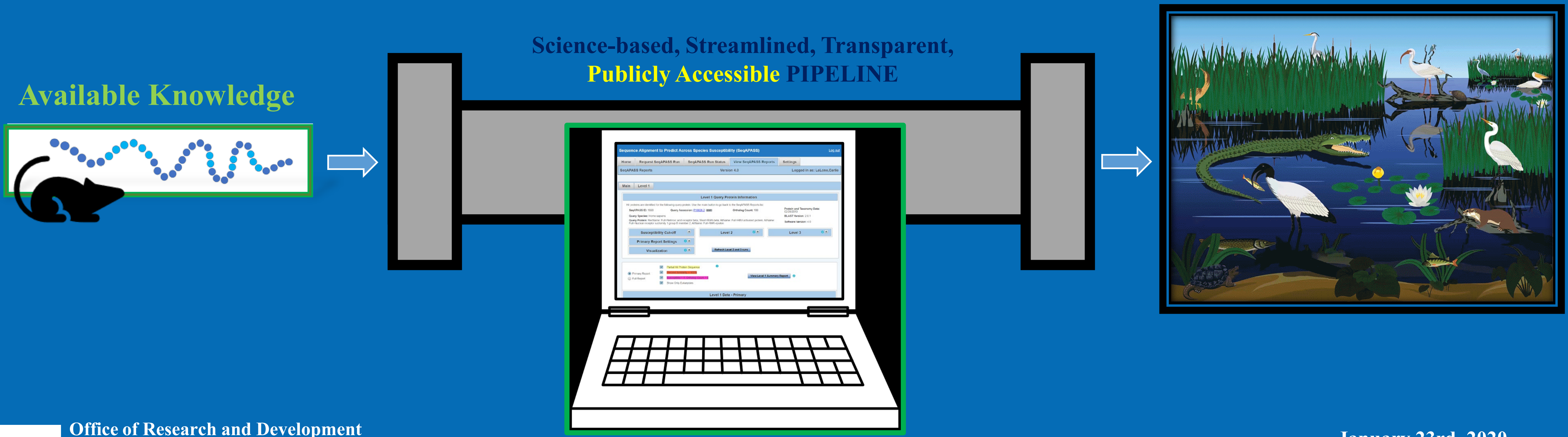
United States
Environmental Protection
Agency

Evolution of the Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) Tool

Carlie A. LaLone



Predict Susceptibility

Science-based, Streamlined, Transparent,
Publicly Accessible PIPELINE

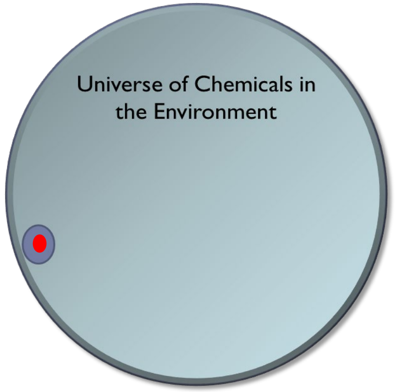
Available Knowledge

Office of Research and Development
Center for Computational Toxicology and Exposure, Great Lakes Toxicology and Ecology Division

January 23rd, 2020

Chemical Safety Evaluation

- Protect human health and the environment
  - Ensure that chemicals in the marketplace are reviewed for safety
- Challenging mission:
  - Tens of thousand of chemicals are currently in use and hundreds are introduced annually
  - Many have not been thoroughly evaluated for potential risk to human health and the environment
    - *Chemicals tested across species: Even more sparse*



Reduce Animal Testing at the US EPA

- EPA Administrator Andrew Wheeler signed directive (Sept. 10th 2019) to reduce animal testing
  - Calls for the Agency to:
    - Reduce its request for, and funding of, mammal studies by 30% by 2025
      - That is ~5 years from today!
    - Eliminate all mammal study requests and funding by 2035
      - That is ~15 years from today

How do we get there?
NAMs

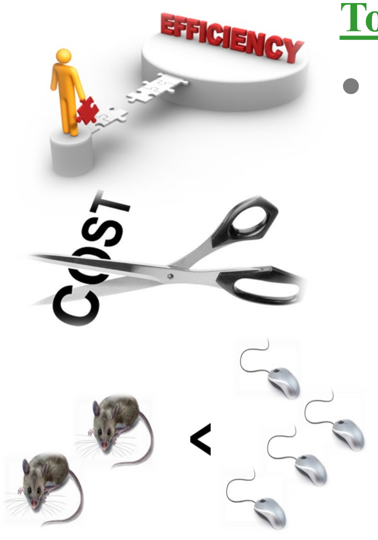
Transformation of Toxicity Testing

Historically:

Whole animal test

- **Observe Toxic Outcome**
  - **- Examples**
    - **tumor development**
    - **mortality**

Resource intensive

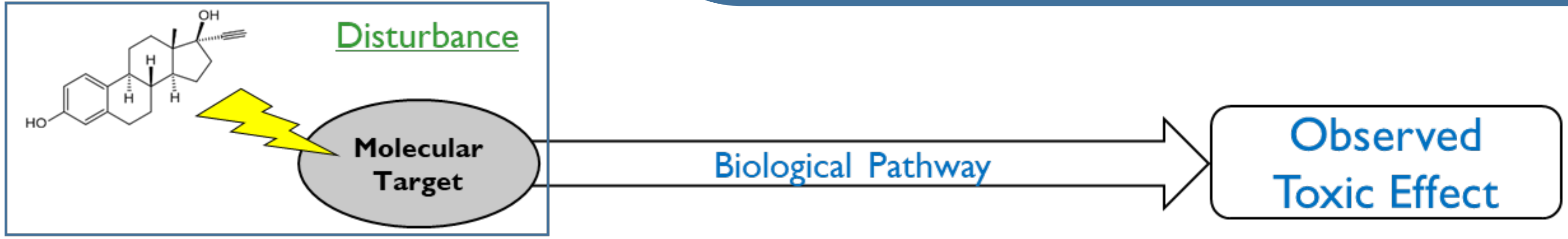


Toxicity Testing in the 21st Century:

- ***In vitro* and *in silico* methods**
  - **Pathway-based approaches**
  - **Focus on disturbance of the biological pathway**
    - **Predictive of the observable toxic effects**

New Approach Methods (NAMs)

- Informatics
- High throughput
- Systems biology
- OMICs



Disturbance

Molecular Target

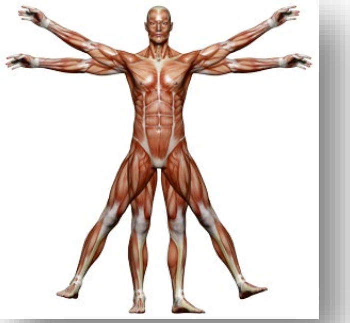
Biological Pathway

Observed Toxic Effect

Enabled by evolution of the science and technology

Model Organisms for Toxicity Testing

- Assumed that sensitivity of species to a chemical is a function of their relatedness



Cannot Test

≅

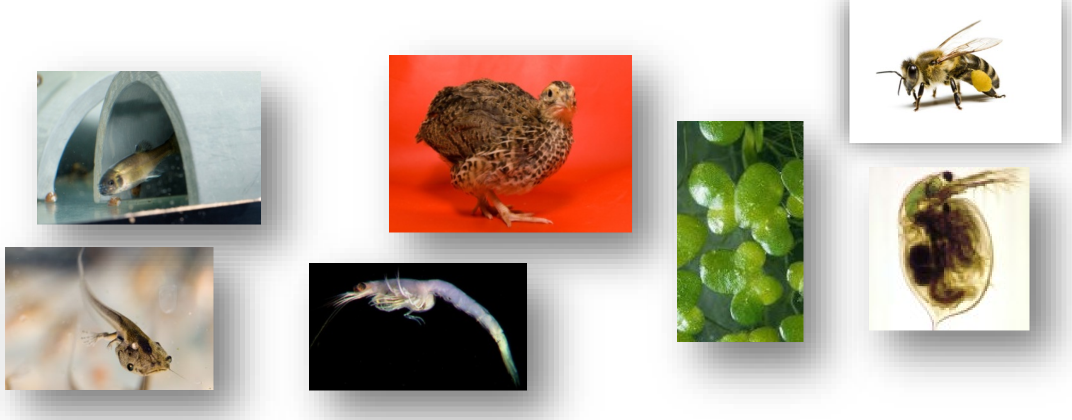
- **Human Health Risk Assessment**

Use of Surrogates

- **Ecological Risk Assessment**

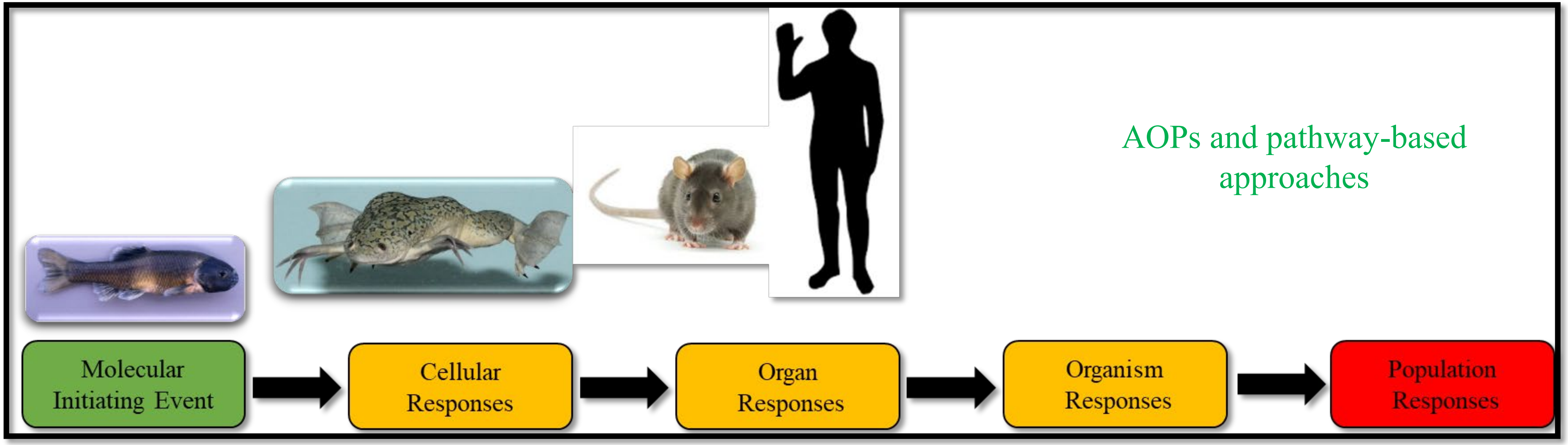
Cannot Test

≅



Representative species across a diversity of organism classes

Surrogates in 21st Century Chemical Safety



Molecular Initiating Event → Cellular Responses → Organ Responses → Organism Responses → Population Responses

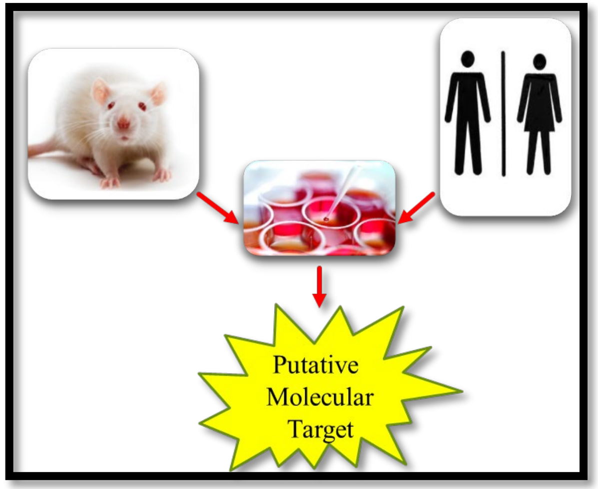
AOPs and pathway-based approaches

AI and Text Mining of Available Toxicity Literature

Is the model organism
A good surrogate for the
Species we are trying to protect?
Maybe

Could we gather evidence rapidly
to help us understand this question?
Absolutely!!!

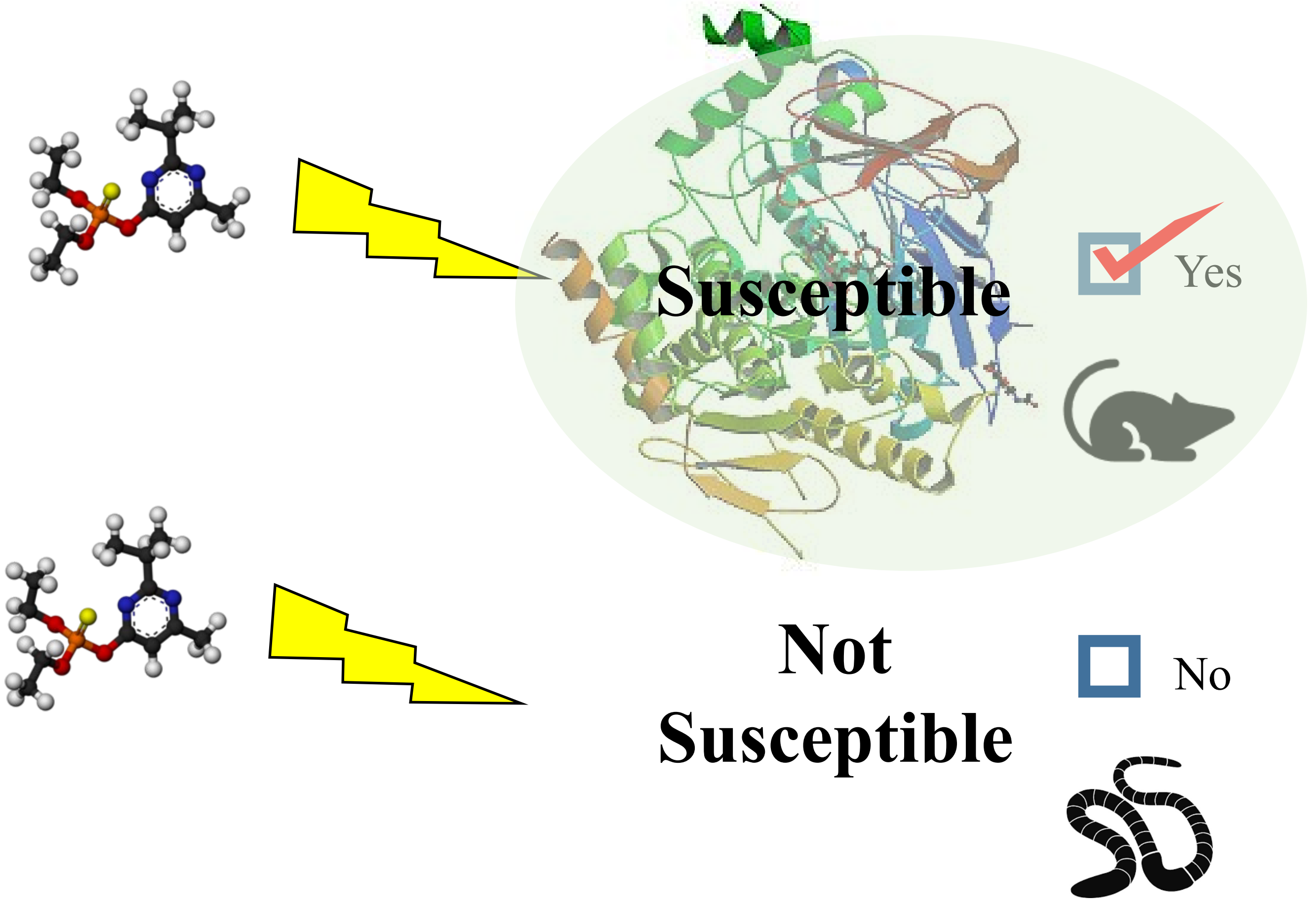
HTS



Considering chemical sensitivity?

Factors that make a species sensitive

- Exposure
- Dose
- ADME
- Target receptor availability
- Life stage
- Life history
- etc.
- etc.



Susceptible ☑ Yes

Not Susceptible ☐ No

Simple question to address:

Is the known chemical target available in a species for a chemical to act upon?

Yes or No

Likely susceptible or Not likely susceptible (at least through the known mechanism)

New Approach Methods: Species Extrapolation

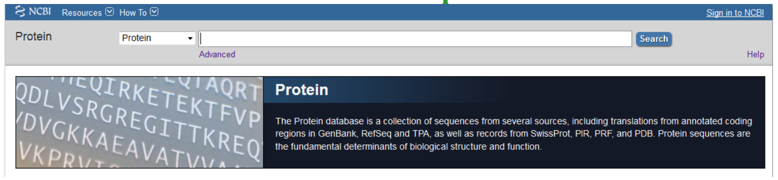
New tools and technologies have emerged

- Improved sequencing technologies
- Large databases of sequence data

As of this week

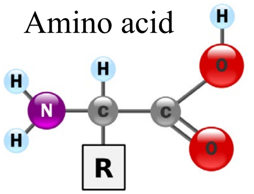
~161 million Proteins

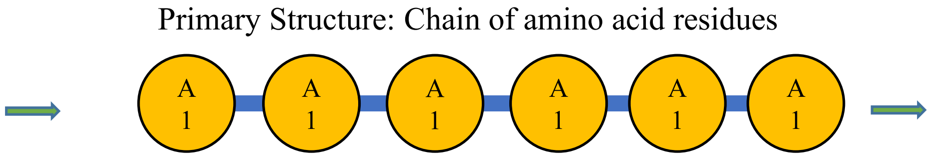
~98 thousand Species

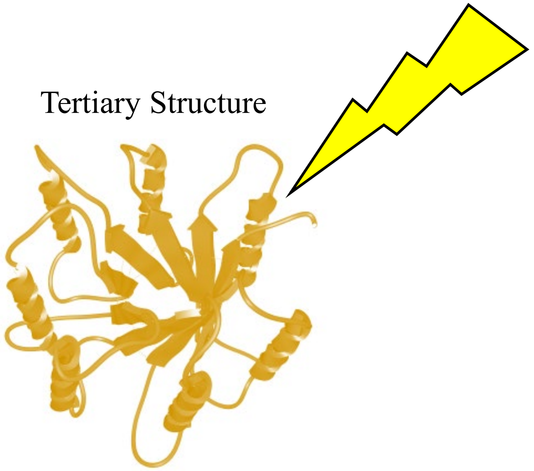


The Protein database is a collection of sequences from several sources, including translations from annotated coding regions in GenBank, RefSeq and TPA, as well as records from SwissProt, PIR, PRF, and PDB. Protein sequences are the fundamental determinants of biological structure and function.

- **Focus on the molecular machine: The Protein**
  - **Large biomolecule assembled from amino acids encoded in genes**



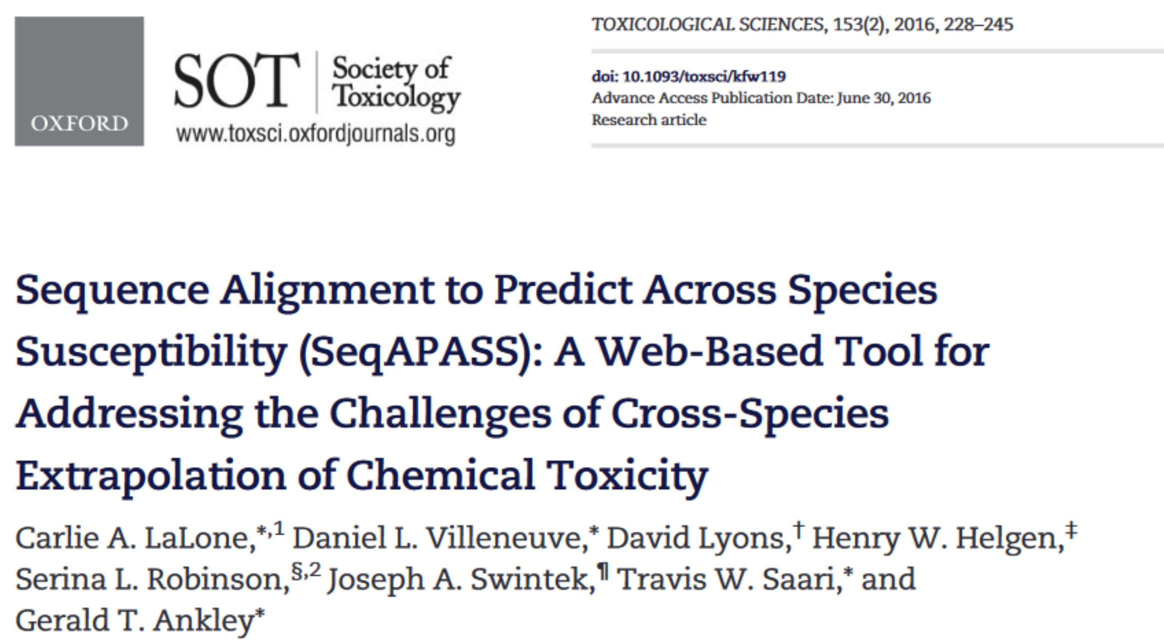




  - **Many functions (e.g., catalyze reactions, structural/mechanical functions, cell signaling, immune response, etc.)**

- **Evaluate protein similarity between species**
  - **Moving away from empirical testing and qualitative understanding of molecular target (protein) conservation to quantitative measures**

EPA
United States
Environmental Protection
Agency



OXFORD
SOT Society of Toxicology
www.toxsci.oxfordjournals.org

TOXICOLOGICAL SCIENCES, 153(2), 2016, 228–245

doi: 10.1093/toxsci/kfw119
Advance Access Publication Date: June 30, 2016
Research article

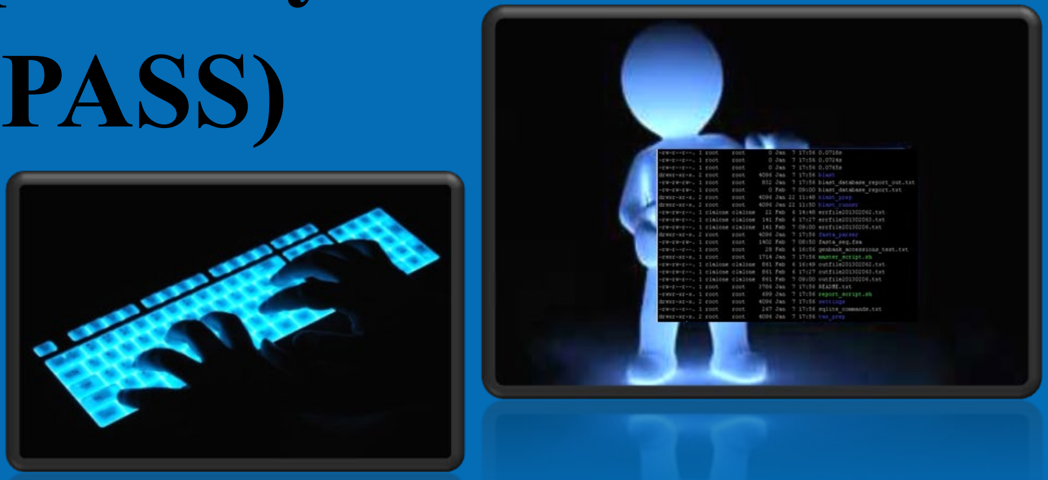
Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): A Web-Based Tool for Addressing the Challenges of Cross-Species Extrapolation of Chemical Toxicity

Carlie A. LaLone,[*,1] Daniel L. Villeneuve,[*] David Lyons,[†] Henry W. Helgen,[‡]
Serina L. Robinson,[§,2] Joseph A. Swintek,[¶] Travis W. Saari,[*] and
Gerald T. Ankley[*]

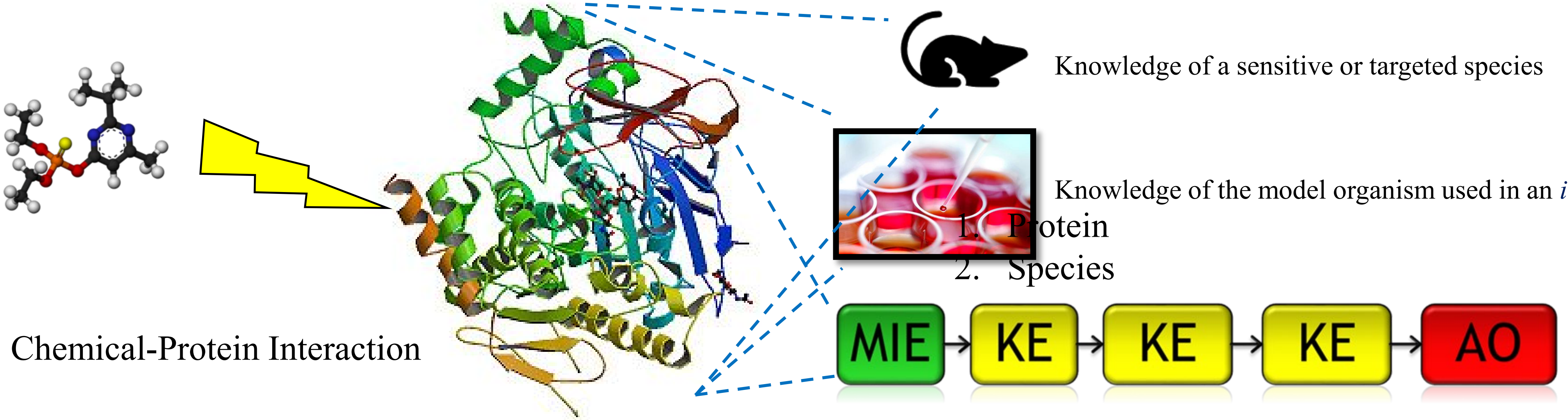
https://seqapass.epa.gov/seqapass/

Sequence Alignment to Predict Across Species Susceptibility

(SeqAPASS)



What information is required for a SeqAPASS query?



