.208
N	Asparagine	Amidic	132.119
P	Proline	Aliphatic	115.132
Q	Glutamine	Amidic	146.146
R	Arginine	Basic	174.203
S	Serine	Hydroxylic	105.093
T	Threonine	Hydroxylic	119.119
V	Valine	Aliphatic	117.148
W	Tryptophan	Aromatic	204.228
Y	Tyrosine	Aromatic	181.191

(1 of 1)      [Download Table:](#)

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

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

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 1<sup>st</sup> 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.

Select Amino Acid Residues

1M	3M
2T	219Y
4T	267H
5L	268K
6H	272D
7T	594T
8K	
9A	

Update Report

Alternatively, the user can enter the amino acid positions in the “Enter Amino Acid Residue Positions” text box (e.g., 351,353,362) and click the “Copy to Residue List” button.

Enter Amino Acid Residue Positions

351,353,362,364,394,524

Copy to Residue List

Upon clicking “Copy to Residue List” the “Select Amino Acid Residues” shuttle box is populated with the position and residues typed. The user can then click the update Report button to produce Level 3 results in the table below.

Select Amino Acid Residues

1M	351D
2T	353E
3M	362K
4T	364V
5L	394R
6H	524H
7T	
8K	

Update Report

Enter Amino Acid Residue Positions

351,353,362,364,394,524

Copy to Residue List

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 1<sup>st</sup> position in the template sequence to the last position in the template sequence.

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### Level 3 Data – Primary Report

The default report is the “Primary Report” and can be recognized as such because the radio button for “Primary Report” above the “Level 3 Data” table is selected.

The “Primary Report” columns for the alignment will be titled “Similar Susceptibility as Template” (“Y” or “N” for yes or no, respectively), followed by Position 1, Amino Acid 1, Total Match 1, Position 2 Amino Acid 2, Total Match 2, Position 3, Amino Acid 3, Total 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 and aligning with the template sequence. The Total Match X, describes whether the amino acid residue matches the template based on side-chain classification and molecular weight, “Y,” for yes, or “N,” for not a match to the template. 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.

	Common Name	Protein Name	Analysis Completed	Similar Susceptibility as Template	Position 1	Amino Acid	Total Match	Position 2	Amino Acid	Total Match
<a href="#">H</a>	Human	estrogen receptor isoform 1	2018 02 27 14:37:54	Y	351	D	Y	353	E	Y
<a href="#">ncllas</a>	Fathead minnow	estrogen receptor alpha	2018 02 27 14:37:54	Y	320	D	Y	322	E	Y
<a href="#">:</a>	Atlantic salmon	PREDICTED: estrogen receptor isoform X2	2018 02 27 14:37:54	Y	316	D	Y	318	E	Y
<a href="#">H</a>	Japanese medaka	estrogen receptor	2018 02 27 14:37:54	Y	355	D	Y	357	E	Y
<a href="#">lykias</a>	Rainbow trout	estrogen receptor isoform X3	2018 02 27 14:37:54	Y	319	D	Y	321	E	Y
	Zebrafish	Estrogen receptor 1	2018 02 27 14:37:54	Y	319	D	Y	321	E	Y

**Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide**  
 Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

*Level 3 Data – Full Report*

The user may choose to view the Full Report for Level 3 data by selecting the radio button above the “Level 3 Data” table for “Full Report.” The table below will automatically update to display all of the alignment details.

The “Full Report” columns for the alignment will be titled “Similar Susceptibility as Template” (“Y” or “N” for yes or no), followed by Position 1, Amino Acid 1, Direct Match 1, Side Chain 1, MW1, MW Match 1, Total Match 1, Position 2, Amino Acid 2, Direct Match 2, Side Chain 2, MW2, MW Match 2, Total Match 2..... 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 Total Match X, describes whether the amino acid residue matches the template based on side-chain classification and molecular weight, “Y,” for yes, or “N,” for not a match to the template. 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.

Completed	Similar Susceptibility as Template	Position 1	Amino Acid 1	Direct Match 1	Side Chain 1	Side Chain Match 1	MW 1	MW Match 1	Total Match 1	Position 2	Amino Acid 2	Direct M
27 14:37:54	Y	351	D	Y	Acidic	Y	133.104	Y	Y	353	E	Y
27 14:37:54	Y	320	D	Y	Acidic	Y	133.104	Y	Y	322	E	Y
27 14:37:54	Y	316	D	Y	Acidic	Y	133.104	Y	Y	318	E	Y
27 14:37:54	Y	355	D	Y	Acidic	Y	133.104	Y	Y	357	E	Y
27 14:37:54	Y	319	D	Y	Acidic	Y	133.104	Y	Y	321	E	Y
27 14:37:54	Y	319	D	Y	Acidic	Y	133.104	Y	Y	321	E	Y

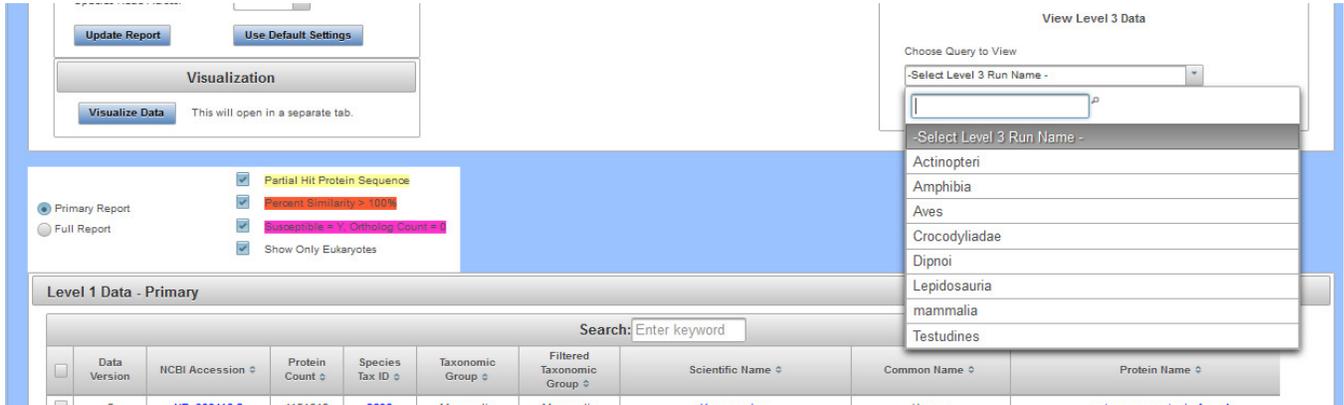
The “Direct Match X” column describes whether the hit amino acid is an exact match to the template amino acid, providing a “Y” or “N” for yes or no, respectively. The “Side Chain X” column indicates the side chain classification for the amino acid residue (click on “Show Amino Acid Info...for more information on classifications). The “Side Chain Match X” column indicates whether the hit side chain has the same classification as the template amino acid, providing a “Y” or “N” for yes or no, respectively. The “MW X” column indicates the molecular weight (g/mol) of the amino acid residue and the “MW Match X” column indicates whether the hit molecular weight has a difference in molecular weight greater than or equal to 30 g/mol compared to the template amino acid, providing a “Y” or “N” for yes or no, respectively. For the “Total Match X” to be “Y,” both “Side Chain Match X” and “MW Match X” should be either “Y” and Y” or one “Y” and one “N,” respectively. Only if both “Side Chain Match X” and “MW Match X” are “N” and “N,” then the “Total Match X” is “N” for no. Ultimately, the Total Match 1, 2, 3, 4... are used to inform the “Similar Susceptibility as Template” column. If there is one or more “N” for Total Match comparing any amino acid residue to the template across a row for a given species, then the “Similar Susceptibility as Template” is “N” for no, indicating that the hit species is predicted NOT to have the same susceptibility prediction as the template sequence. However, if all “Total Match X” are “Y” for yes, then the “Similar Susceptibility as Template” is “Y” indicating that the hit species is predicted to have the same susceptibility prediction as the template sequence.