Knowledge of a sensitive or targeted species

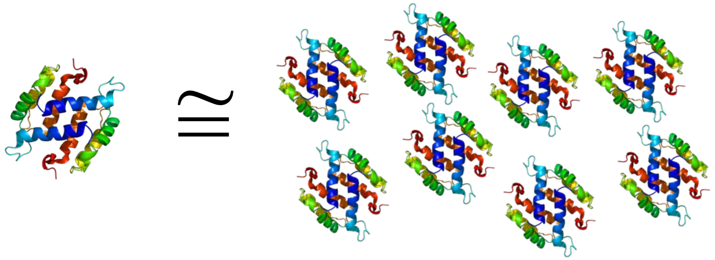
Knowledge of the model organism used in an *in vitro* assay

1. Protein
2. Species

Chemical-Protein Interaction

Knowledge of the species for which the Key Event was developed

Chemical Molecular Target
in Target Species

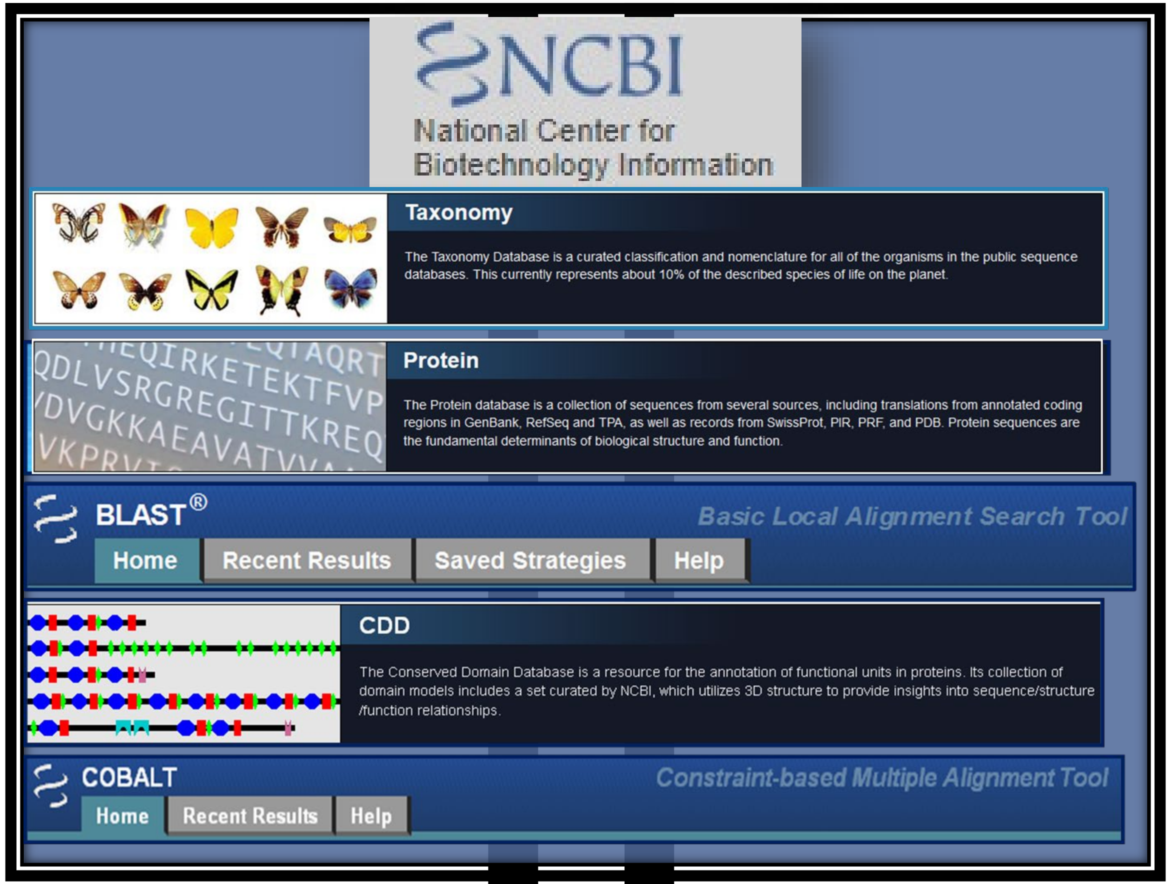


Compare to Millions of Proteins
From Thousands of Species

Greater similarity = Greater likelihood that chemical can act on the protein
Line of Evidence: Predict Potential Chemical Susceptibility Across Species

Animation by: Miguel Moravec (EPA CSS) & Andrew Patterson

SeqAPASS



Taxonomy

The Taxonomy Database is a curated classification and nomenclature for all of the organisms in the public sequence databases. This currently represents about 10% of the described species of life on the planet.

Protein

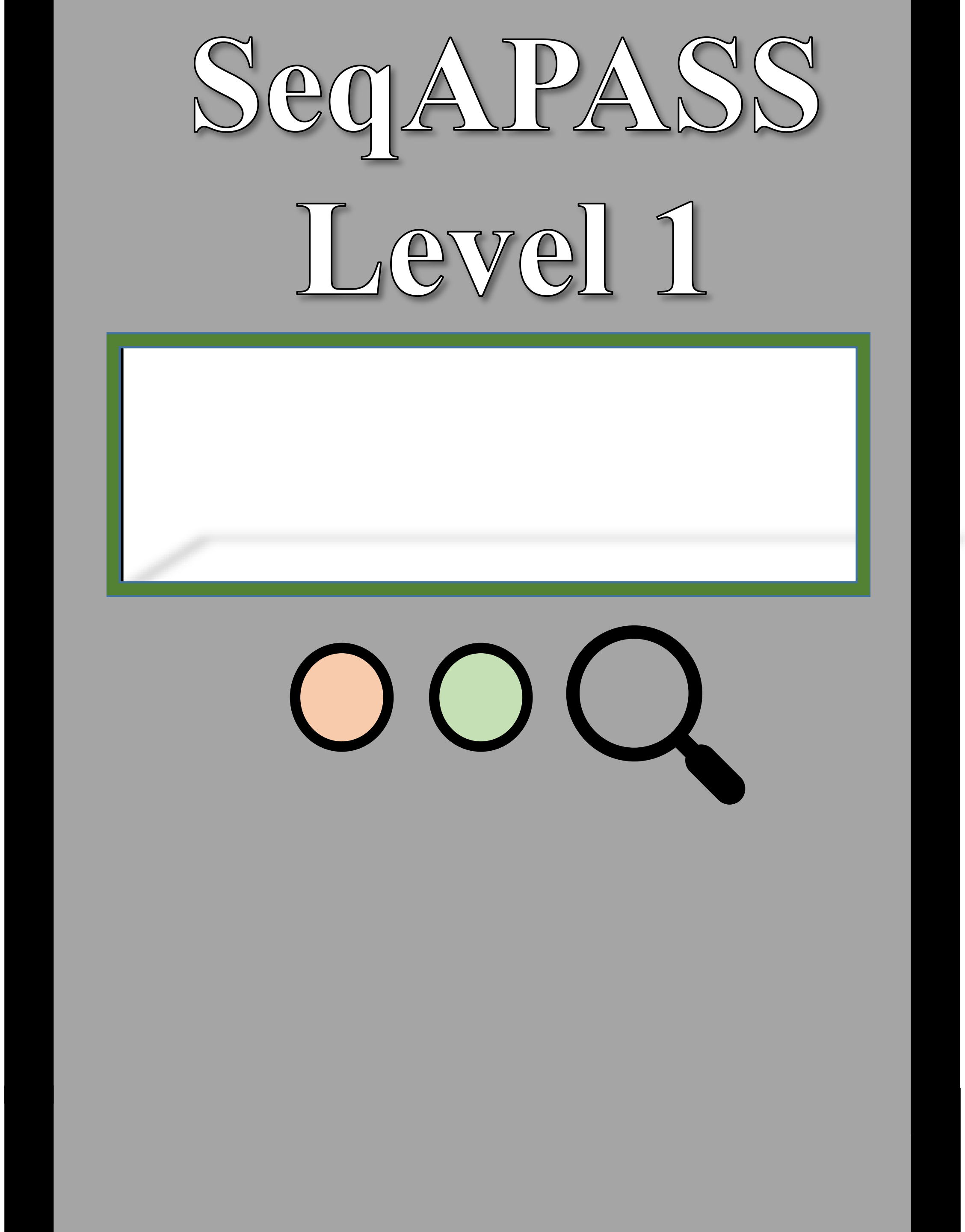
The Protein database is a collection of sequences from several sources, including translations from annotated coding regions in GenBank, RefSeq and TPA, as well as records from SwissProt, PIR, PRF, and PDB. Protein sequences are the fundamental determinants of biological structure and function.

CDD

The Conserved Domain Database is a resource for the annotation of functional units in proteins. Its collection of domain models includes a set curated by NCBI, which utilizes 3D structure to provide insights into sequence/structure /function relationships.