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

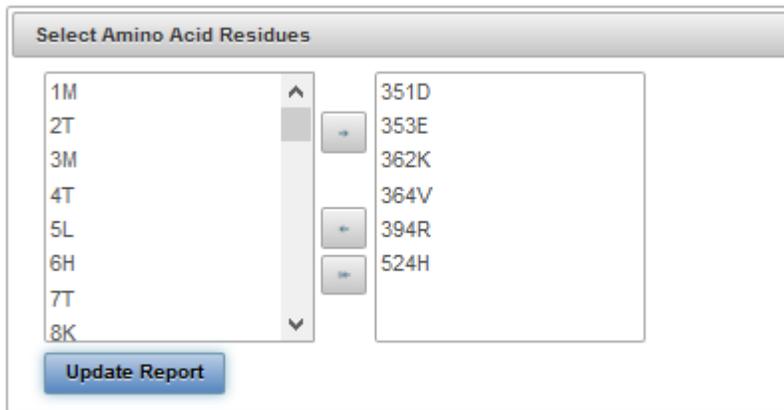
Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Multiple Level 3 Runs Requiring the Same Amino Acid Residue Comparisons

Typically, Level 3 individual amino acid residue alignments are submitted repetitively, comparing species from one taxonomic group at a time to the template amino acid residue(s).



Therefore, to increase efficiency in submitting the same alignments in Level 3 over and over again, the user can take advantage of the “Copy to Residue List” button. For the first alignment of amino acid residues, the user would select the amino acid residues to align and click the “Update Report” button.



By clicking “Update Report” the residues that were selected will be copied into the “Enter Amino Acid Residue Positions” text box. When the user selects a new Level 3 Run Name from the same Level 1 query accession) to view by using the “View Level 3 Data” dropdown and clicking the “View Level 3 Data” button on the Level 1 Query Protein Information page, the “Enter Amino Acid Residue Positions” text box will be populated with the amino acid residues selected from the previous run.



The user can keep, add, or delete, residue positions in this box and click “Copy to Residue List” button. The amino acid residues will then be moved to the “Select Amino Acid Residues Shuttle” and the user can then click “Update Report” to view the data in the table below.

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

Updated 3/06/18; 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 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

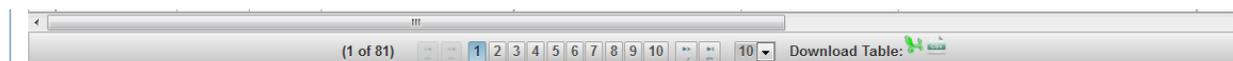
## Search, View, and Download Data Tables

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.

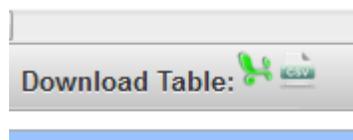
Search using text box on top of tables:



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 .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 interface. On the left, there is a graph with a red line and a 'View Cutoff' button. In the center, there are report options: 'Primary Report' (selected), 'Full Report', 'Partial Hit Protein Sequence', 'Percent Similarity > 10%', 'Susceptible = Y Ortholog Count = 0', and 'Show Only Eukaryotes'. To the right, there are settings for 'E-value' (10.0), 'Sorted by Taxonomic Group' (class), and 'Species Read-Across' (Yes). Below these is a table titled 'Level 2 Data - Primary' with columns: Data Version, NCBI Accession, Protein Count, Species Tax ID, Taxonomic Group, Filtered Taxonomic Group, Scientific Name, Common Name, and Protein Name. The table lists 10 rows of data for various species including Homo sapiens, Hylobates lar, Callithrix jacchus, Cebus capucinus imitator, Gorilla gorilla gorilla, Pan paniscus, Pan troglodytes, Ateles paniscus, Mandrillus leucophaeus, and Pongo abelii. At the bottom right, a dialog box titled 'Opening SeqAPASS\_Level2\_Primary\_Report.csv' is open, showing the file name and a dropdown menu set to 'Open with: Microsoft Excel (default)'. The dialog also has 'OK' and 'Cancel' buttons.