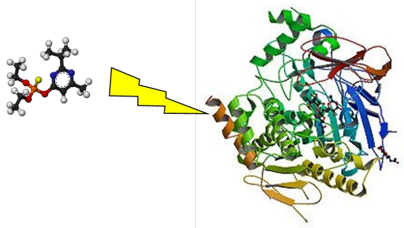
SeqAPASS

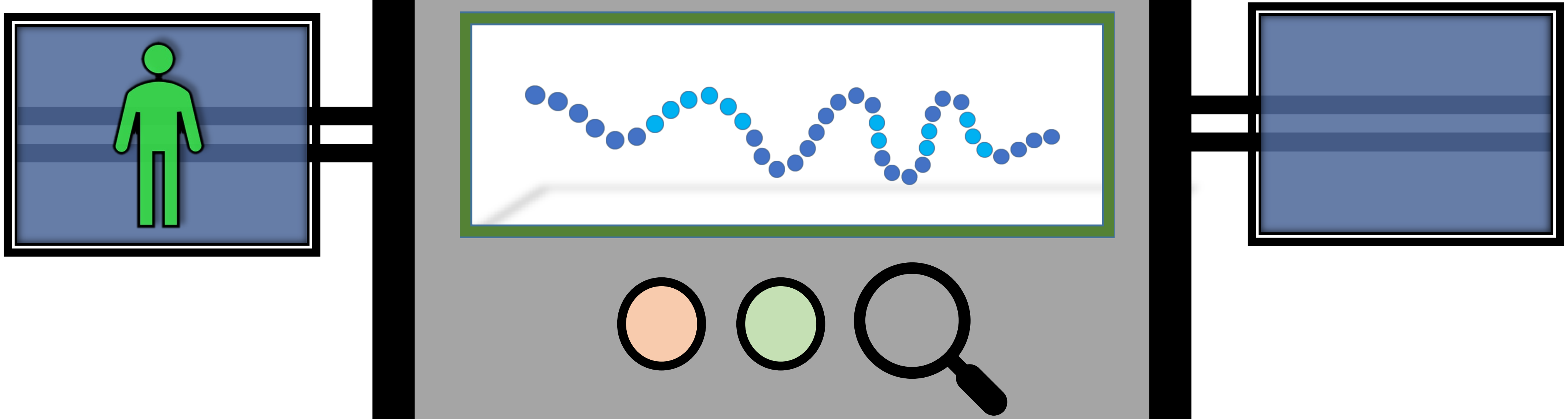
SeqAPASS
Level 1



SeqAPASS Level 1

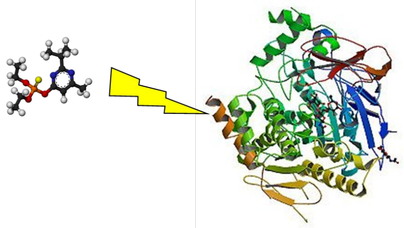
Human Protein Target



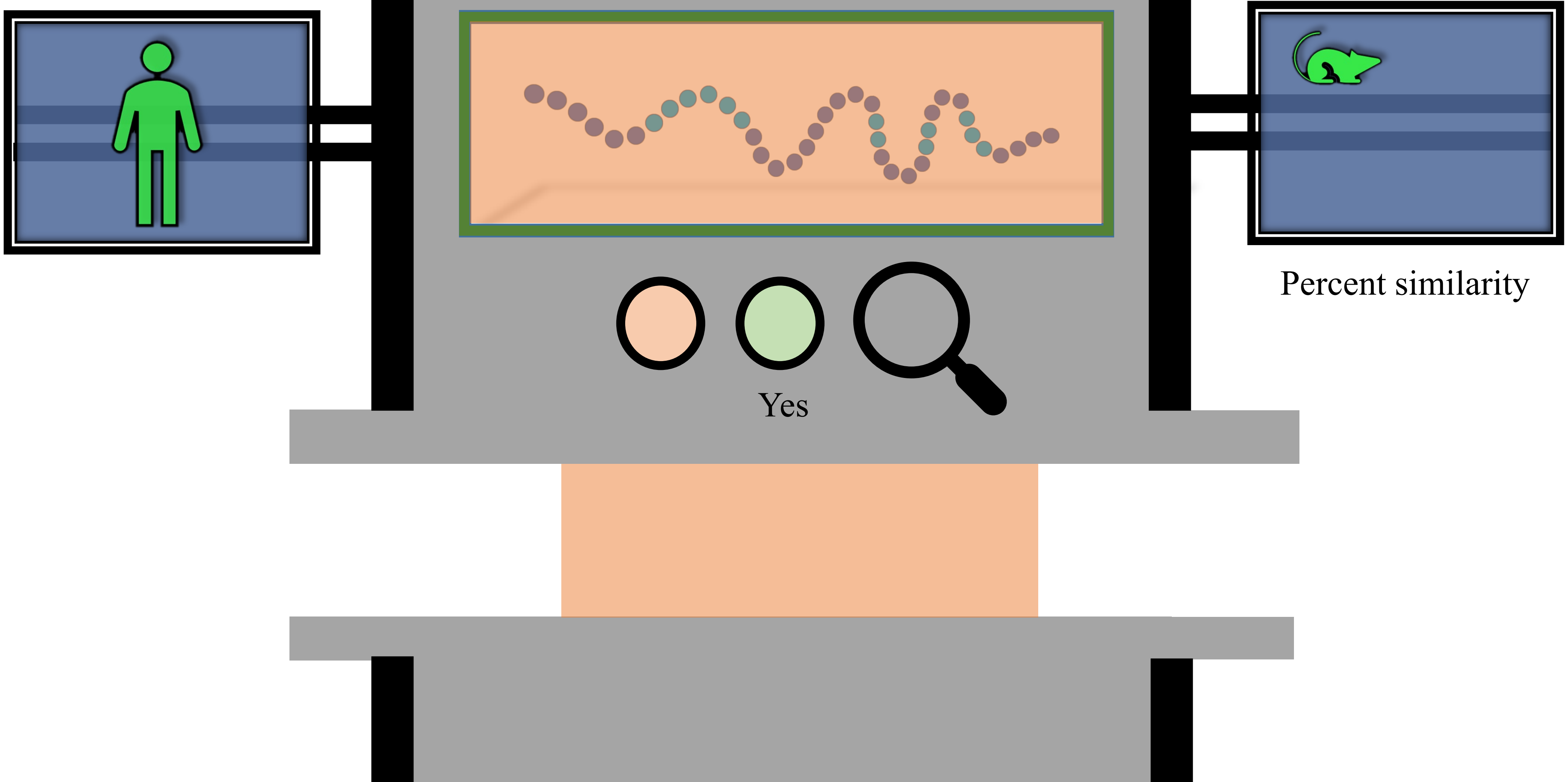


SeqAPASS Level 1

Human Protein Target



Line of Evidence:
Primary amino acid sequence
Conserved

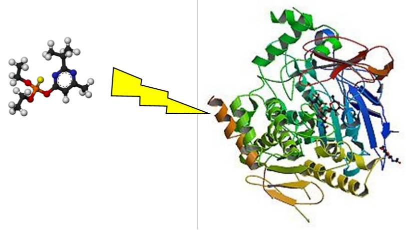


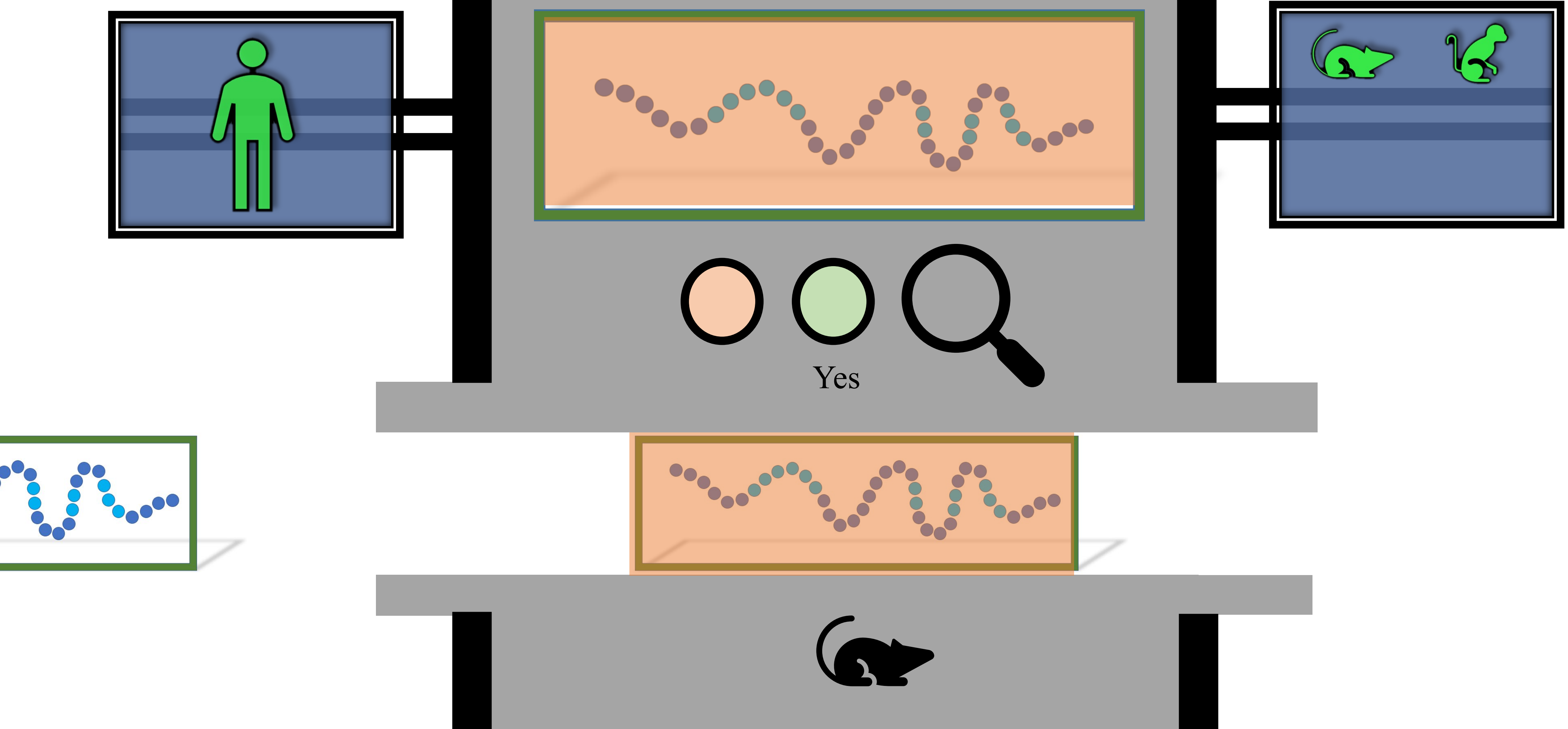
Percent similarity

Yes

SeqAPASS Level 1

Human Protein Target



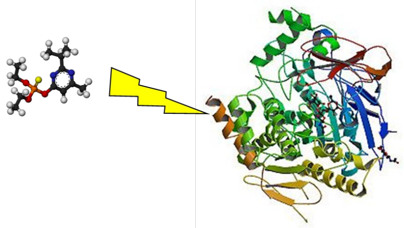


Line of Evidence:
Primary amino acid sequence
Conserved

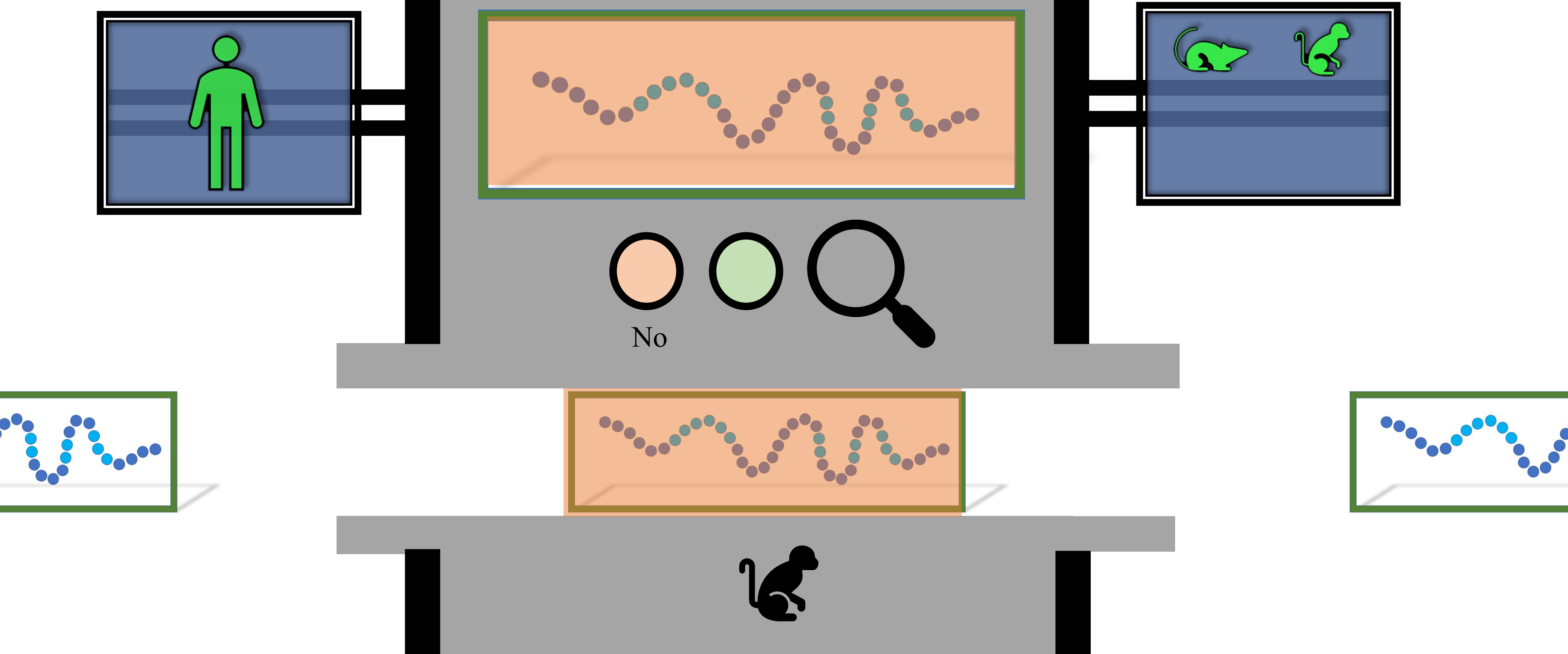
Yes

SeqAPASS
Level 1

Human Protein Target



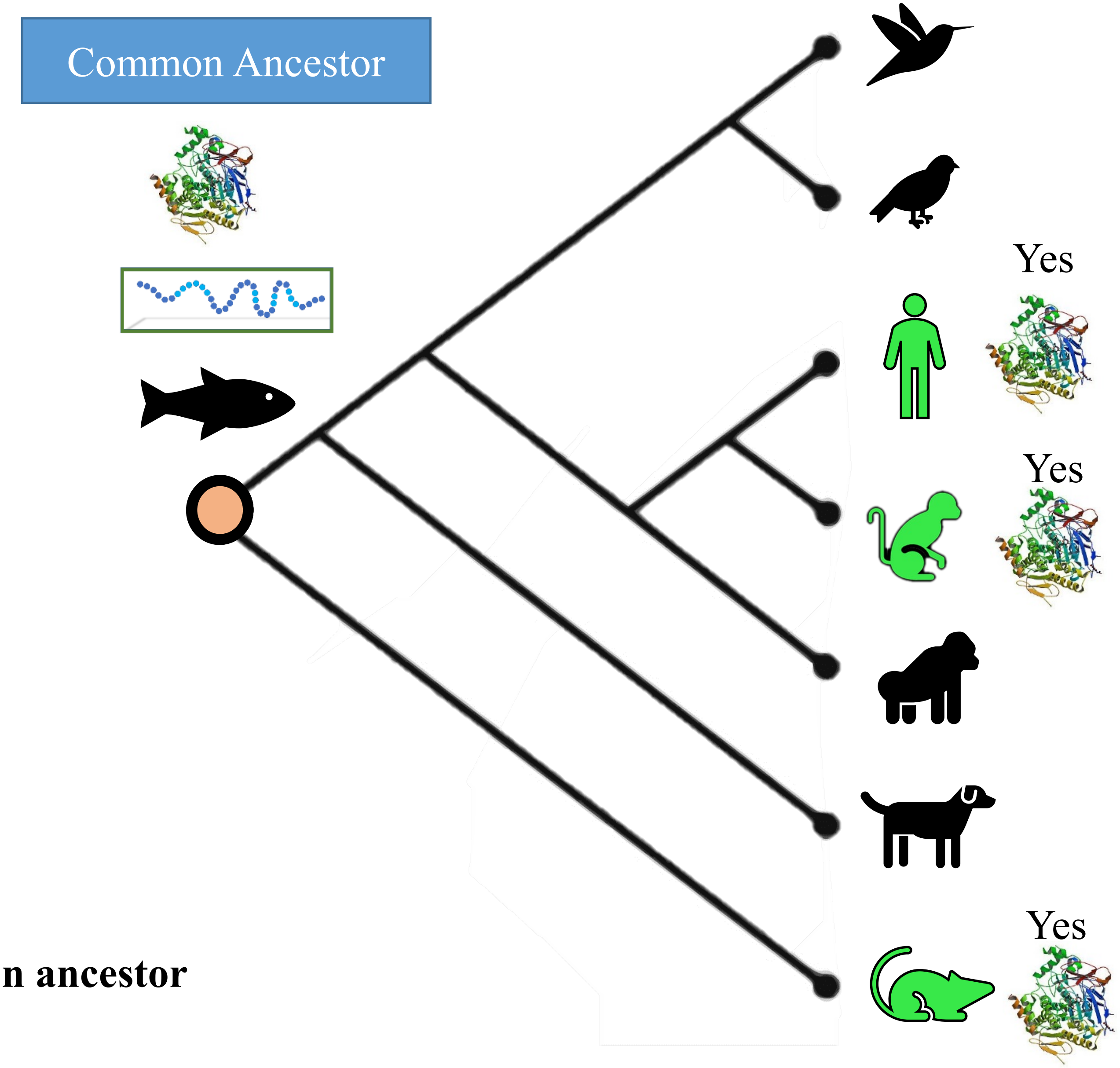
Line of Evidence:
Primary amino acid sequence
Conserved



SeqAPASS Level 1

Ortholog Candidate Identification

Yes



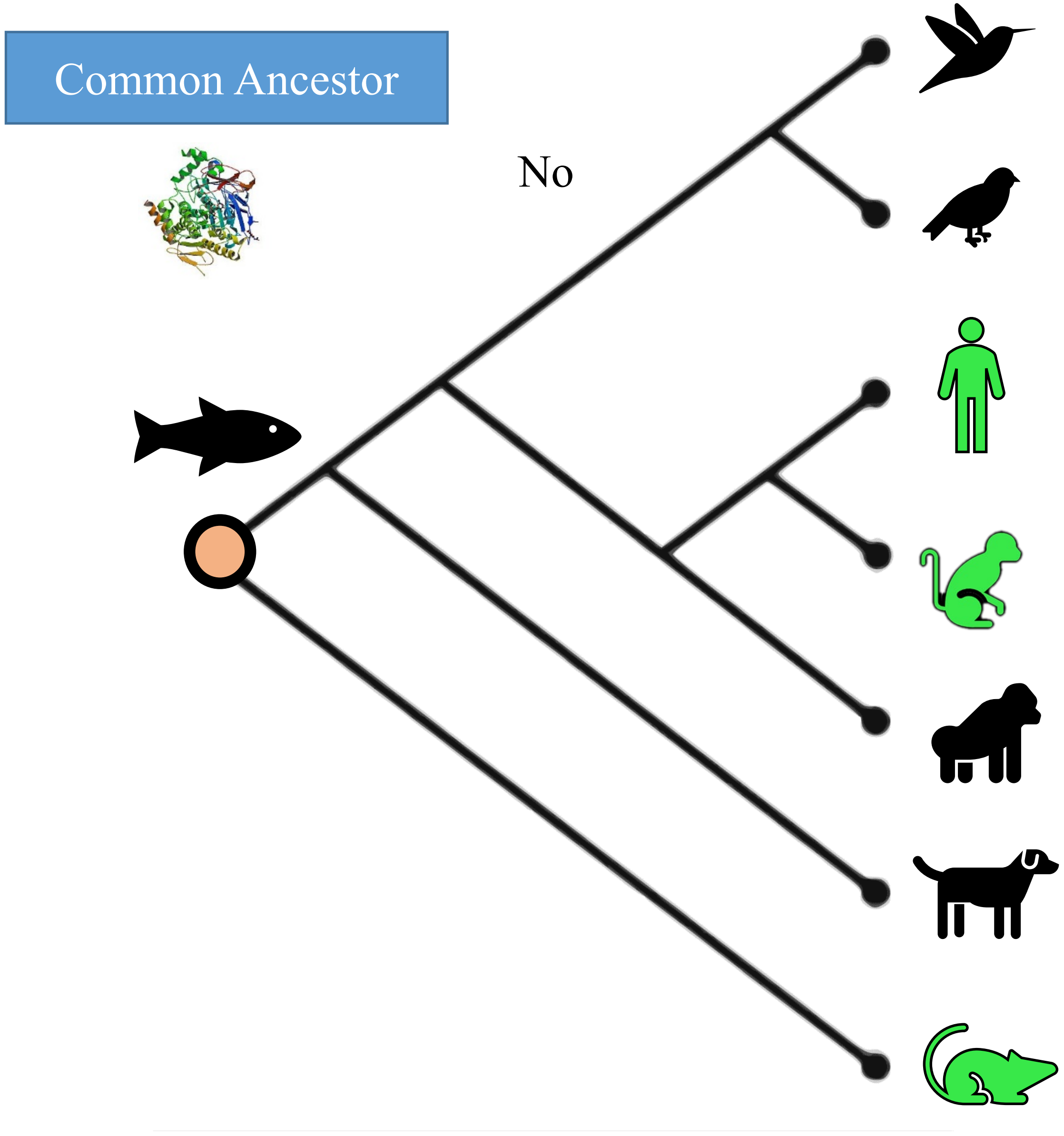
Proteins in different species that evolved from a common ancestor

Typically maintain similar function

SeqAPASS Level 1

Ortholog Candidate Identification

No



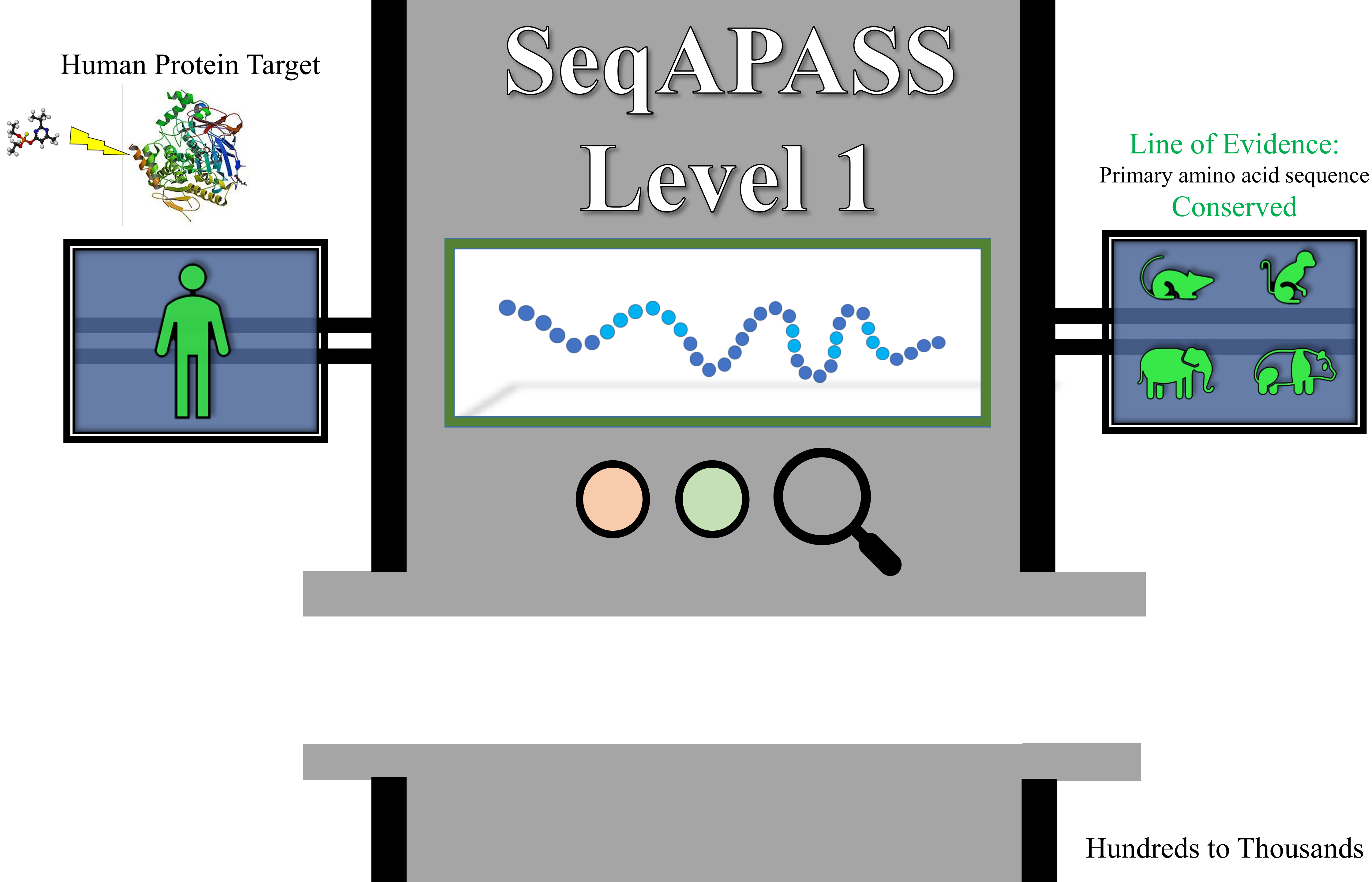
SeqAPASS Level 1

Common Name	Ortholog Candidate	Cut-off	Percent Similarity
Human	Y	33.15	100
Florida manatee	Y	33.15	98.8
Mallard	Y	33.15	82.29
Rock pigeon	Y	33.15	80.93
Green anole	Y	33.15	80.65
Pacific transparent sea squirt	Y	33.15	33.15
Yesso scallop	N	33.15	32.87
Purple sea urchin	N	33.15	26.05
Human whipworm	N	33.15	23.53
Bed bug	N	33.15	21.62

Lowest % Similarity that is still an ortholog

Example:
Susceptibility Cut-off: Set at 33.15
Above cut-off: More likely to be susceptible base on similar **FUNCTION**

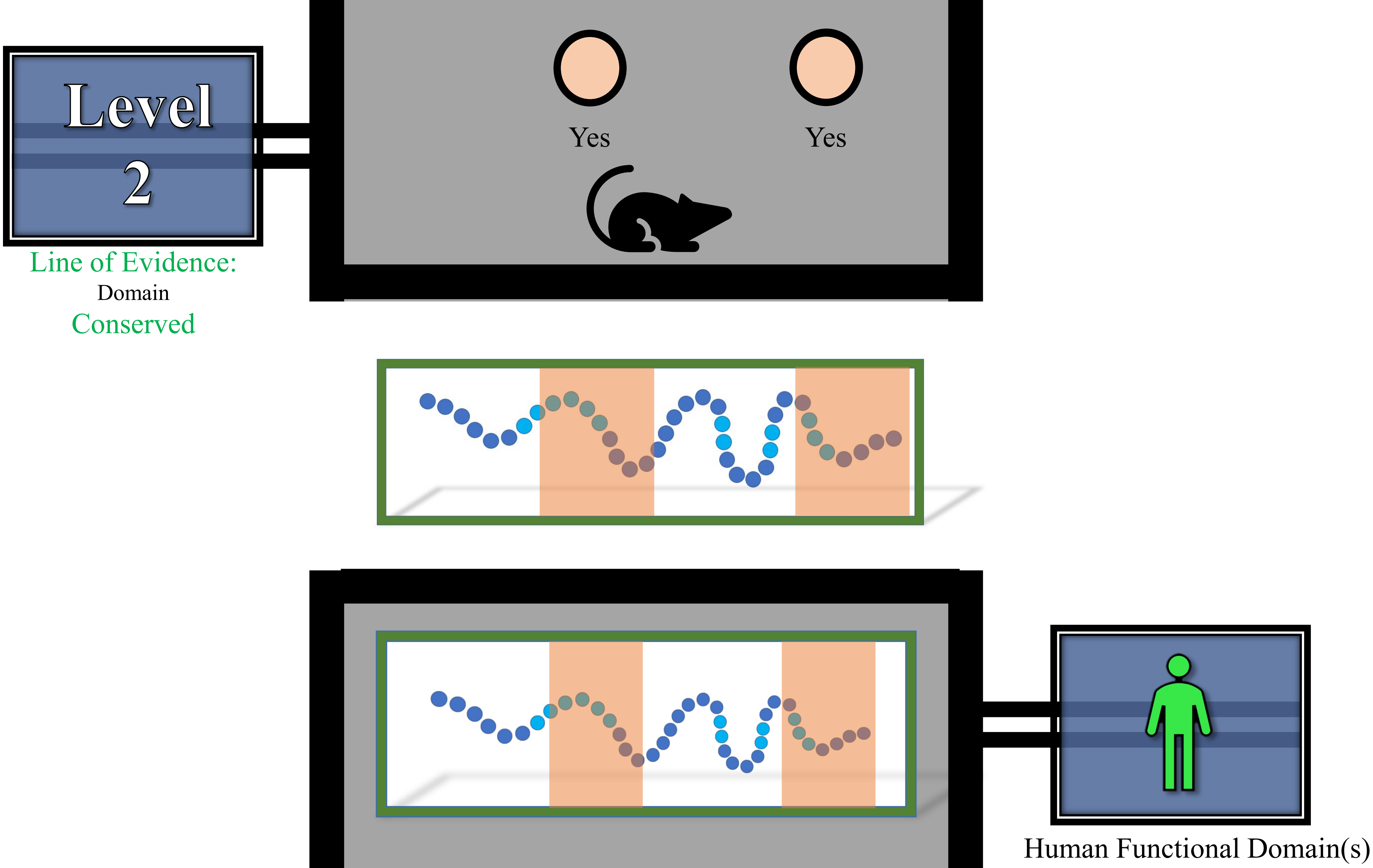
SeqAPASS Level 1



Human Protein Target

Line of Evidence:
Primary amino acid sequence
Conserved

Hundreds to Thousands of Species



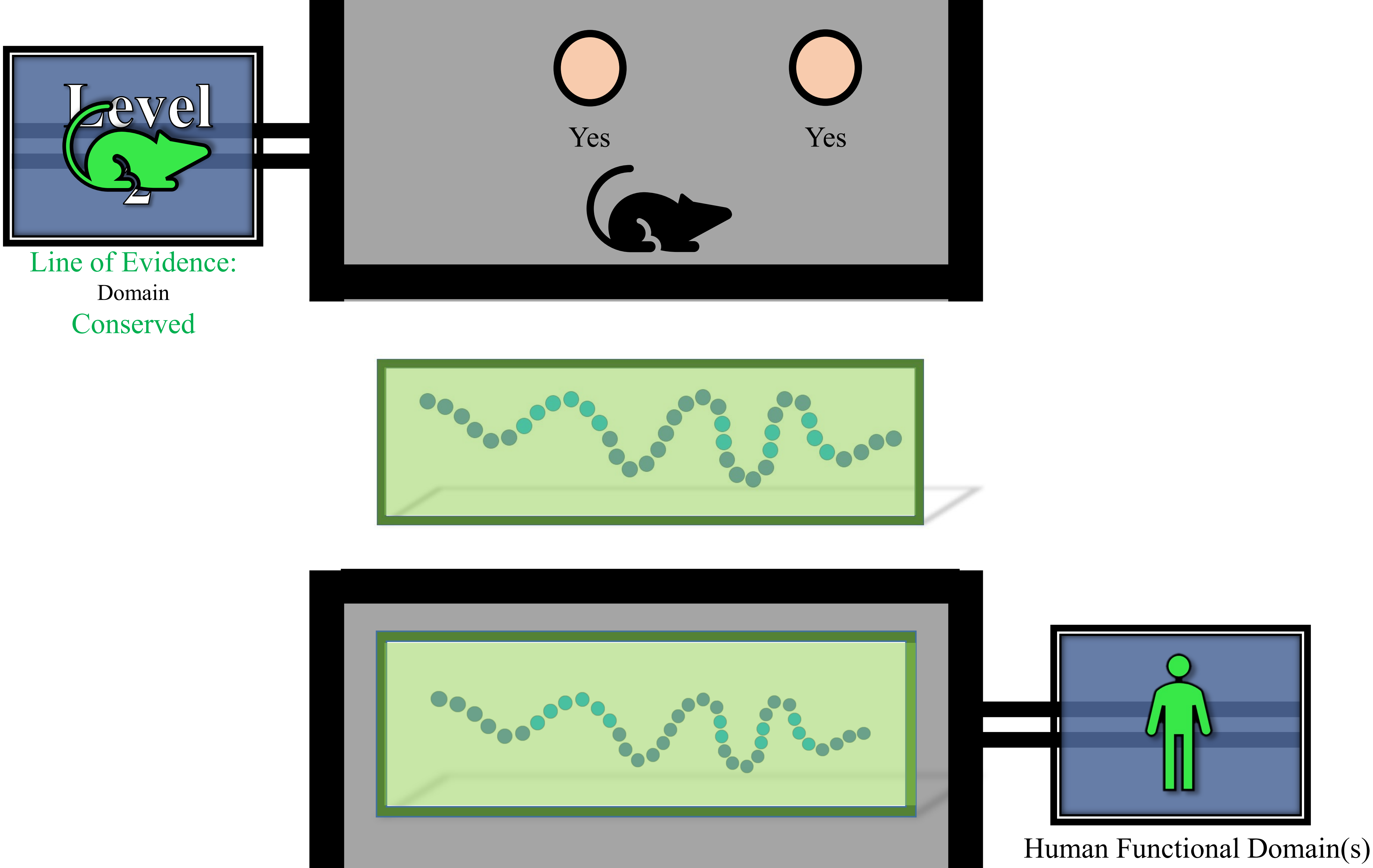
Level 2

Line of Evidence:
Domain
Conserved

Yes

Yes

Human Functional Domain(s)



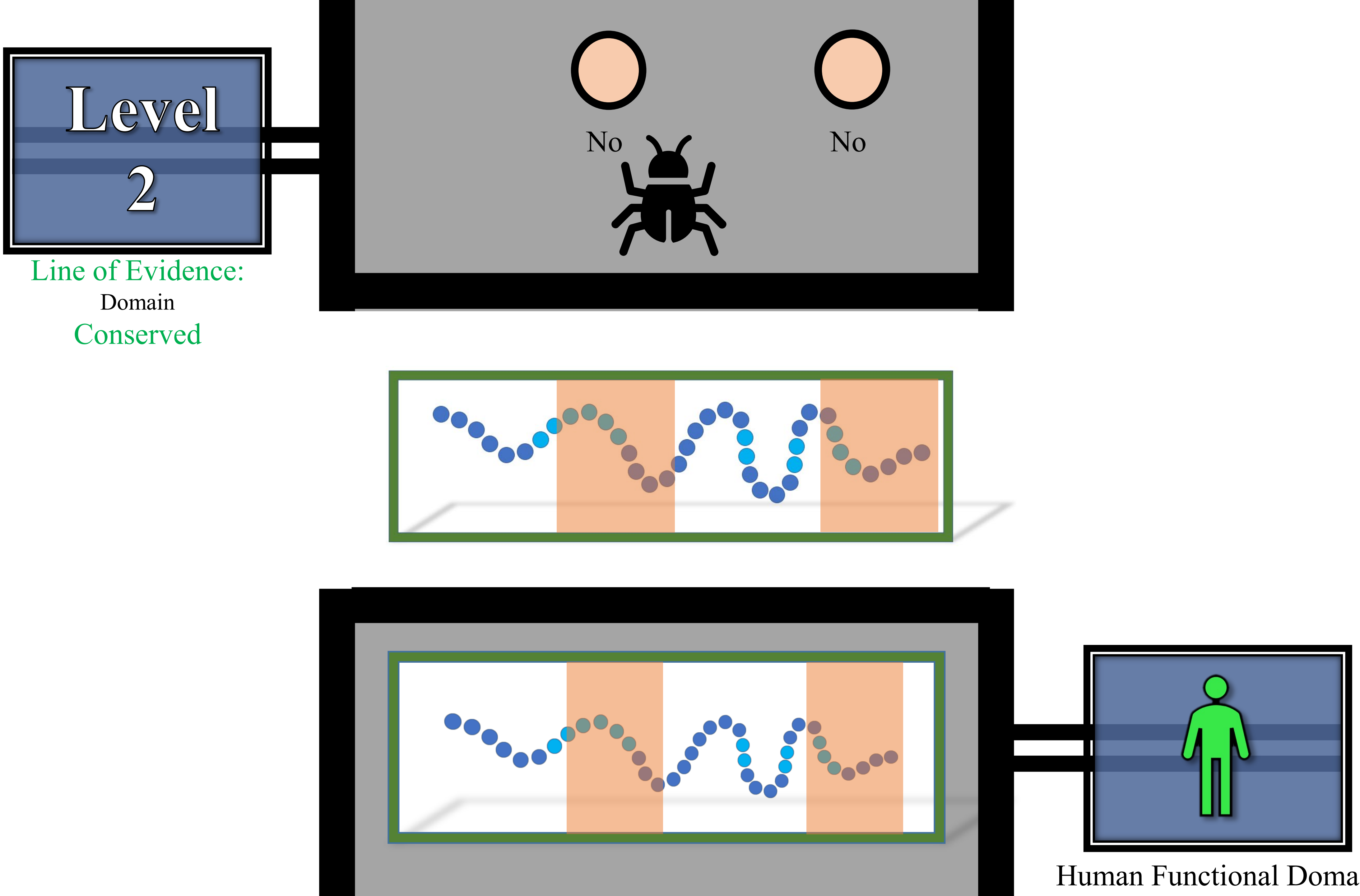
Level 2

Line of Evidence:
Domain
Conserved

Yes

Yes

Human Functional Domain(s)



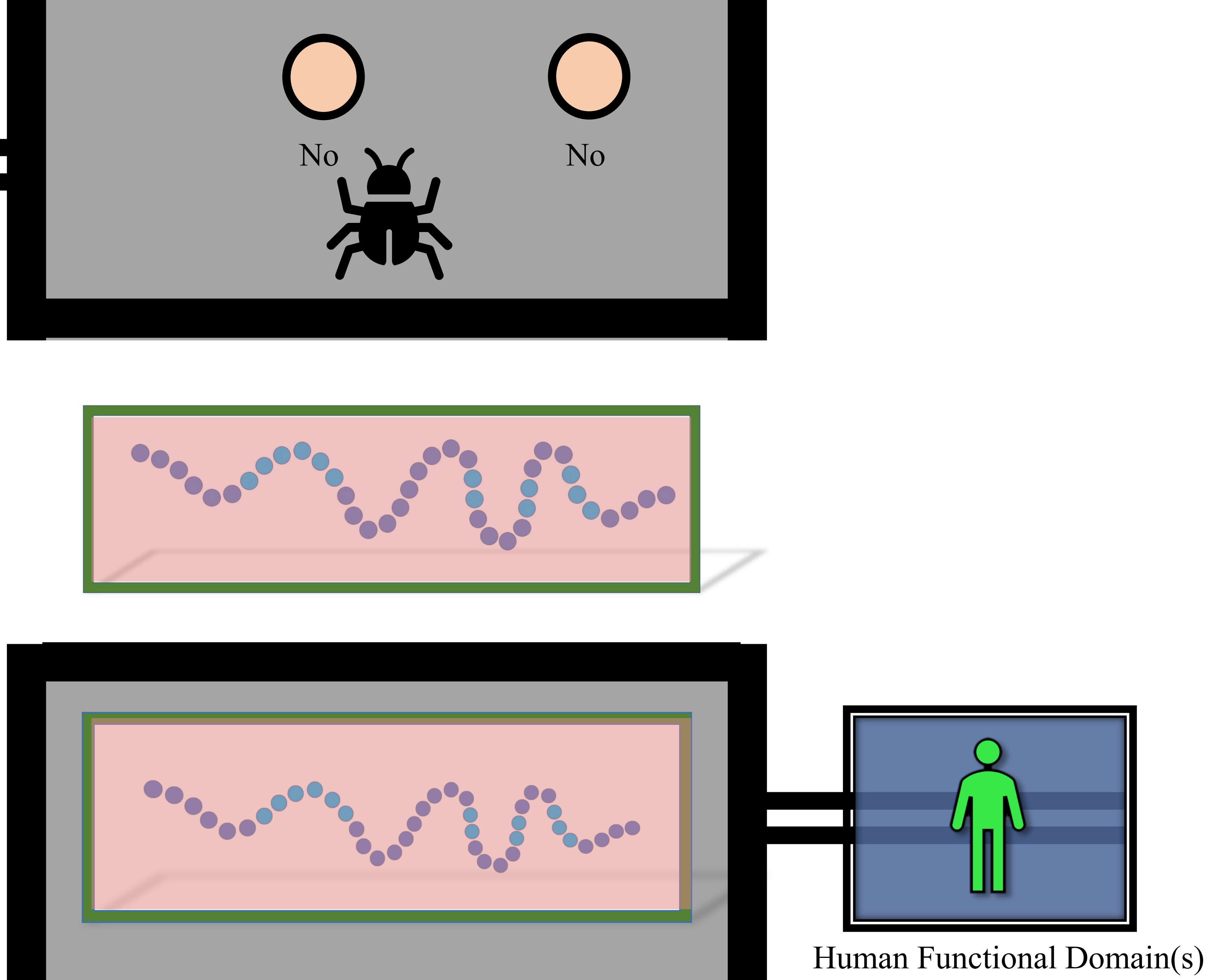
Level 2

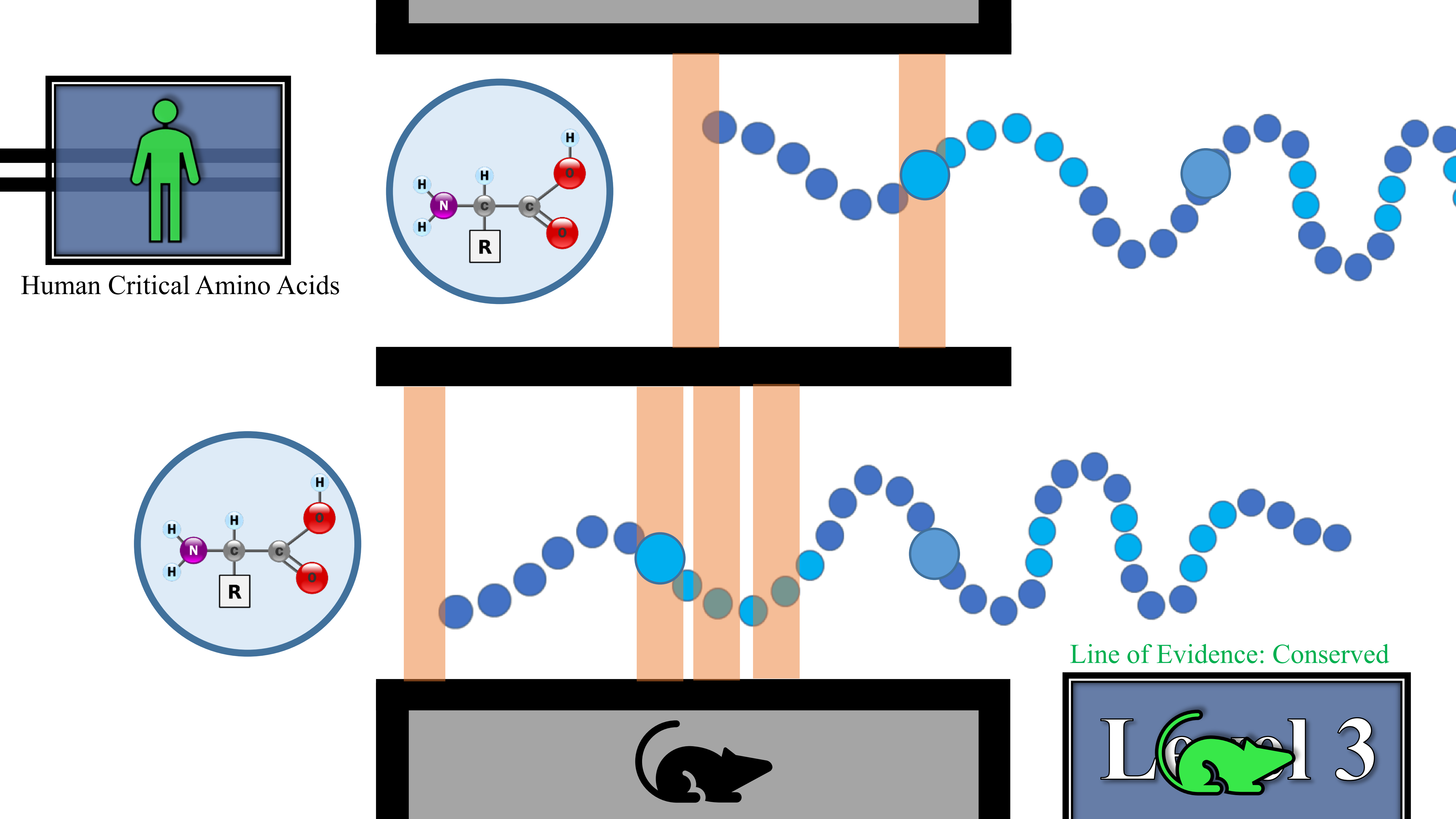
Line of Evidence:
Domain
Conserved

Human Functional Domain(s)

Level 2

Line of Evidence:
Domain
Not Conserved

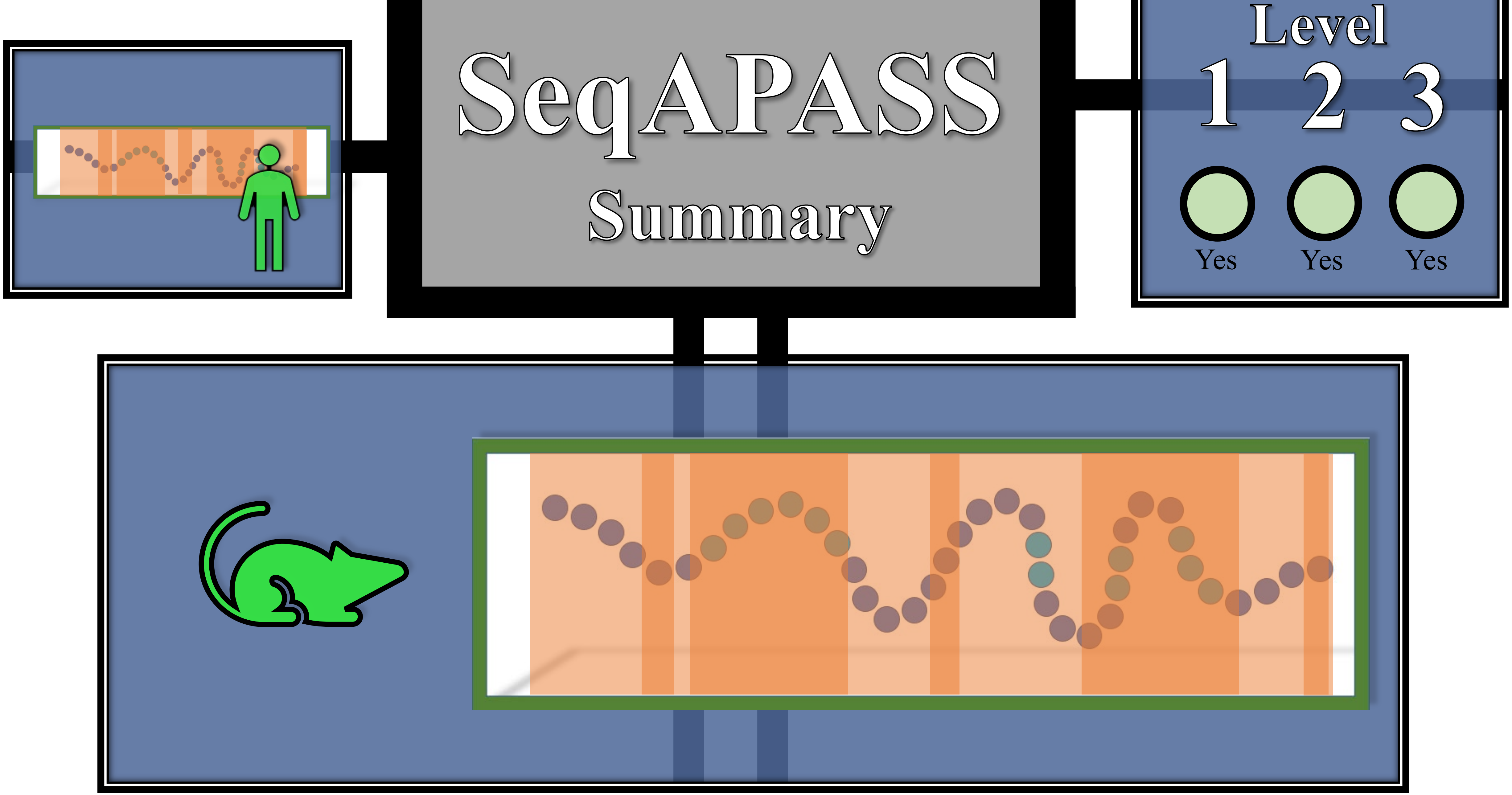




Human Critical Amino Acids

Line of Evidence: Conserved

Level 3



SeqAPASS

Summary

Level

1 2 3

Yes Yes Yes

Gather Lines of Evidence for Conservation of Protein Target:
Susceptibility Prediction: Yes or No

SeqAPASS DEMO:

Take a whirlwind tour of the tool and updated features

How to get started!

Process:

Define your question – problem formulation

Literature review to understand protein target and species

How well has the protein been characterized?

Do we know anything about functional domains?

Do we have any information about the chemical-protein interaction?

Use SeqAPASS to Guide you through the evaluation:

SeqAPASS Level 1 Query – view results

SeqAPASS Level 2 Query – view results

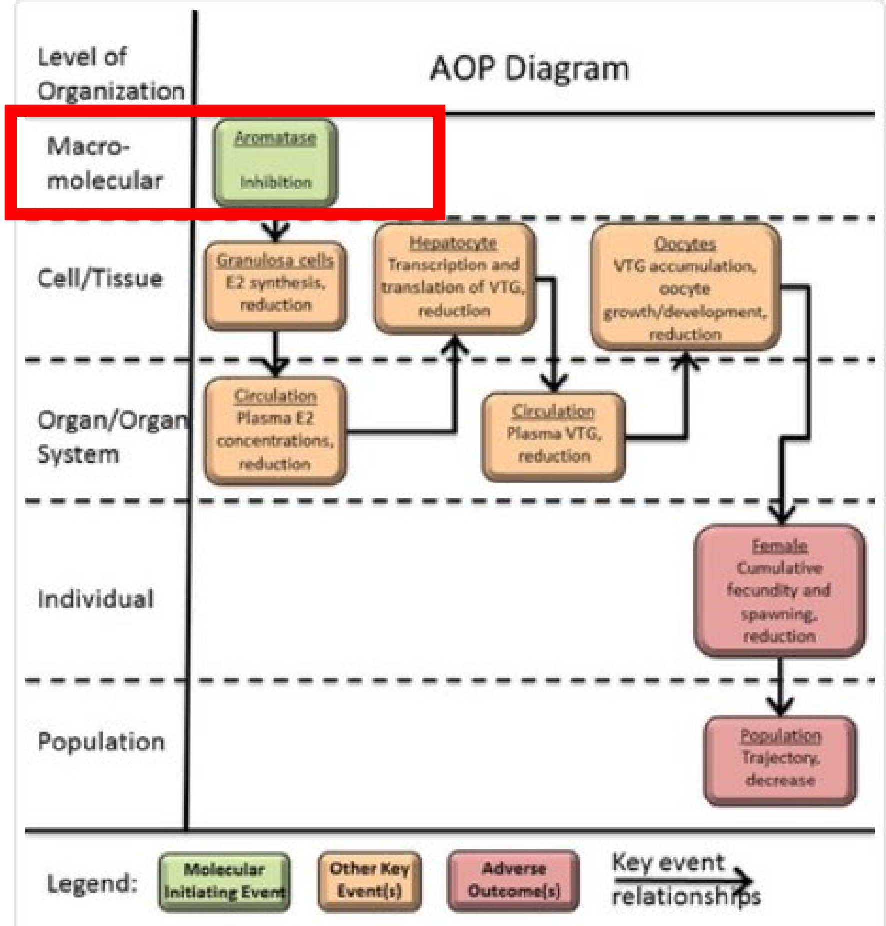
SeqAPASS Level 3 Query – view results

Address challenge in cross species extrapolation

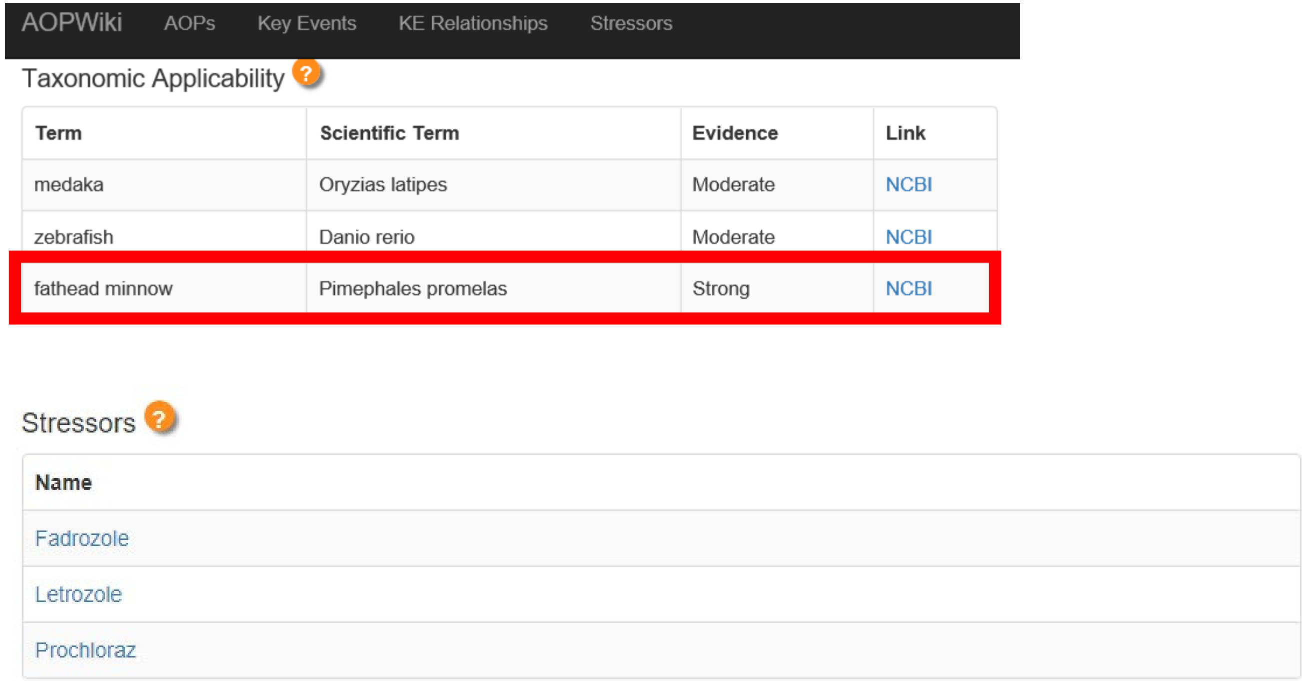
SeqAPASS was developed for you!

SeqAPASS Case Study

Case Study: SeqAPASS Evaluation to define the taxonomic relevance of an MIE in an AOP



AOP:25 Aromatase inhibition leading to reproductive dysfunction (https://aopwiki.org/aops/25)



Taxonomic Applicability

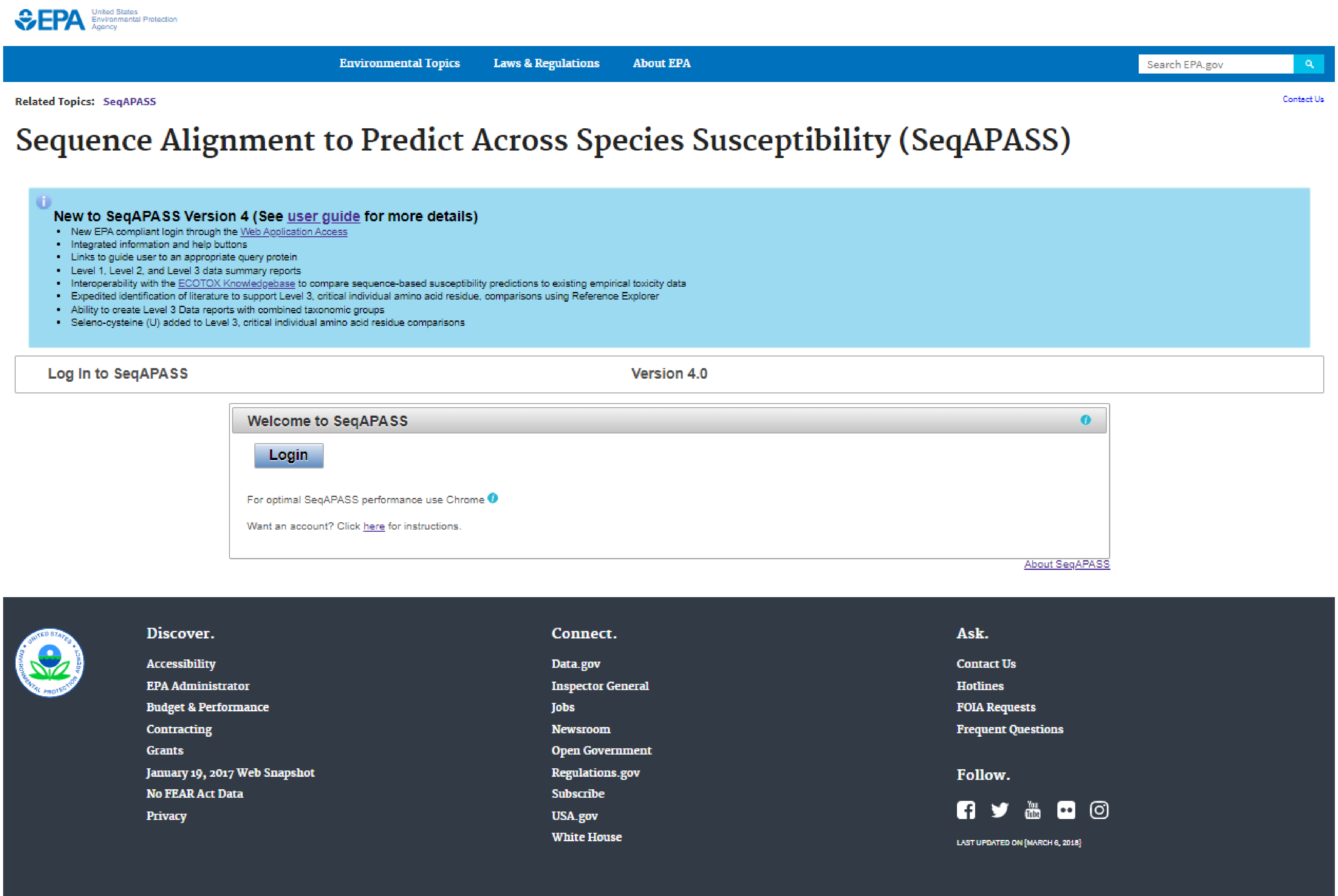
Term	Scientific Term	Evidence	Link
medaka	Oryzias latipes	Moderate	NCBI
zebrafish	Danio rerio	Moderate	NCBI
fathead minnow	Pimephales promelas	Strong	NCBI

Stressors

Name
Fadrozole
Letrozole
Prochloraz

Question(s): How broadly an I anticipate extrapolating this MIE?
OR from a chemical specific perspective,
What other species are likely susceptible to Fadrozole?