Data Version	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Filtered Taxonomic Group	Scientific Name	Common Name	Protein Name
2	NP_000116.2	1055918	9205	Mammalia	Mammalia	Homo sapiens	human	estrogen receptor isoform 1
2	ARY64719.1	649	9580	Mammalia	Mammalia	Hylobates lar	common gibbon	estrogen receptor alpha
2	XP_008993525.1	64219	9483	Mammalia	Mammalia	Callithrix jacchus	white-lufted-ear marmoset	PREDICTED: estrogen receptor isoform X1
2	XP_017393067.1	53175	1737458	Mammalia	Mammalia	Cebus capucinus imitator	white-faced sapajou	PREDICTED: estrogen receptor isoform X1
2	XP_018884801.1	68319	9595	Mammalia	Mammalia	Gorilla gorilla gorilla	western lowland gorilla	PREDICTED: estrogen receptor isoform X2
2	XP_003811544.1	49145	9597	Mammalia	Mammalia	Pan paniscus	pygmy chimpanzee	PREDICTED: estrogen receptor isoform X2
2	XP_003311598.1	113964	9598	Mammalia	Mammalia	Pan troglodytes	chimpanzee	PREDICTED: estrogen receptor isoform X2
2	ARY64724.1	64	9510	Mammalia	Mammalia	Ateles paniscus	black spider monkey	estrogen receptor alpha
2	XP_011852190.1	38584	9568	Mammalia	Mammalia	Mandrillus leucophaeus	drill	PREDICTED: estrogen receptor isoform X2
2	XP_002817538.1	44332	9901	Mammalia	Mammalia	Pongo abelii	Sumatran orangutan	PREDICTED: estrogen receptor isoform X2

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

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 displays the SeqAPASS web application interface. At the top left, there is a line graph showing 'Percent Similarity' vs 'Percent Identity' with three options: Default, Second Local Minimum, and User Defined. A 'View Cutoff' button is below the graph. To the right is the 'Primary Report Settings' panel with fields for E-value (10.0), Sorted by Taxonomic Group (class), and Species Read-Across (Yes). Below these are 'Update Report' and 'Use Default Settings' buttons.

Below the settings is a filter panel with checkboxes for:
 

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

 Radio buttons for 'Primary Report' (selected) and 'Full Report' are also present.

The main content area is titled 'Level 2 Data - Primary' and features a search bar. Below is a table with the following columns: Data Version, NCBI Accession, Protein Count, Species Tax ID, Taxonomic Group, Filtered Taxonomic Group, Scientific Name, Common Name, and Protein Name. The table contains 10 rows of data for various species including Homo sapiens, Hylobates lar, Callithrix jacchus, Cebus opaninus imitator, Gorilla gorilla gorilla, Pan paniscus, Pan troglodytes, Ateles paniscus, Mandrillus leucophaeus, and Pongo abelii.

At the bottom right, a dialog box titled 'Opening SeqAPASS\_Level2\_Primary\_Report.xls' is open, showing the file name and options to 'Open with Microsoft Excel (default)', 'Save File', or 'Do this automatically for files like this from now on'. 'OK' and 'Cancel' buttons are at the bottom.

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

Updated 3/06/18; 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 top navigation bar of the SeqAPASS application. It features a blue header with the application name and a 'Log out' link. Below this is a white navigation bar with five tabs: 'Home', 'Request SeqAPASS Run', 'SeqAPASS Run Status', 'View SeqAPASS Reports', and 'Settings'. The 'View SeqAPASS Reports' tab is currently selected. At the bottom of the navigation bar, there is a status bar containing the text 'Welcome to SeqAPASS', 'Version 2.0', and 'Logged in as: LaLone.Carlie@epa.gov'.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) <span style="float: right;">Log out</span>				
Home	Request SeqAPASS Run	SeqAPASS Run Status	View SeqAPASS Reports	Settings
Welcome to SeqAPASS		Version 2.0	Logged in as: LaLone.Carlie@epa.gov	

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

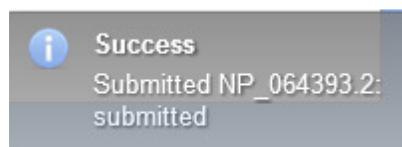
### Pop-up Messages

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.

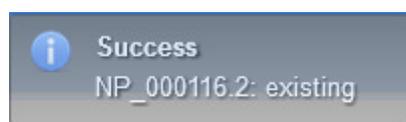


Pop-up messages are meant to guide the user to submit the correct information for a query, inform the user of a successful or failed query submission, or otherwise inform the user of an error. All pop-up messages will appear for 10 seconds on the upper right-hand side of the screen, and then disappear. 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.

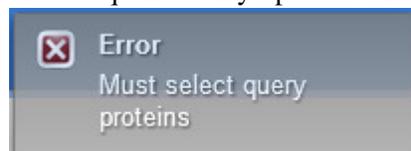
In the “Request SeqAPASS Run” tab, Compare Primary Amino Acid Sequences “By Species” page, a successful Level 1 query submission will display a pop-up message indicating that the query has been submitted to the run queue or if “existing” message appears indicating that the accession has been ran previously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



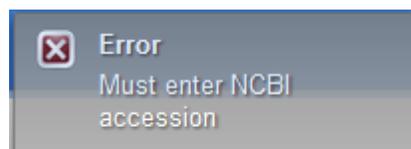
OR



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.



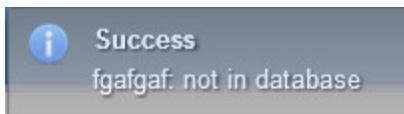
OR



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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

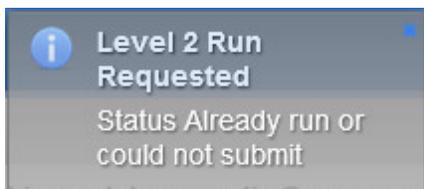
If the user enters non-sense text (or any text that is not an NCBI accession) into the “NCBI Protein Accession” text box for submitting a Level 1 query in the “Request SeqAPASS Run” tab, in the Compare Primary Amino Acid Sequences “By Accession” page, and clicked “Request Run” button, the message below will pop-up indicating that the Accession entered is not in the SeqAPASS database.



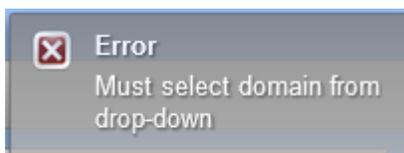
In the “View SeqAPASS Reports” tab, Level 1 page, if a user clicks “View Level 2 Data,” a successful Level 2 query submission will display a pop-up message indicating that the query has entered the run queue.



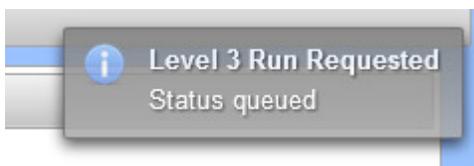
In the “View SeqAPASS Reports” tab, Level 1 page, if a user selects a domain that has already been submitted (but not completed) and clicks “Request Domain Run” a message for successful Level 2 query submission will display a pop-up message indicating that the query has entered the run queue



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.



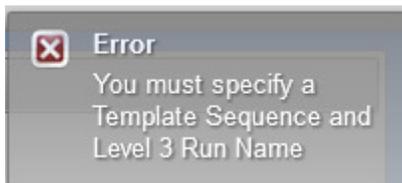
In the “View SeqAPASS Reports” tab, Level 1 page, a successful Level 3 query submission will display a pop-up message indicating that the query has entered the run queue.



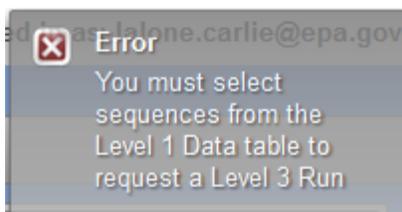
## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

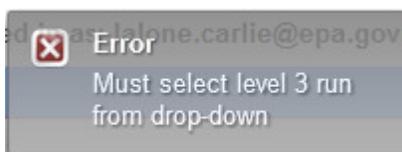
In the “View SeqAPASS Reports” tab, Level 1 page, if a user fails to type a user defined Level 3 Run Name, the message below will pop-up to indicate that the user must do so.



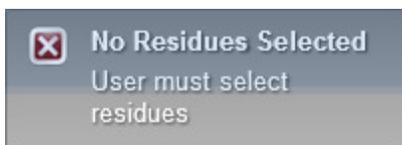
In the “View SeqAPASS Reports” tab, Level 1 page, if a user fails to select species from the Level 1 Data table to be compared with the template sequence, the message below will pop-up.