United States
Environmental Protection
Agency



Related Topics: SeqAPASS

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)

New to SeqAPASS Version 4 (See user guide for more details)

- New EPA compliant login through the Web Application Access
- Integrated information and help buttons
- Links to guide user to an appropriate query protein
- Level 1, Level 2, and Level 3 data summary reports
- Interoperability with the ECOTOX Knowledgebase to compare sequence-based susceptibility predictions to existing empirical toxicity data
- Expedited identification of literature to support Level 3, critical individual amino acid residue, comparisons using Reference Explorer
- Ability to create Level 3 Data reports with combined taxonomic groups
- Seleno-cysteine (U) added to Level 3, critical individual amino acid residue comparisons

Log In to SeqAPASS Version 4.0

Welcome to SeqAPASS

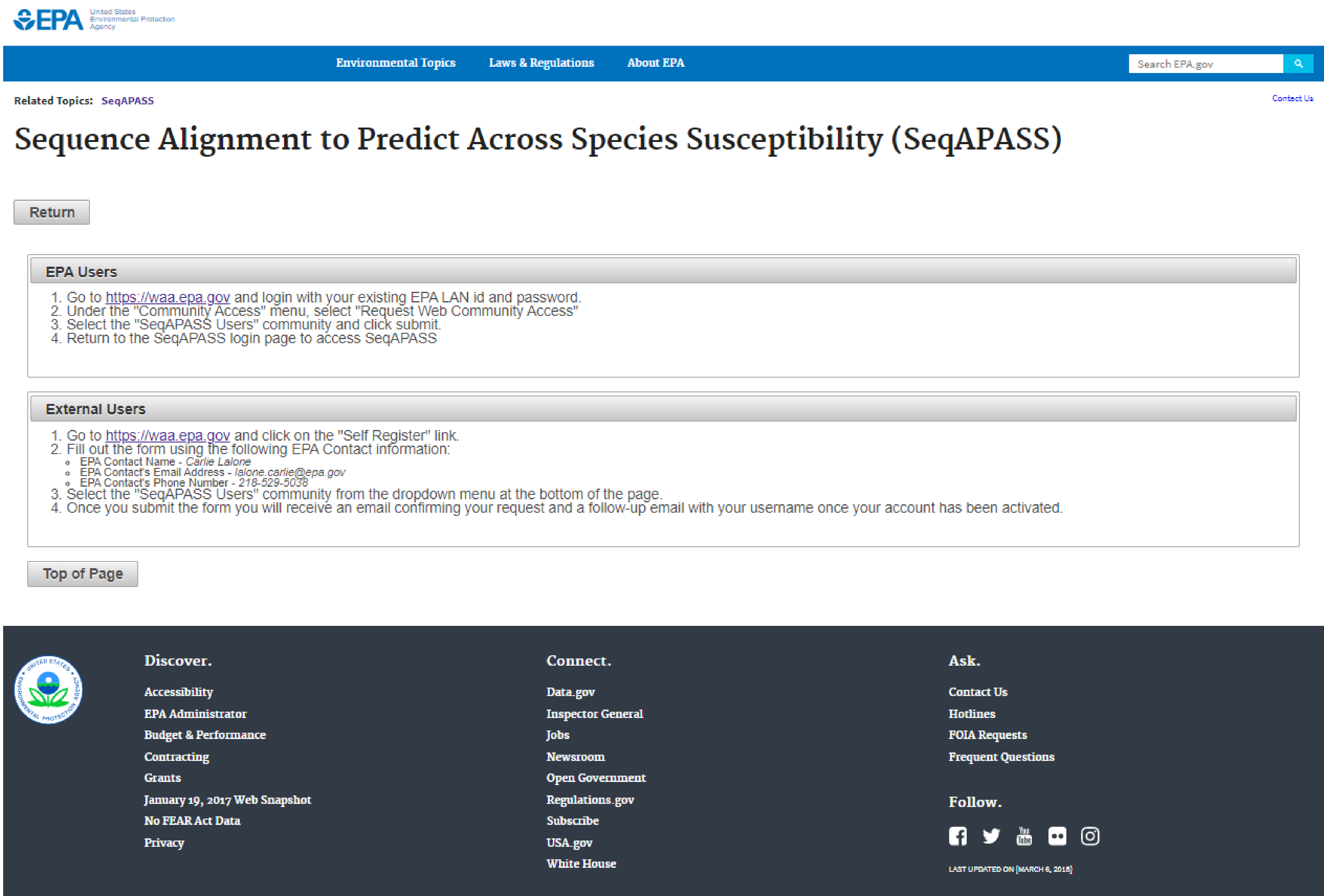
For optimal SeqAPASS performance use Chrome

Want an account? Click here for instructions.

About SeqAPASS

LAST UPDATED ON [MARCH 6, 2018]

EPA
United States
Environmental Protection
Agency



Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)

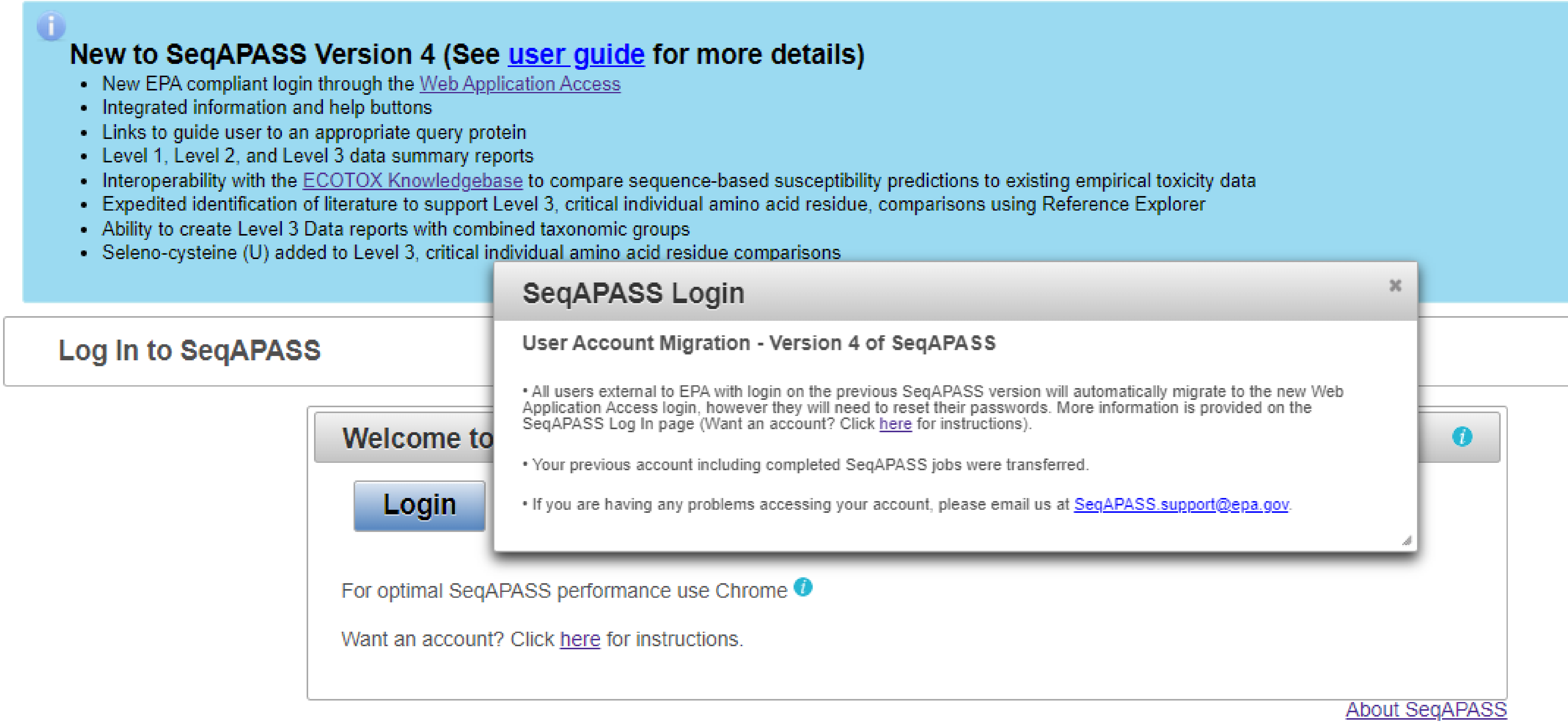
EPA Users

1. Go to https://waa.epa.gov and login with your existing EPA LAN id and password.
2. Under the "Community Access" menu, select "Request Web Community Access"
3. Select the "SeqAPASS Users" community and click submit.
4. Return to the SeqAPASS login page to access SeqAPASS

External Users

1. Go to https://waa.epa.gov and click on the "Self Register" link.
2. Fill out the form using the following EPA Contact information:
   - EPA Contact Name - *Carlie Lalone*
   - EPA Contact's Email Address - *lalone.carlie@epa.gov*
   - EPA Contact's Phone Number - *218-529-5038*
3. Select the "SeqAPASS Users" community from the dropdown menu at the bottom of the page.
4. Once you submit the form you will receive an email confirming your request and a follow-up email with your username once your account has been activated.

LAST UPDATED ON [MARCH 6, 2018]



New to SeqAPASS Version 4 (See user guide for more details)

- New EPA compliant login through the Web Application Access
- Integrated information and help buttons
- Links to guide user to an appropriate query protein
- Level 1, Level 2, and Level 3 data summary reports
- Interoperability with the ECOTOX Knowledgebase to compare sequence-based susceptibility predictions to existing empirical toxicity data
- Expedited identification of literature to support Level 3, critical individual amino acid residue, comparisons using Reference Explorer
- Ability to create Level 3 Data reports with combined taxonomic groups
- Seleno-cysteine (U) added to Level 3, critical individual amino acid residue comparisons

SeqAPASS Login

User Account Migration - Version 4 of SeqAPASS

• All users external to EPA with login on the previous SeqAPASS version will automatically migrate to the new Web Application Access login, however they will need to reset their passwords. More information is provided on the SeqAPASS Log In page (Want an account? Click here for instructions).

• Your previous account including completed SeqAPASS jobs were transferred.

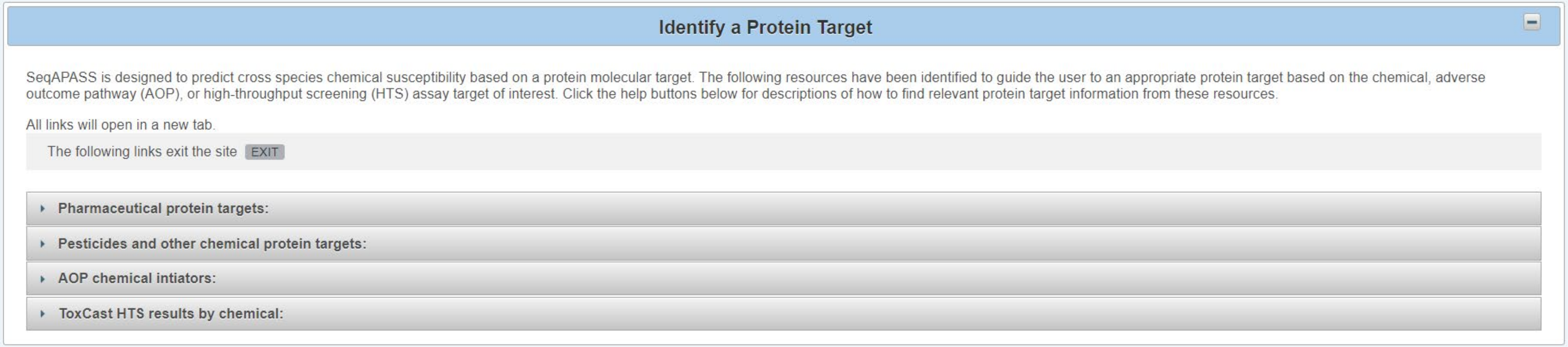
• If you are having any problems accessing your account, please email us at SeqAPASS.support@epa.gov.

Log In to SeqAPASS

Welcome to

For optimal SeqAPASS performance use Chrome

Want an account? Click here for instructions.



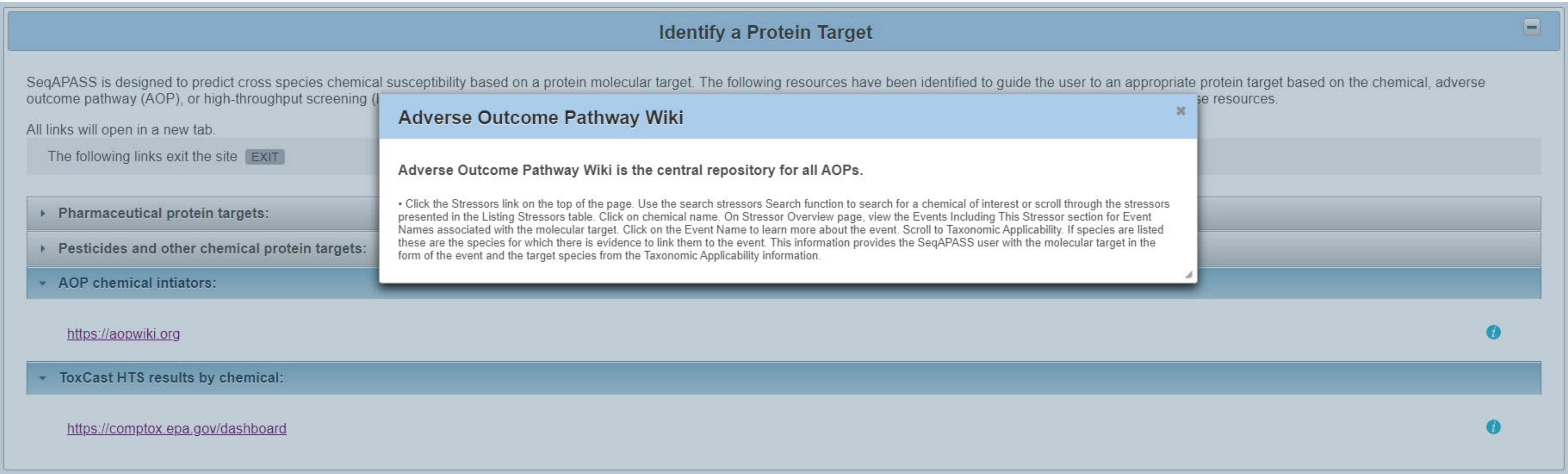
Identify a Protein Target

SeqAPASS is designed to predict cross species chemical susceptibility based on a protein molecular target. The following resources have been identified to guide the user to an appropriate protein target based on the chemical, adverse outcome pathway (AOP), or high-throughput screening (HTS) assay target of interest. Click the help buttons below for descriptions of how to find relevant protein target information from these resources.

All links will open in a new tab.

The following links exit the site EXIT

- **Pharmaceutical protein targets:**
- **Pesticides and other chemical protein targets:**
- **AOP chemical intiators:**
- **ToxCast HTS results by chemical:**



Identify a Protein Target

SeqAPASS is designed to predict cross species chemical susceptibility based on a protein molecular target. The following resources have been identified to guide the user to an appropriate protein target based on the chemical, adverse outcome pathway (AOP), or high-throughput screening (

All links will open in a new tab.

The following links exit the site EXIT

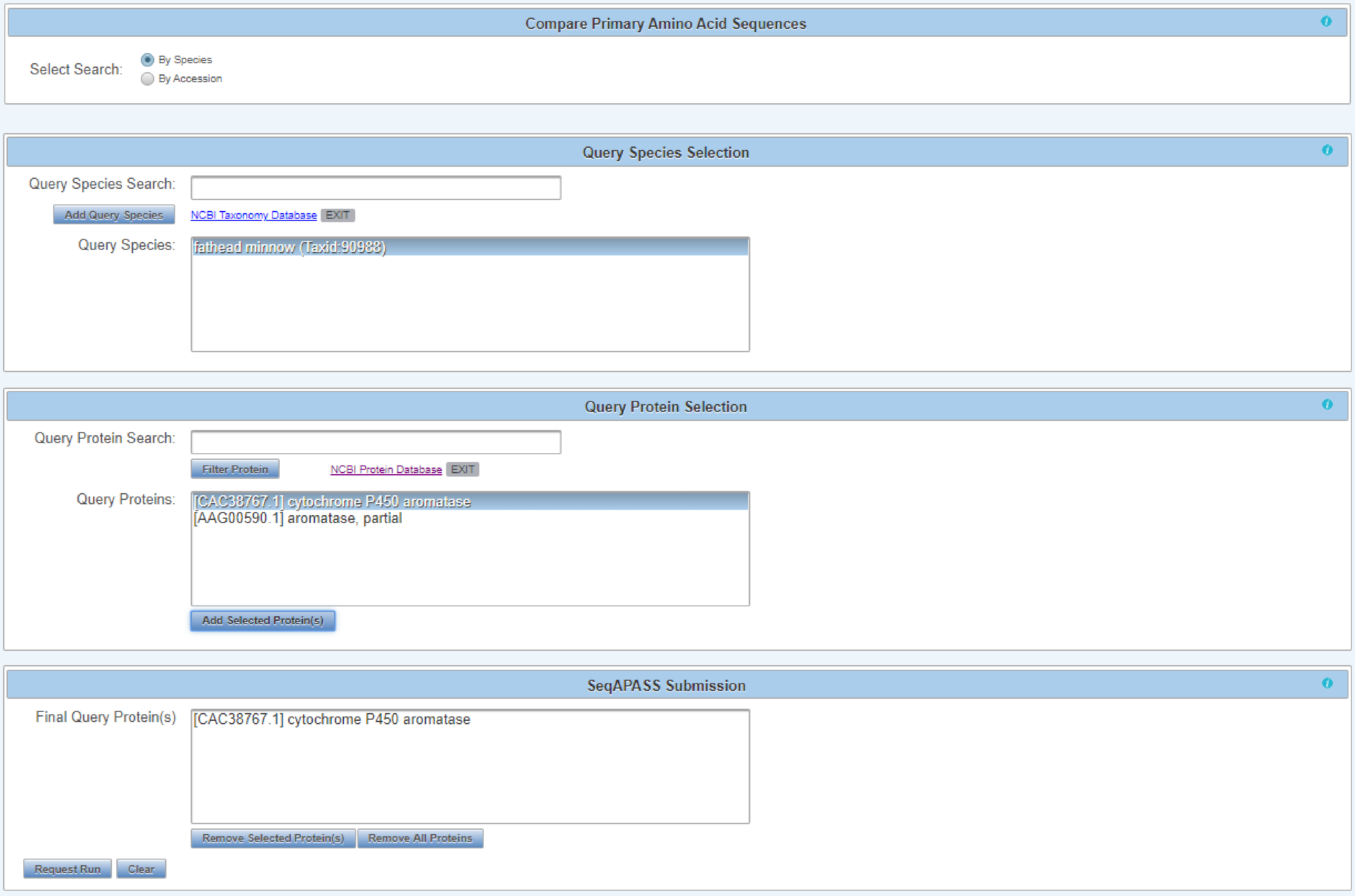
Adverse Outcome Pathway Wiki

Adverse Outcome Pathway Wiki is the central repository for all AOPs.

- Click the Stressors link on the top of the page. Use the search stressors Search function to search for a chemical of interest or scroll through the stressors presented in the Listing Stressors table. Click on chemical name. On Stressor Overview page, view the Events Including This Stressor section for Event Names associated with the molecular target. Click on the Event Name to learn more about the event. Scroll to Taxonomic Applicability. If species are listed these are the species for which there is evidence to link them to the event. This information provides the SeqAPASS user with the molecular target in the form of the event and the target species from the Taxonomic Applicability information.