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 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

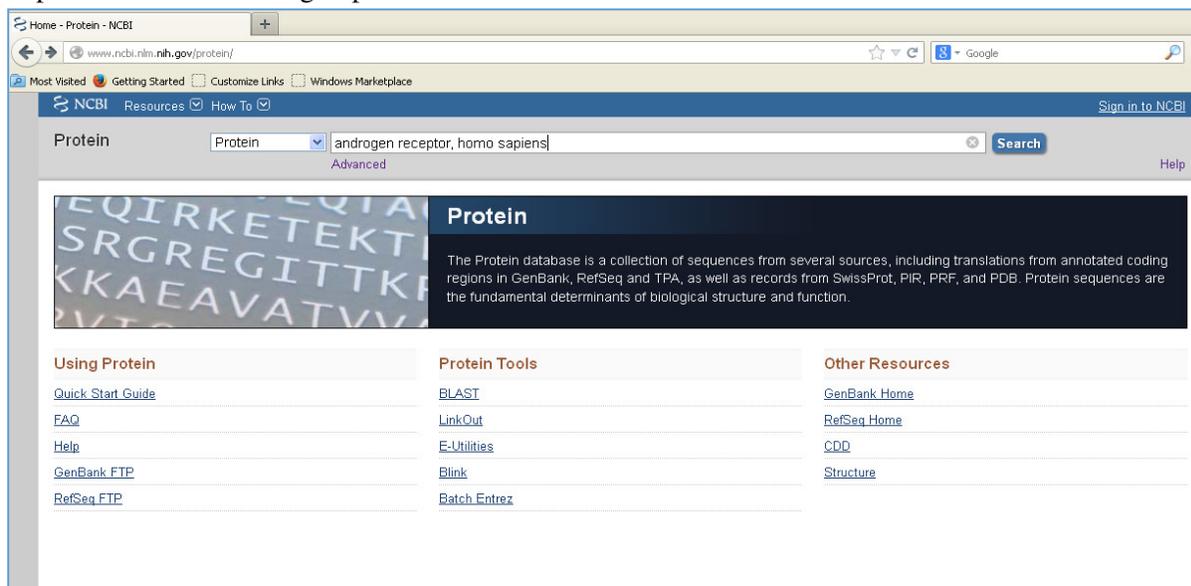
### Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) Documentation

**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 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 query. 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 interface. The search bar contains the text "androgen receptor, homo sapiens" and the search button is labeled "Search". Below the search bar, there is a section titled "Protein" 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 description, 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).

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

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

The screenshot shows the NCBI protein search interface. The search term is 'androgen receptor, homo sapiens'. The results are displayed in a list format, showing the first five hits. Each hit includes the protein name, accession number, and a link to view the protein details. The first hit is 'androgen receptor [Homo sapiens]' with accession number AAA51771.1. The right sidebar contains filters for organisms, related data, search details, and recent activity.

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

The screenshot shows the detailed view of the protein 'androgen receptor [Homo sapiens]' (GenBank AAA51771.1). The page includes a summary of the protein's properties, such as its length (917 aa) and accession number. It also provides a list of references, including the original publication by Tilley et al. (1989) and a more recent study by Marcelli et al. (1990). The right sidebar offers options to analyze the sequence, view the protein 3D structure, and find related articles.

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

Updated 3/06/18; 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.

Important Note: To identify which query protein has the greatest number of Ortholog Candidates the user can choose to submit multiple proteins with the same species and protein. Upon the Level 1 runs completing for those similar proteins, the user can then select the “View SeqAPASS Reports” tab and look at the table for “Ortholog Count” the protein with the highest number is likely to be the most appropriate query species for a SeqAPASS evaluation.

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

## Example: Androgen receptor, *Homo sapiens*

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 31-OCT-1994

DEFINITION androgen receptor [Homo sapiens].

ACCESSION AA051771

VERSION AA051771.1 GI:176872

RESOURCE locus HOMOLOC accession [M21748.1](#)

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)

PUBMED 2311578

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)

PUBMED 2339426

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="MR\_LBD\_AR"

/note="MR-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="MR\_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 gqgmpgal