- **Pharmaceutical protein targets:**
- **Pesticides and other chemical protein targets:**
- **AOP chemical intiators:**
  - https://aopwiki.org
- **ToxCast HTS results by chemical:**
  - https://comptox.epa.gov/dashboard

United States
Environmental Protection
Agency

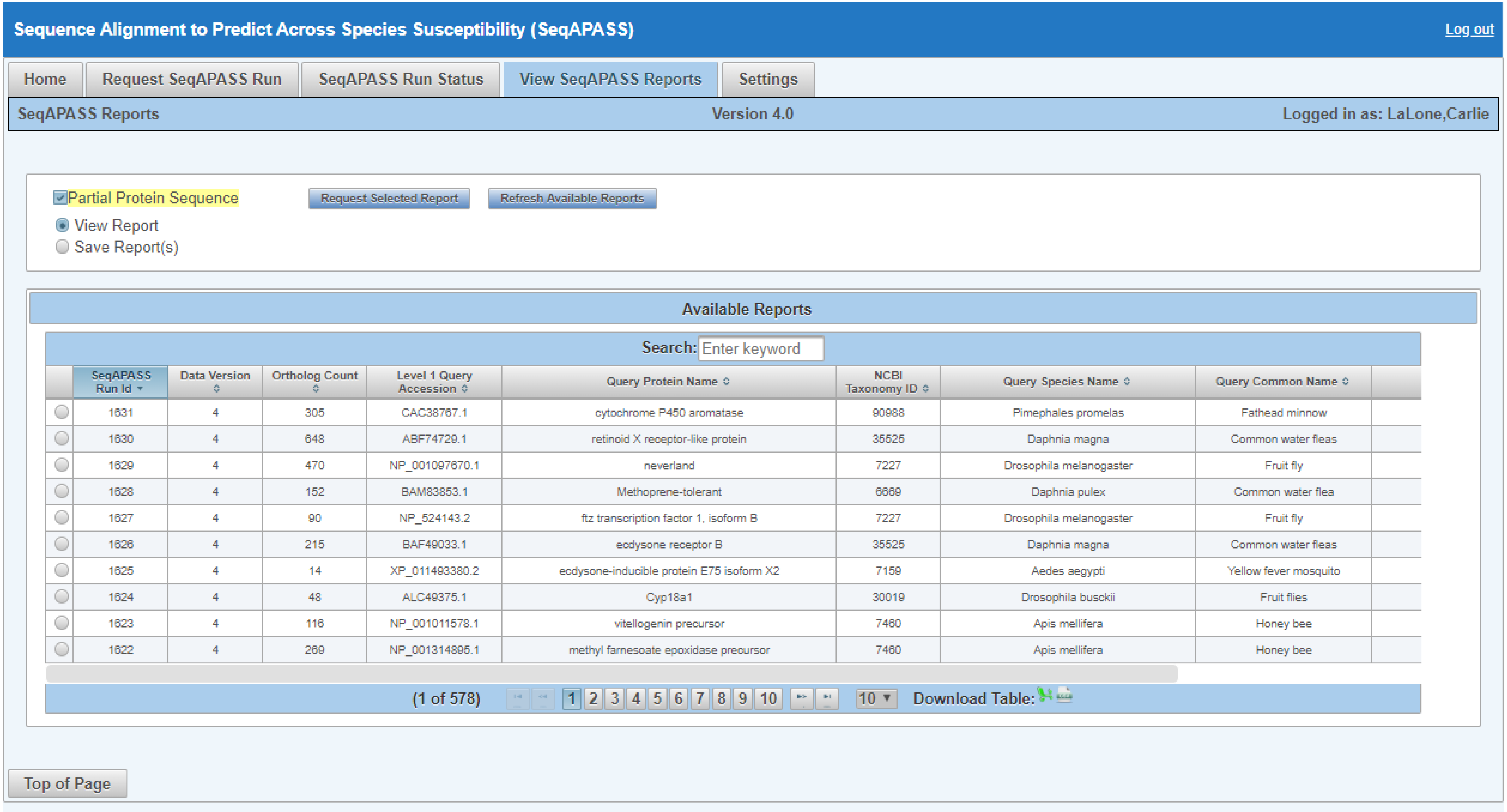


Select Search: ● By Species ○ By Accession

Query Species: fathead minnow (Taxid:90988)

Query Proteins:
[CAC38767.1] cytochrome P450 aromatase
[AAG00590.1] aromatase, partial

Final Query Protein(s): [CAC38767.1] cytochrome P450 aromatase



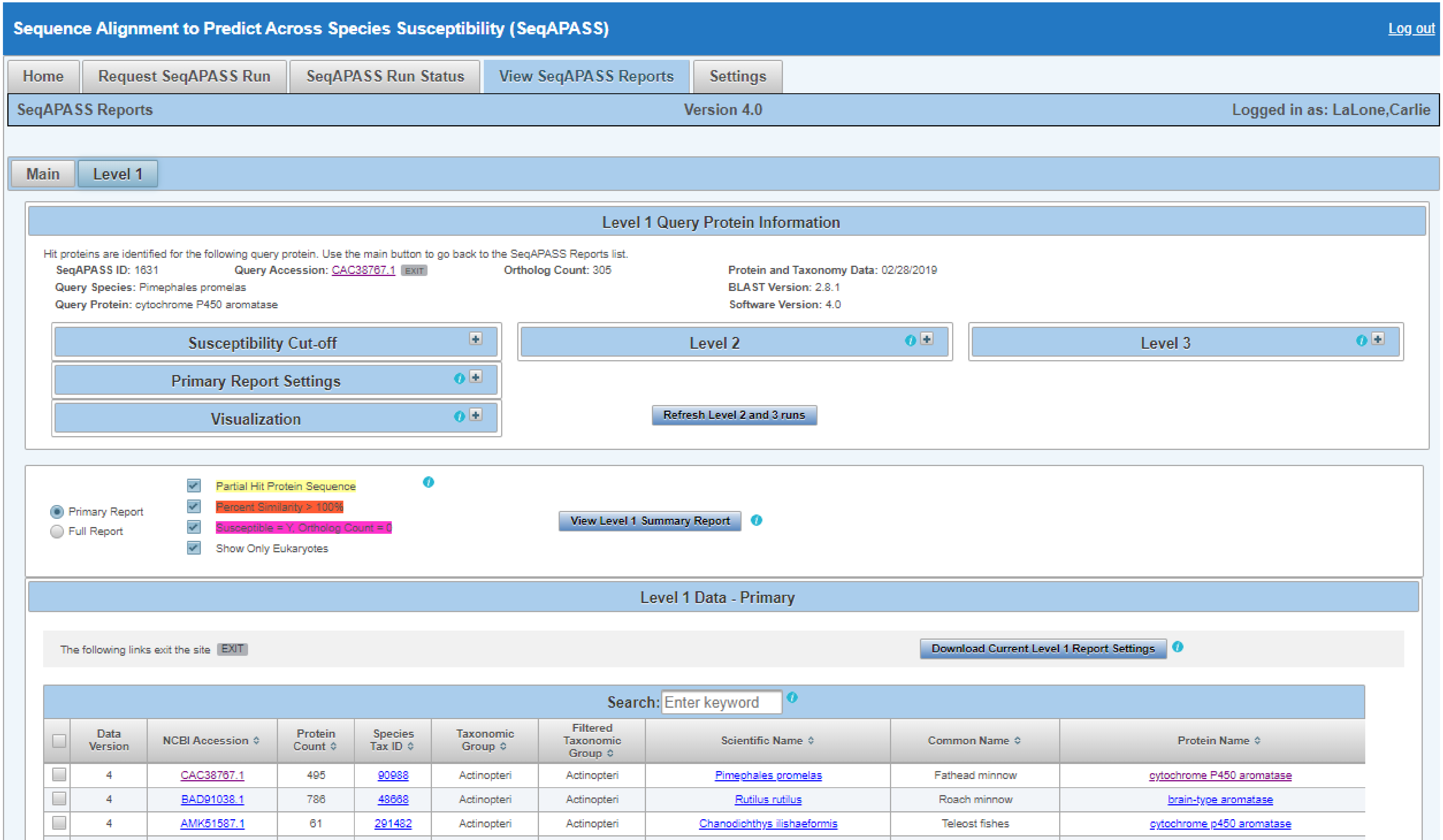
☑ Partial Protein Sequence

● View Report

○ Save Report(s)

	SeqAPASS Run Id	Data Version	Ortholog Count	Level 1 Query Accession	Query Protein Name	NCBI Taxonomy ID	Query Species Name	Query Common Name
○	1631	4	305	CAC38767.1	cytochrome P450 aromatase	90988	Pimephales promelas	Fathead minnow
○	1630	4	648	ABF74729.1	retinoid X receptor-like protein	35525	Daphnia magna	Common water fleas
○	1629	4	470	NP_001097670.1	neverland	7227	Drosophila melanogaster	Fruit fly
○	1628	4	152	BAM83853.1	Methoprene-tolerant	6669	Daphnia pulex	Common water flea
○	1627	4	90	NP_524143.2	ftz transcription factor 1, isoform B	7227	Drosophila melanogaster	Fruit fly
○	1626	4	215	BAF49033.1	ecdysone receptor B	35525	Daphnia magna	Common water fleas
○	1625	4	14	XP_011493380.2	ecdysone-inducible protein E75 isoform X2	7159	Aedes aegypti	Yellow fever mosquito
○	1624	4	48	ALC49375.1	Cyp18a1	30019	Drosophila busckii	Fruit flies
○	1623	4	116	NP_001011578.1	vitellogenin precursor	7460	Apis mellifera	Honey bee
○	1622	4	269	NP_001314895.1	methyl farnesoate epoxidase precursor	7460	Apis mellifera	Honey bee

(1 of 578) 1 2 3 4 5 6 7 8 9 10 | 10 ▼ | Download Table:



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: 1631
Query Accession: CAC38767.1 EXIT
Ortholog Count: 305
Protein and Taxonomy Data: 02/28/2019
Query Species: Pimephales promelas
BLAST Version: 2.8.1
Query Protein: cytochrome P450 aromatase
Software Version: 4.0

● Primary Report
○ Full Report

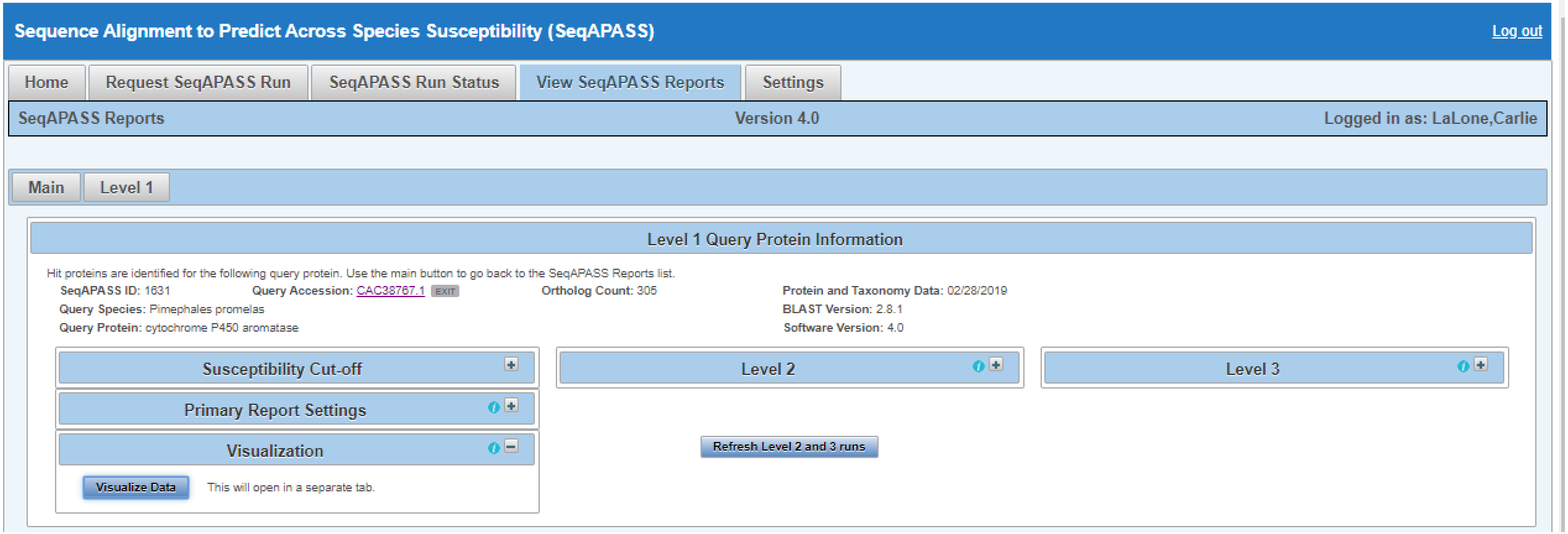
☑ Partial Hit Protein Sequence
☑ Percent Similarity > 100%
☑ Susceptible = Y, Ortholog Count = 0
☑ Show Only Eukaryotes

Level 1 Data - Primary

The following links exit the site EXIT

☐	Data Version	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Filtered Taxonomic Group	Scientific Name	Common Name	Protein Name
☐	4	CAC38767.1	495	90988	Actinopteri	Actinopteri	Pimephales promelas	Fathead minnow	cytochrome P450 aromatase
☐	4	BAD91038.1	786	48668	Actinopteri	Actinopteri	Rutilus rutilus	Roach minnow	brain-type aromatase
☐	4	AMK51587.1	61	291482	Actinopteri	Actinopteri	Chanodichthys ilishaeformis	Teleost fishes	cytochrome p450 aromatase

United States
Environmental Protection
Agency

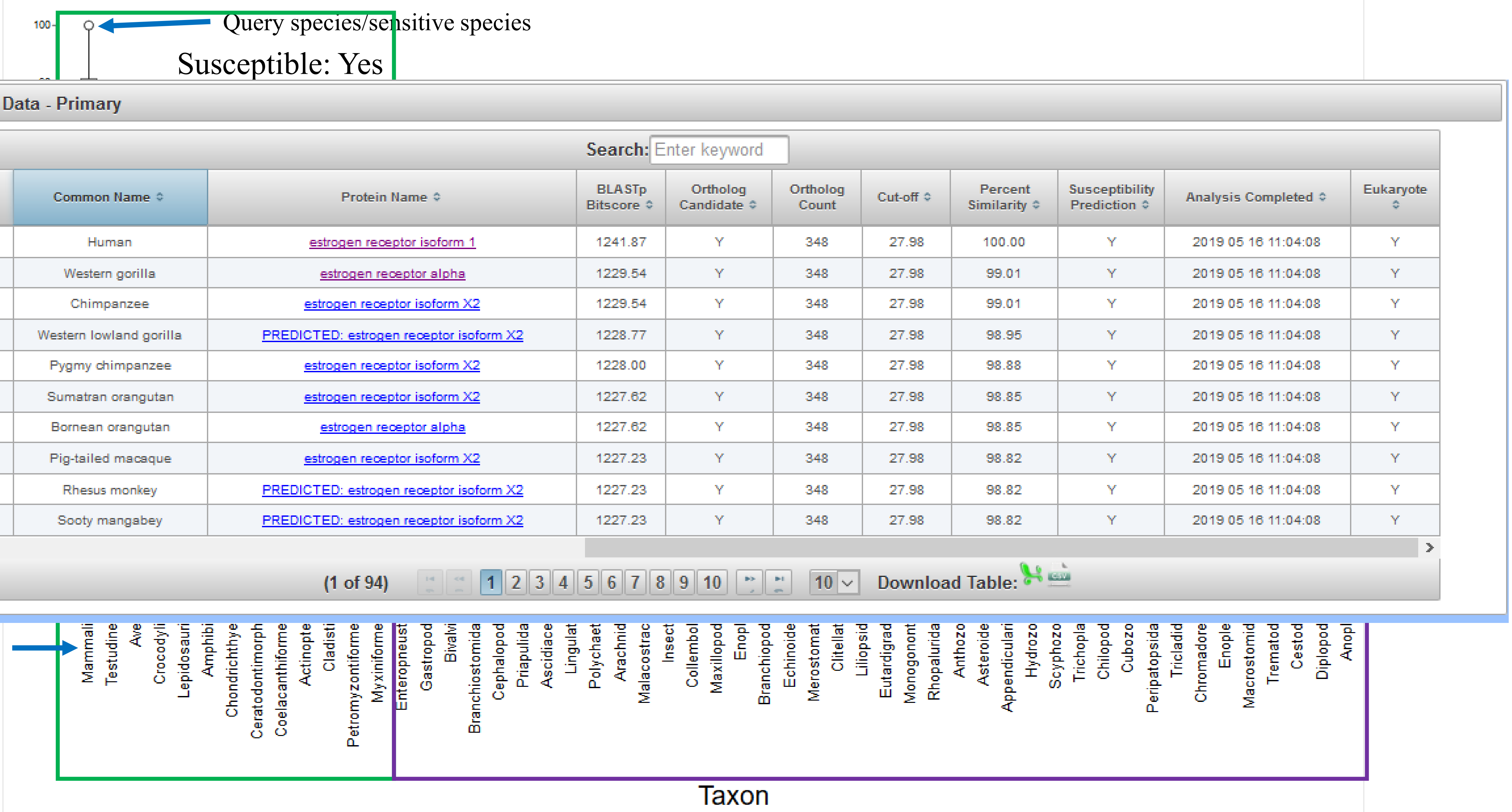


Hit proteins are identified for the following query protein. Use the main button to go back to the SeqAPASS Reports list.

SeqAPASS ID: 1631
Query Accession: CAC38767.1 EXIT
Ortholog Count: 305
Protein and Taxonomy Data: 02/28/2019
Query Species: Pimephales promelas
BLAST Version: 2.8.1
Query Protein: cytochrome P450 aromatase
Software Version: 4.0

This will open in a separate tab.

Interactive Data Visualization



Query species/sensitive species

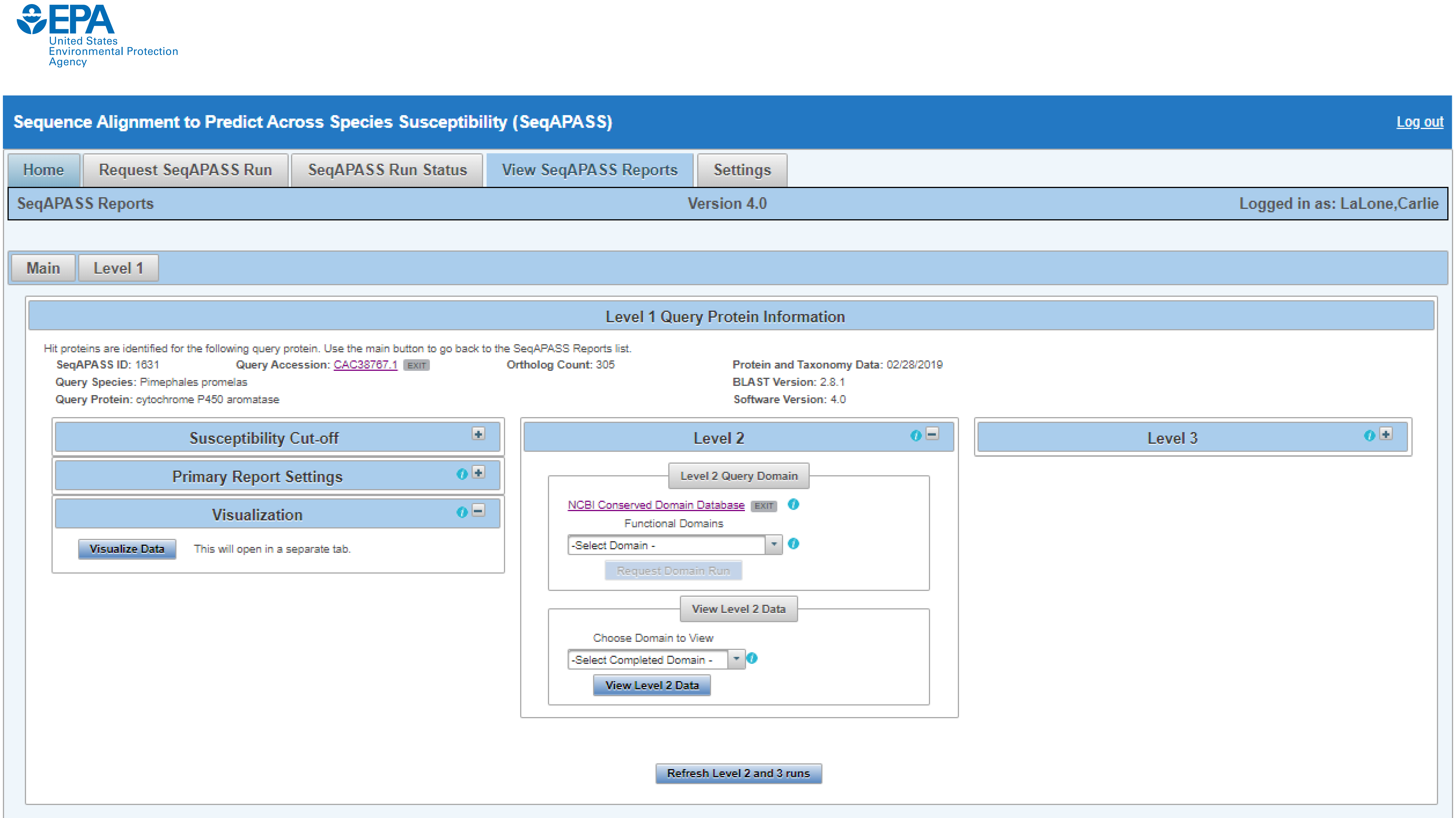
Susceptible: Yes

Taxonomic group: Class

Level 1 Data - Primary

Common Name ⇕	Protein Name ⇕	BLASTp Bitscore ⇕	Ortholog Candidate ⇕	Ortholog Count	Cut-off ⇕	Percent Similarity ⇕	Susceptibility Prediction ⇕	Analysis Completed ⇕	Eukaryote ⇕
Human	estrogen receptor isoform 1	1241.87	Y	348	27.98	100.00	Y	2019 05 16 11:04:08	Y
Western gorilla	estrogen receptor alpha	1229.54	Y	348	27.98	99.01	Y	2019 05 16 11:04:08	Y
Chimpanzee	estrogen receptor isoform X2	1229.54	Y	348	27.98	99.01	Y	2019 05 16 11:04:08	Y
Western lowland gorilla	PREDICTED: estrogen receptor isoform X2	1228.77	Y	348	27.98	98.95	Y	2019 05 16 11:04:08	Y
Pygmy chimpanzee	estrogen receptor isoform X2	1228.00	Y	348	27.98	98.88	Y	2019 05 16 11:04:08	Y
Sumatran orangutan	estrogen receptor isoform X2	1227.62	Y	348	27.98	98.85	Y	2019 05 16 11:04:08	Y
Bornean orangutan	estrogen receptor alpha	1227.62	Y	348	27.98	98.85	Y	2019 05 16 11:04:08	Y
Pig-tailed macaque	estrogen receptor isoform X2	1227.23	Y	348	27.98	98.82	Y	2019 05 16 11:04:08	Y
Rhesus monkey	PREDICTED: estrogen receptor isoform X2	1227.23	Y	348	27.98	98.82	Y	2019 05 16 11:04:08	Y
Sooty mangabey	PREDICTED: estrogen receptor isoform X2	1227.23	Y	348	27.98	98.82	Y	2019 05 16 11:04:08	Y

(1 of 94) 1 2 3 4 5 6 7 8 9 10



Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)

SeqAPASS Reports

Version 4.0

Logged in as: LaLone,Carlie

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: 1631

Query Accession: CAC38767.1

Ortholog Count: 305

Protein and Taxonomy Data: 02/28/2019

Query Species: Pimephales promelas

BLAST Version: 2.8.1

Query Protein: cytochrome P450 aromatase

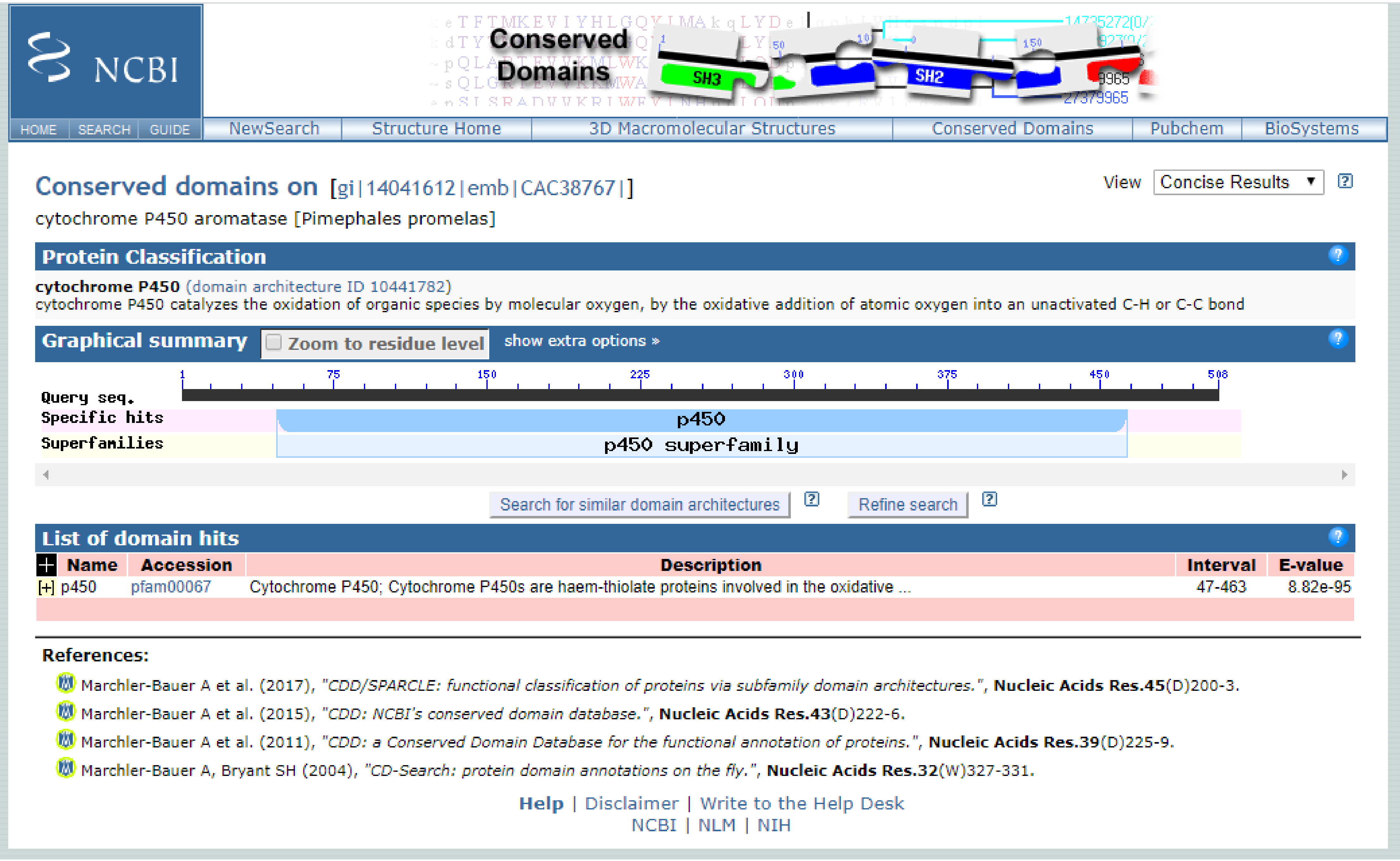
Software Version: 4.0

This will open in a separate tab.

NCBI Conserved Domain Database

Functional Domains

Choose Domain to View



Conserved domains on [gi|14041612|emb|CAC38767|]

cytochrome P450 aromatase [Pimephales promelas]

Protein Classification

cytochrome P450 (domain architecture ID 10441782)
cytochrome P450 catalyzes the oxidation of organic species by molecular oxygen, by the oxidative addition of atomic oxygen into an unactivated C-H or C-C bond

Graphical summary

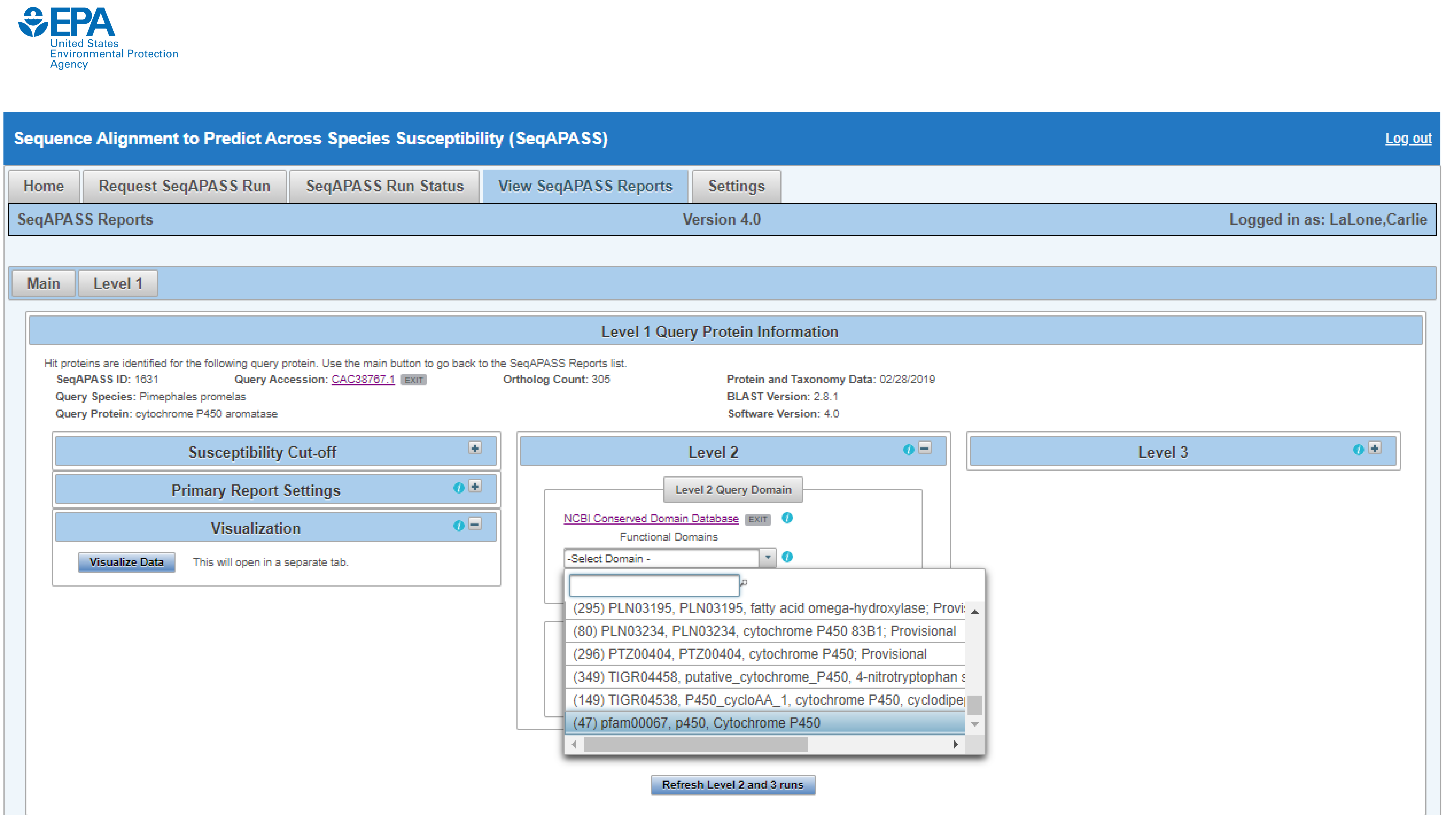
List of domain hits

Name	Accession	Description	Interval	E-value
[+] p450	pfam00067	Cytochrome P450; Cytochrome P450s are haem-thiolate proteins involved in the oxidative ...	47-463	8.82e-95

References:

- Marchler-Bauer A et al. (2017), *"CDD/SPARCLE: functional classification of proteins via subfamily domain architectures."*, **Nucleic Acids Res.45**(D)200-3.
- Marchler-Bauer A et al. (2015), *"CDD: NCBI's conserved domain database."*, **Nucleic Acids Res.43**(D)222-6.
- Marchler-Bauer A et al. (2011), *"CDD: a Conserved Domain Database for the functional annotation of proteins."*, **Nucleic Acids Res.39**(D)225-9.
- Marchler-Bauer A, Bryant SH (2004), *"CD-Search: protein domain annotations on the fly."*, **Nucleic Acids Res.32**(W)327-331.

Help | Disclaimer | Write to the Help Desk
NCBI | NLM | NIH



Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)

SeqAPASS Reports | Version 4.0 | Logged in as: LaLone,Carlie

Level 1 Query Protein Information

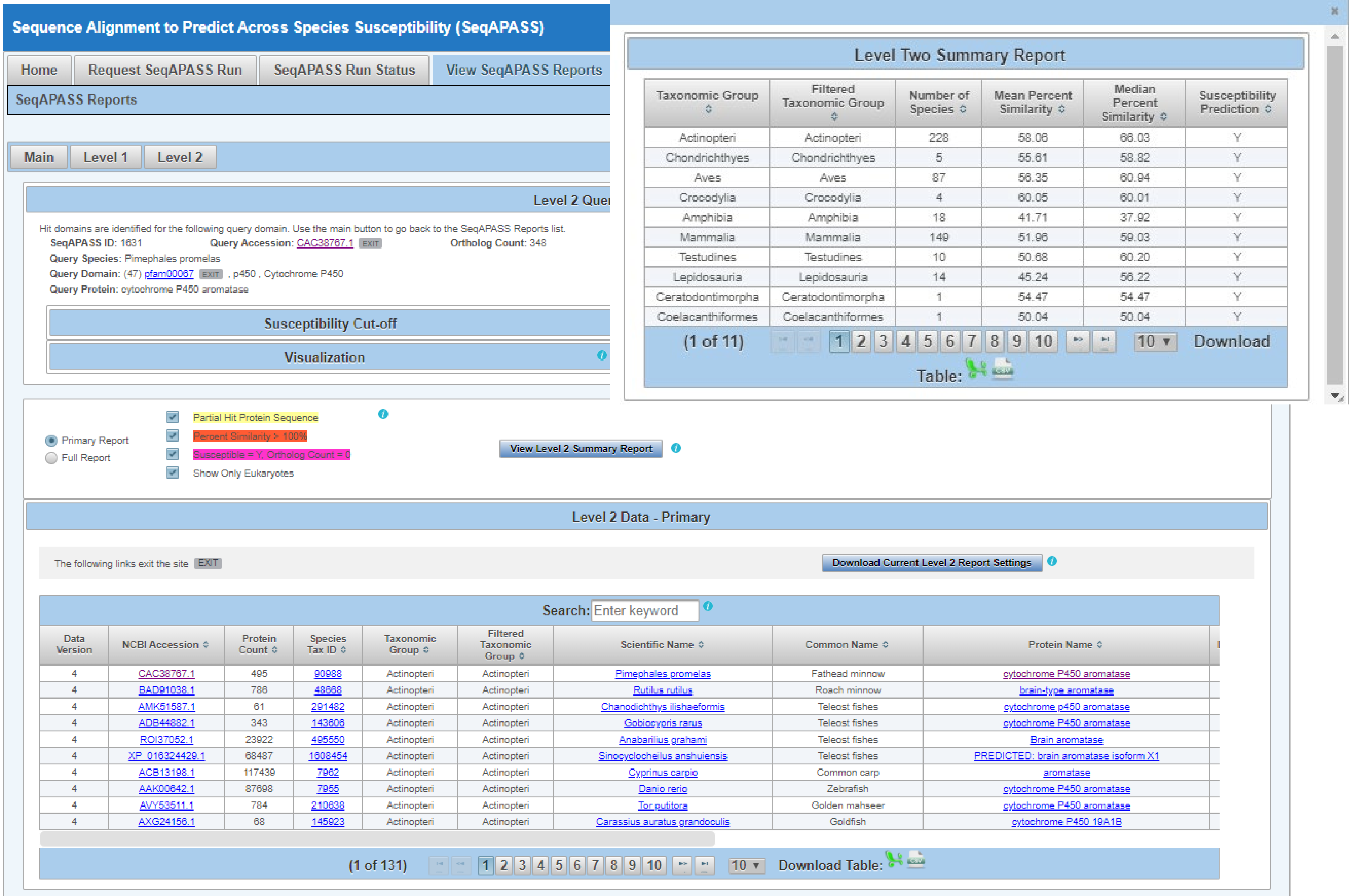
Hit proteins are identified for the following query protein. Use the main button to go back to the SeqAPASS Reports list.

SeqAPASS ID: 1631
Query Accession: CAC38767.1 EXIT
Ortholog Count: 305
Protein and Taxonomy Data: 02/28/2019
Query Species: Pimephales promelas
BLAST Version: 2.8.1
Query Protein: cytochrome P450 aromatase
Software Version: 4.0

This will open in a separate tab.

(295) PLN03195, PLN03195, fatty acid omega-hydroxylase; Provi
(80) PLN03234, PLN03234, cytochrome P450 83B1; Provisional
(296) PTZ00404, PTZ00404, cytochrome P450; Provisional
(349) TIGR04458, putative_cytochrome_P450, 4-nitrotryptophan s
(149) TIGR04538, P450_cycloAA_1, cytochrome P450, cyclodipe
(47) pfam00067, p450, Cytochrome P450

EPA United States Environmental Protection Agency



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

SeqAPASS ID: 1631 **Query Accession:** CAC38767.1 EXIT **Ortholog Count:** 348

Query Species: Pimephales promelas

Query Domain: (47) pfam00067 EXIT , p450 , Cytochrome P450

Query Protein: cytochrome P450 aromatase

● Primary Report
○ Full Report

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

Level Two Summary Report

Taxonomic Group	Filtered Taxonomic Group	Number of Species	Mean Percent Similarity	Median Percent Similarity	Susceptibility Prediction
Actinopteri	Actinopteri	228	58.06	66.03	Y
Chondrichthyes	Chondrichthyes	5	55.61	58.82	Y
Aves	Aves	87	56.35	60.94	Y
Crocodylia	Crocodylia	4	60.05	60.01	Y
Amphibia	Amphibia	18	41.71	37.92	Y
Mammalia	Mammalia	149	51.96	59.03	Y
Testudines	Testudines	10	50.68	60.20	Y
Lepidosauria	Lepidosauria	14	45.24	56.22	Y
Ceratodontimorpha	Ceratodontimorpha	1	54.47	54.47	Y
Coelacanthiformes	Coelacanthiformes	1	50.04	50.04	Y

(1 of 11) 1 2 3 4 5 6 7 8 9 10 · 10 ▼ Download Table:

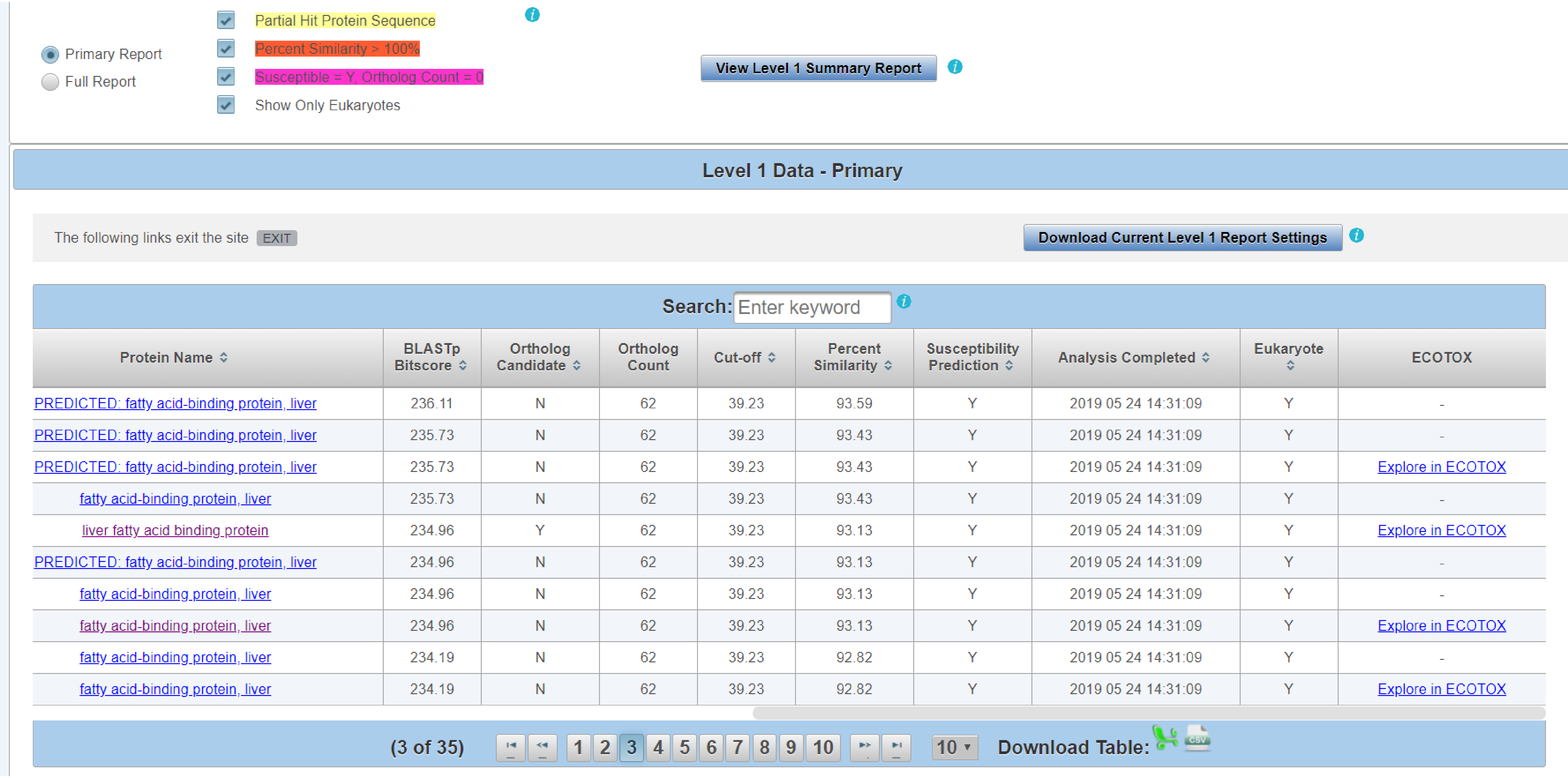
Level 2 Data - Primary

The following links exit the site

Search: Enter keyword

Data Version	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Filtered Taxonomic Group	Scientific Name	Common Name	Protein Name
4	CAC38767.1	495	90988	Actinopteri	Actinopteri	Pimephales promelas	Fathead minnow	cytochrome P450 aromatase
4	BAD91038.1	786	48668	Actinopteri	Actinopteri	Rutilus rutilus	Roach minnow	brain-type aromatase
4	AMK51587.1	61	291482	Actinopteri	Actinopteri	Chanodichthys ilishaeformis	Teleost fishes	cytochrome p450 aromatase
4	ADB44882.1	343	143606	Actinopteri	Actinopteri	Gobiocypris rarus	Teleost fishes	cytochrome P450 aromatase
4	ROI37052.1	23922	495550	Actinopteri	Actinopteri	Anabarilius grahami	Teleost fishes	Brain aromatase
4	XP_016324429.1	68487	1608454	Actinopteri	Actinopteri	Sinocyclocheilus anshuiensis	Teleost fishes	PREDICTED: brain aromatase isoform X1
4	ACB13198.1	117439	7962	Actinopteri	Actinopteri	Cyprinus carpio	Common carp	aromatase
4	AAK00642.1	87698	7955	Actinopteri	Actinopteri	Danio rerio	Zebrafish	cytochrome P450 aromatase
4	AVY53511.1	784	210638	Actinopteri	Actinopteri	Tor putitora	Golden mahseer	cytochrome P450 aromatase
4	AXG24156.1	68	145923	Actinopteri	Actinopteri	Carassius auratus grandoculis	Goldfish	cytochrome P450 19A1B

(1 of 131) 1 2 3 4 5 6 7 8 9 10 · 10 ▼ Download Table:



● Primary Report
○ Full Report

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

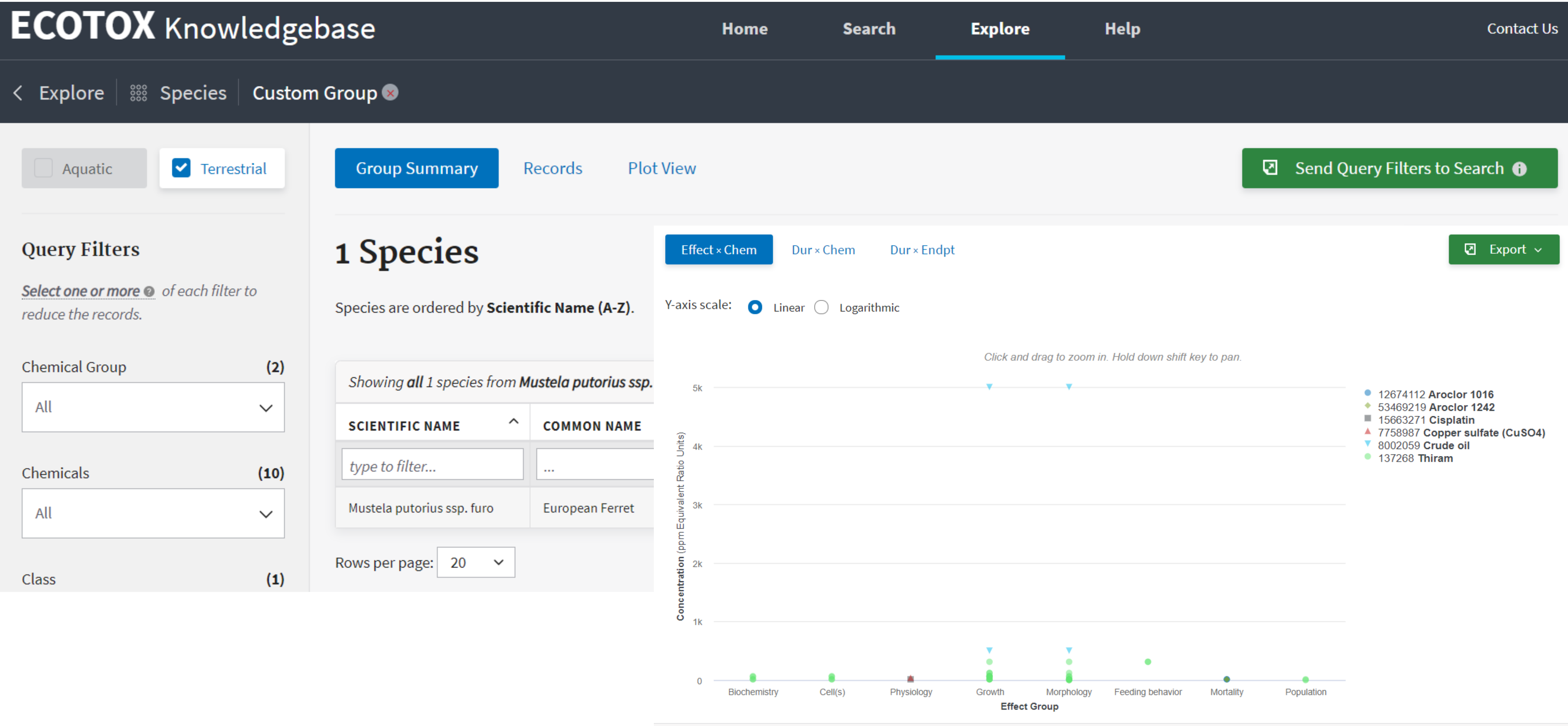
Level 1 Data - Primary

The following links exit the site EXIT

Search: Enter keyword

Protein Name	BLASTp Bitscore	Ortholog Candidate	Ortholog Count	Cut-off	Percent Similarity	Susceptibility Prediction	Analysis Completed	Eukaryote	ECOTOX
PREDICTED: fatty acid-binding protein, liver	236.11	N	62	39.23	93.59	Y	2019 05 24 14:31:09	Y	-
PREDICTED: fatty acid-binding protein, liver	235.73	N	62	39.23	93.43	Y	2019 05 24 14:31:09	Y	-
PREDICTED: fatty acid-binding protein, liver	235.73	N	62	39.23	93.43	Y	2019 05 24 14:31:09	Y	Explore in ECOTOX
fatty acid-binding protein, liver	235.73	N	62	39.23	93.43	Y	2019 05 24 14:31:09	Y	-
liver fatty acid binding protein	234.96	Y	62	39.23	93.13	Y	2019 05 24 14:31:09	Y	Explore in ECOTOX
PREDICTED: fatty acid-binding protein, liver	234.96	N	62	39.23	93.13	Y	2019 05 24 14:31:09	Y	-
fatty acid-binding protein, liver	234.96	N	62	39.23	93.13	Y	2019 05 24 14:31:09	Y	-
fatty acid-binding protein, liver	234.96	N	62	39.23	93.13	Y	2019 05 24 14:31:09	Y	Explore in ECOTOX
fatty acid-binding protein, liver	234.19	N	62	39.23	92.82	Y	2019 05 24 14:31:09	Y	-
fatty acid-binding protein, liver	234.19	N	62	39.23	92.82	Y	2019 05 24 14:31:09	Y	Explore in ECOTOX

(3 of 35) 1 2 3 4 5 6 7 8 9 10 10 ▾ Download Table:



ECOTOX Knowledgebase

Home Search **Explore** Help Contact Us

Explore | Species | Custom Group

☐ Aquatic ☑ Terrestrial

Group Summary | Records | Plot View

Send Query Filters to Search

Query Filters

Select one or more of each filter to reduce the records.