121 nchpergcpv ppaavaakh glpqlpapp dehdraapst llllgtlps ltrradlkd

181 llrartnml lggggggvss ggggggare rrgggsdhd rylggstss dshelckav

241 svnmglgvea lehlrpgql rgdnyapll gppavpqp ctaieckag; llldagktr

301 edkaeyptk gytgkleg elgcgaaa gsgtelrp tlrlxkag deaayqrd

361 yynplalag ppppppppp harikterpl dyaawaaa acrygdla lhgaaagp

421 sgp:aaas: sbeclitae gilycpqgg gggggggggg gggggggggg tagavapyy

481 tpppqlagq eddtapdw ypggvsrv yppctvks wggmdyys pygdrleta

541 rdlvpldy fpppaelic gbaagchgy altpgclqv hkaaeqk ylaarnde

601 ldrtdmcp rlylcyra gmlgarklk hnglkige qaarvcp eecqglvov

661 hieytcpi flrvtaiep gvcaqhdn qdifaall; sinelgerl vhwvmtkal

721 pgrlrvhdv qnavqyrom glvvtanqr stvsvml ytapdlvne ymbkmys

781 qcvamhlgv efgolqtpq eflnkal11 fsiipvdgk ngtfdelm rylkeldrri

841 ackerlptc rryfqlckl ldsvapiare lhgtdflll khwvsvdp emaeiivg

901 vpkilgqvk pythq

//

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

h. If multiple proteins appear to be the best query protein for SeqAPASS, the sequences can be aligned using NCBI's COBALT. Enter (copy and paste from NCBI protein search list) accessions and align.

Alignment page will be generated

Accession	Description	Links
<a href="#">P10275.2</a>	RecName: Full=Androgen receptor; AltName: Full=Dihydrotestosterone receptor; AltName: Full=Nucle	<a href="#">GM</a>
<a href="#">AAA51772.1</a>	androgen receptor [Homo sapiens] >gb AAA51771.1  androgen receptor [Homo sapiens]	<a href="#">G</a>
<a href="#">AAA51780.1</a>	androgen-receptor [Homo sapiens]	<a href="#">GM</a>
<a href="#">AAA51771.1</a>	androgen receptor [Homo sapiens] >gb AAA51772.1  androgen receptor [Homo sapiens]	<a href="#">GM</a>
<a href="#">AAA51729.1</a>	androgen receptor [Homo sapiens] >sp P10275.2 ANDR_HUMAN RecName: Full=Androgen receptor;	<a href="#">GM</a> <a href="#">PubChem BioAssay Info linked to AAA51729.1</a>
<a href="#">AAD45921.1</a>	androgen receptor [Homo sapiens]	<a href="#">GM</a>
<a href="#">AAA51886.1</a>	androgen receptor [Homo sapiens] >sp P10275.2 ANDR_HUMAN RecName: Full=Androgen receptor;	<a href="#">G</a> <a href="#">PubChem BioAssay Info linked to AAA51886.1</a>

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

Updated 3/06/18; 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 lists P10275.2, AAA51772.1, AAA51780.1, AAA51771.1, AAA51729.1, AAD45921.1, and AAA51886.1. The 'Description' column provides details for each accession. The 'Links' column contains icons for UniGene, GEO, Gene, Structure, and Map Viewer. Below the table, there's an 'Alignments' section with a dropdown menu for 'Conservation Setting' set to 'Identity'. The alignment view shows sequence alignments for each accession, with the 'Conservation Setting' dropdown menu highlighted in red.

Accession	Description	Links
<input checked="" type="checkbox"/> P10275.2	RecName: Full=Androgen receptor; AltName: Full=Dihydrotestosterone receptor; AltName: Full=Nucle	<a href="#">G</a> <a href="#">M</a>
<input checked="" type="checkbox"/> AAA51772.1	androgen receptor [Homo sapiens] >gb AAA51771.1  androgen receptor [Homo sapiens]	<a href="#">G</a>
<input checked="" type="checkbox"/> AAA51780.1	androgen-receptor [Homo sapiens]	<a href="#">G</a> <a href="#">M</a>
<input checked="" type="checkbox"/> AAA51771.1	androgen receptor [Homo sapiens] >gb AAA51772.1  androgen receptor [Homo sapiens]	<a href="#">G</a> <a href="#">M</a>
<input checked="" type="checkbox"/> AAA51729.1	androgen receptor [Homo sapiens] >sp P10275.2 ANDR_HUMAN RecName: Full=Androgen receptor;	<a href="#">G</a> <a href="#">M</a> <a href="#">PubChem BioAssay Info linked to AAA51729.1</a>
<input checked="" type="checkbox"/> AAD45921.1	androgen receptor [Homo sapiens]	<a href="#">G</a> <a href="#">M</a>
<input checked="" type="checkbox"/> AAA51886.1	androgen receptor [Homo sapiens] >sp P10275.2 ANDR_HUMAN RecName: Full=Androgen receptor;	<a href="#">G</a> <a href="#">M</a> <a href="#">PubChem BioAssay Info linked to AAA51886.1</a>

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 3/06/18; 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, Evalue, 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 queried protein selected by the user. For each sequence queried, the Level 1, top row query sequence 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: SeqAPASS v2.0 and newer parse the BLASTP query and hit accessions to identify all the species/accessions from the identical proteins. Therefore, if a hit sequence represents multiple species, all species with the identical sequence will be found in the data table for Level 1 and Level 2. To determine which sequence/species was identified from BLASTP as a hit and which sequence/species was parsed from the identical sequence, view the "Full Report" for Level 1 or Level 2, column "Identical Protein," Where "N" is indicative of the original hit sequence and "Y" is the parsed sequence.

### *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.

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

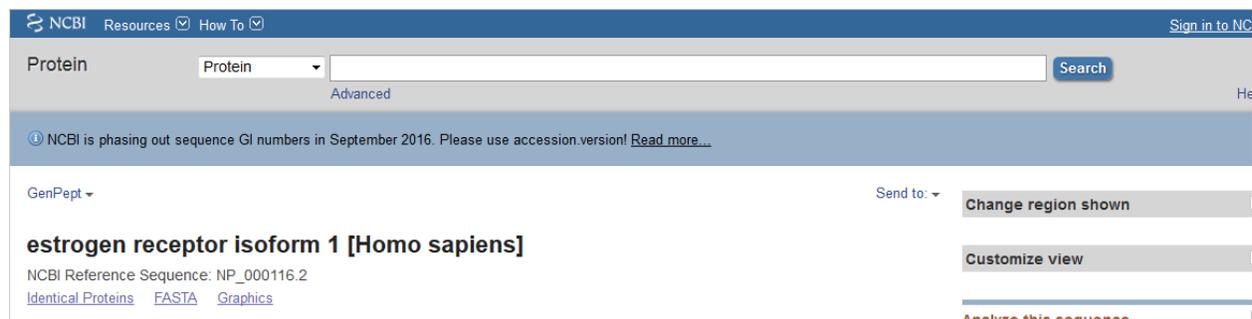
Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

### *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 or one of its identical protein matches 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.

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. 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 Proteins.”

Click the “Identical Proteins” link and look for a sequence in the list from the user defined query species.



The screenshot shows the NCBI protein page for "estrogen receptor isoform 1 [Homo sapiens]". The page includes the NCBI logo, navigation links (Resources, How To), and a search bar. Below the search bar, there is a message about NCBI phasing out sequence GI numbers. The main content area shows the protein name, NCBI Reference Sequence (NP\_000116.2), and links for "Identical Proteins", "FASTA", and "Graphics". There are also buttons for "Change region shown" and "Customize view".

Note: If the top hit is a Protein DataBank (PDB) code (e.g., 1AHR\_A) from RBH BLAST there will be no ortholog candidates identified. BLASTP when ran against all accessions for a given species does not return PDB codes. It is recommended that the user identify a similar/identical sequence to the PDB code and use that sequence as the query sequence.

### *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” or “Full Report” and identifying the local minimums in the data. The default cut-off is determined by taking the 1<sup>st</sup> 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 (when “Primary Report Settings” is set to “Species Read-Across:” Yes):*

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

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

“yes”. This criterion 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  $\leq 0.01$  and Common Domain Count  $\geq 1$ .

*Criteria for Susceptibility Prediction (when “Primary Report Settings” is set to “Species Read-Across:” No):*

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 and not identified as an ortholog candidate (Ortholog Candidate =N, for “no,”), the hit is predicted to not be susceptible; therefore, Susceptibility Prediction = N for “no”

### *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” or “Full Report” and identifying the local minimums in the data. The default cut-off is determined by taking the 1<sup>st</sup> 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.

*Level 2 Criteria for Susceptibility Prediction (when “Primary Report Settings” is set to “Species Read-Across:” Yes):*

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

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

Updated 3/06/18; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

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 criterion 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  $\leq 0.01$  and Common Domain Count  $\geq 1$ .

*Level 2 Criteria for Susceptibility Prediction (when “Primary Report Settings” is set to “Species Read-Across:” No):*

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 and not identified as an ortholog candidate (Ortholog Candidate =N, for “no,”), 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.