Chemical Group (2)
All

Chemicals (10)
All

Class (1)

1 Species

Species are ordered by **Scientific Name (A-Z)**.

*Showing **all** 1 species from **Mustela putorius ssp.***

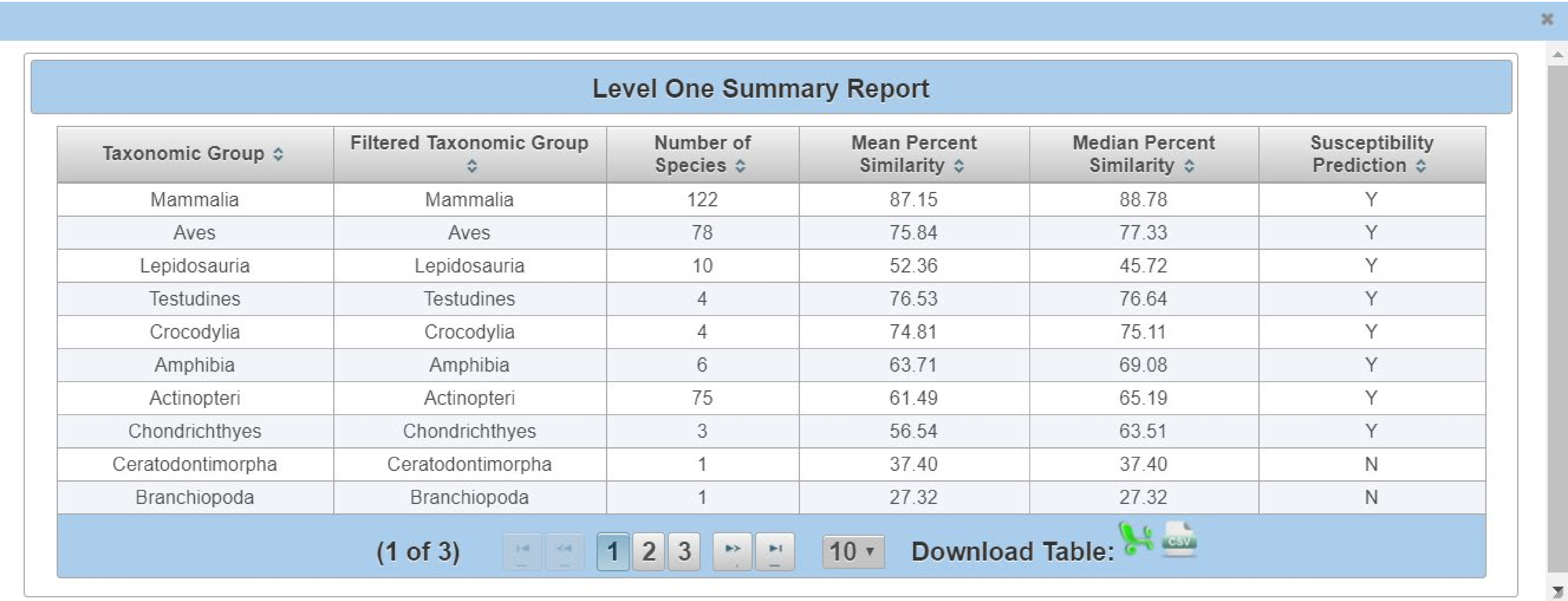
SCIENTIFIC NAME	COMMON NAME
Mustela putorius ssp. furo	European Ferret

Rows per page: 20

Effect × Chem | Dur × Chem | Dur × Endpt

Export

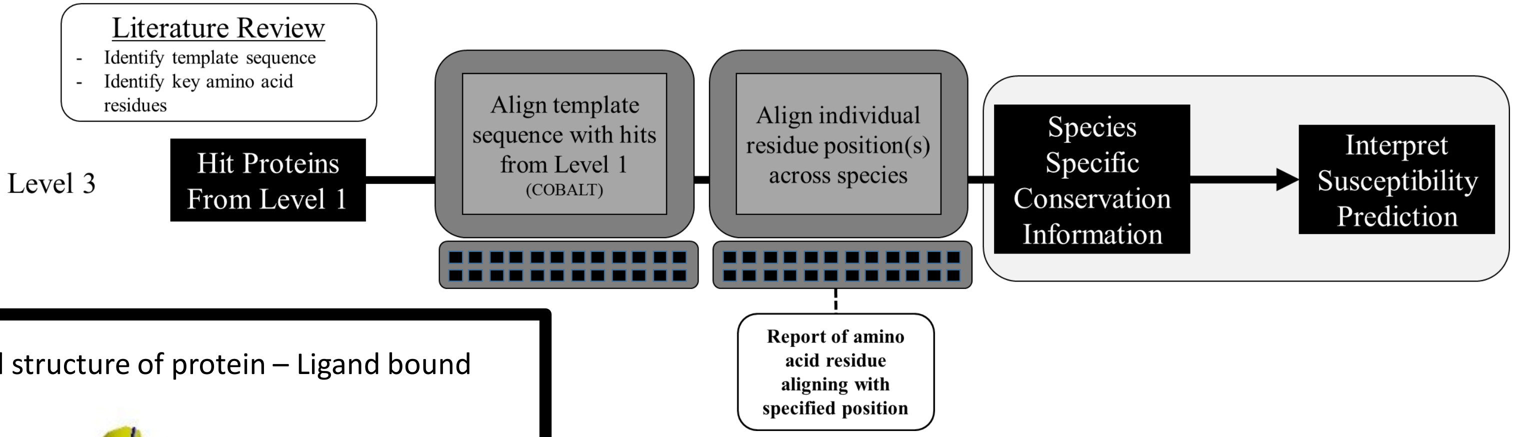
Y-axis scale: ● Linear ○ Logarithmic



Level One Summary Report

Taxonomic Group	Filtered Taxonomic Group	Number of Species	Mean Percent Similarity	Median Percent Similarity	Susceptibility Prediction
Mammalia	Mammalia	122	87.15	88.78	Y
Aves	Aves	78	75.84	77.33	Y
Lepidosauria	Lepidosauria	10	52.36	45.72	Y
Testudines	Testudines	4	76.53	76.64	Y
Crocodylia	Crocodylia	4	74.81	75.11	Y
Amphibia	Amphibia	6	63.71	69.08	Y
Actinopteri	Actinopteri	75	61.49	65.19	Y
Chondrichthyes	Chondrichthyes	3	56.54	63.51	Y
Ceratodontimorpha	Ceratodontimorpha	1	37.40	37.40	N
Branchiopoda	Branchiopoda	1	27.32	27.32	N

Level 3: Individual Amino Acid Residue Comparison



<u>Literature Review</u>
- Identify template sequence
- Identify key amino acid residues

Level 3

Hit Proteins From Level 1

Align template sequence with hits from Level 1 (COBALT)

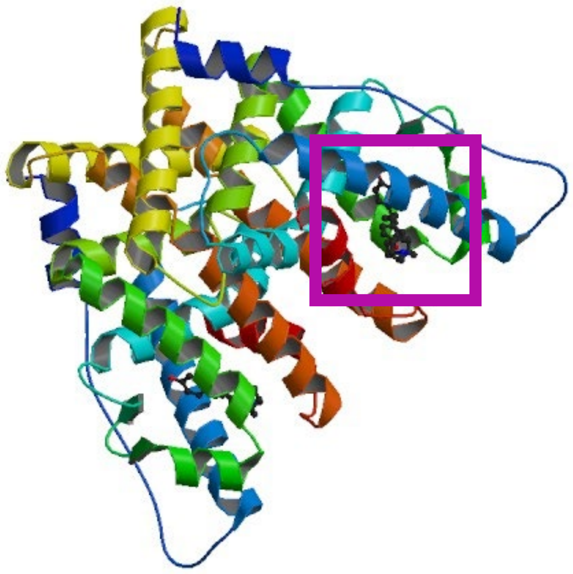
Align individual residue position(s) across species

Species Specific Conservation Information

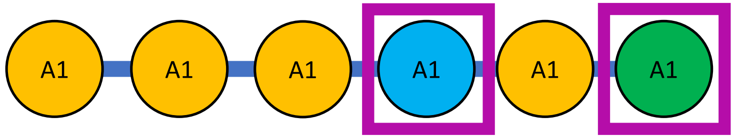
Interpret Susceptibility Prediction

Report of amino acid residue aligning with specified position

Crystal structure of protein – Ligand bound

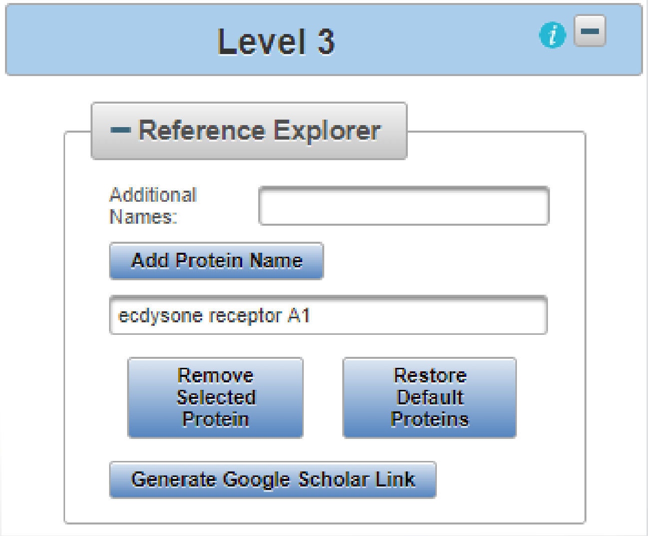


Amino acid residues that interact with the chemical



Where can I find this information?

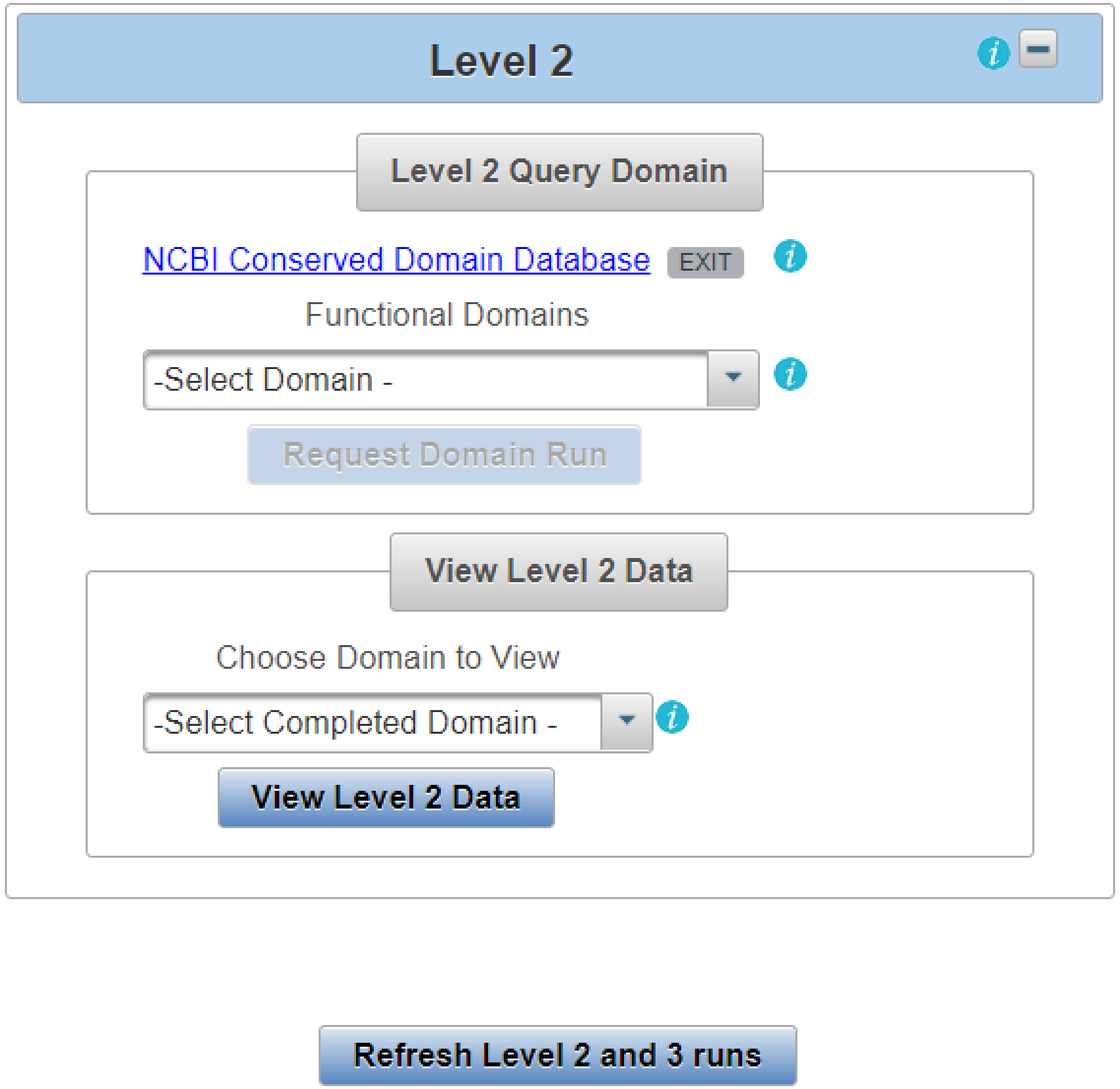
- Literature Review
  - Types of studies I'd be looking for:
    - Site-directed mutagenesis
    - Field resistance (pesticides)
    - Studies of x-ray crystallography
    - Homology modeling



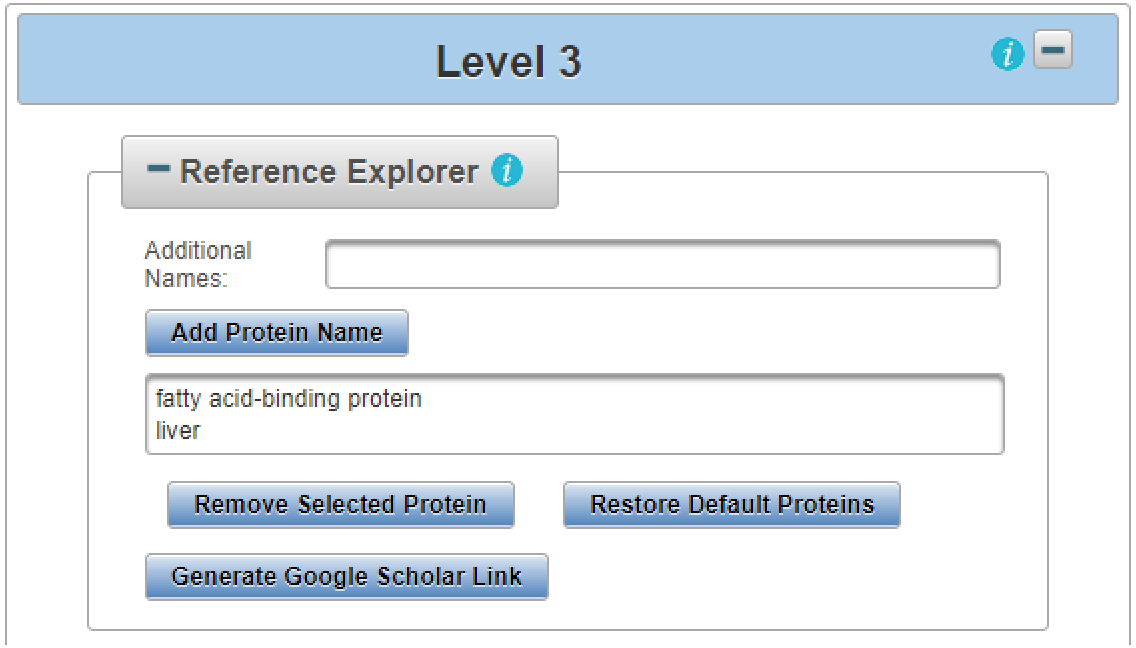
Additional Names:

ecdysone receptor A1

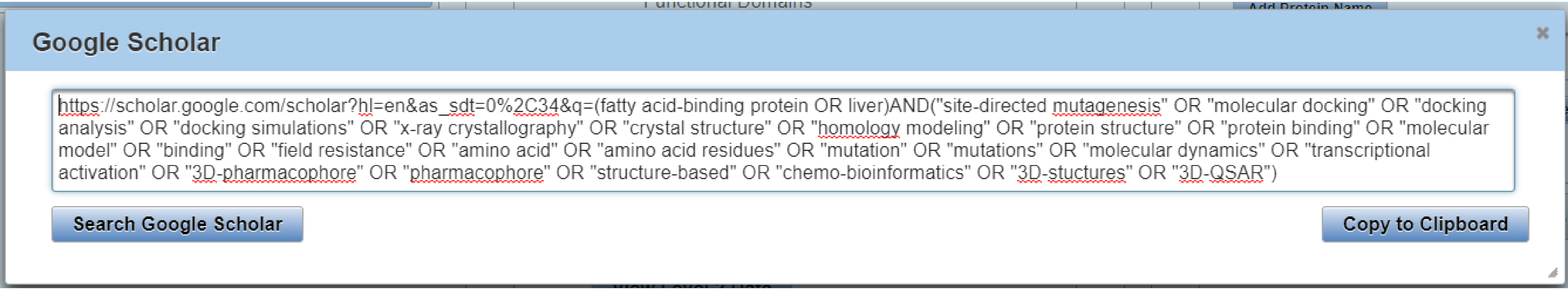
This is not trivial….it can take some time to identify literature and formulate the Level 3 query


Level 2
Level 2 Query Domain
NCBI Conserved Domain Database
EXIT
Functional Domains
-Select Domain -
Request Domain Run
View Level 2 Data
Choose Domain to View
-Select Completed Domain -
View Level 2 Data
Refresh Level 2 and 3 runs

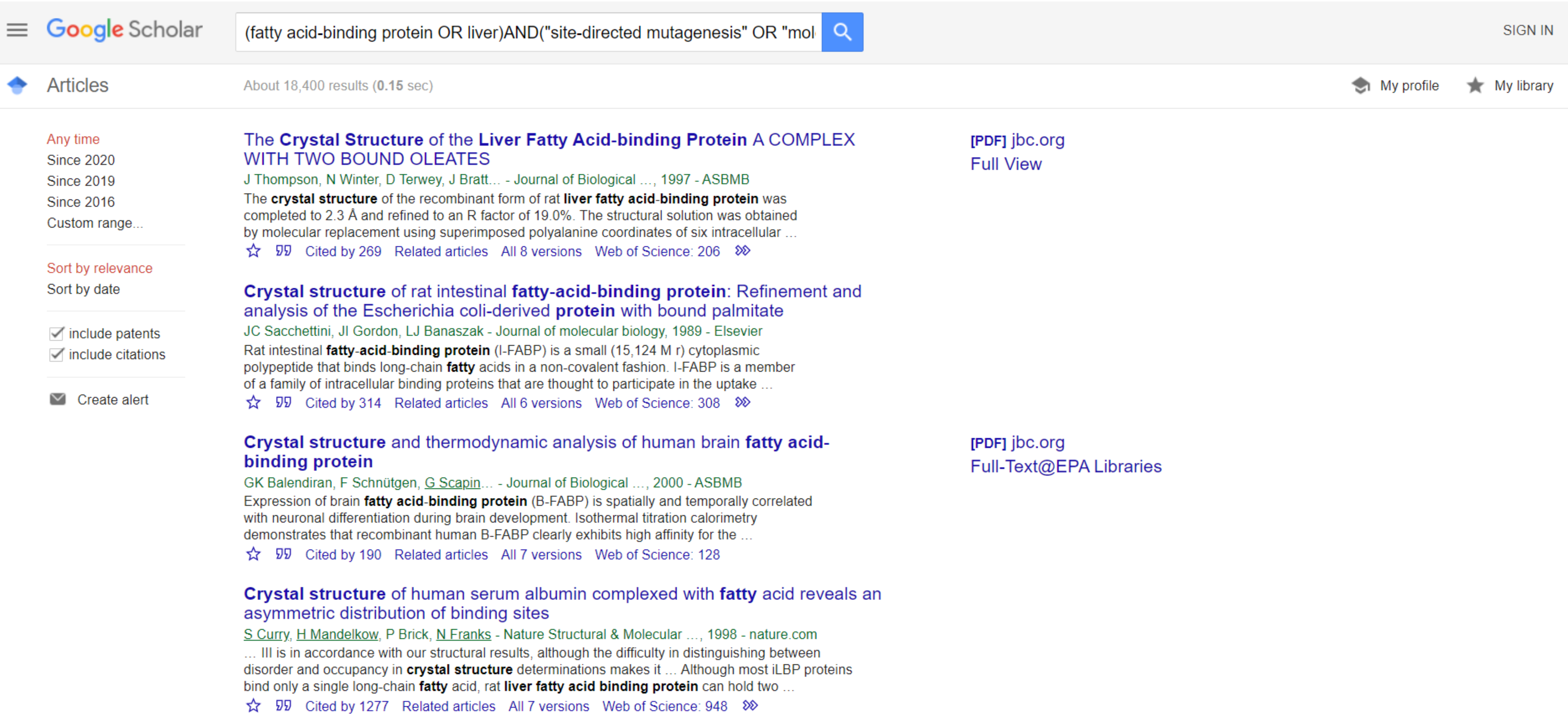
EPA
United States
Environmental Protection
Agency



fatty acid-binding protein
liver



https://scholar.google.com/scholar?hl=en&as_sdt=0%2C34&q=(fatty acid-binding protein OR liver)AND("site-directed mutagenesis" OR "molecular docking" OR "docking analysis" OR "docking simulations" OR "x-ray crystallography" OR "crystal structure" OR "homology modeling" OR "protein structure" OR "protein binding" OR "molecular model" OR "binding" OR "field resistance" OR "amino acid" OR "amino acid residues" OR "mutation" OR "mutations" OR "molecular dynamics" OR "transcriptional activation" OR "3D-pharmacophore" OR "pharmacophore" OR "structure-based" OR "chemo-bioinformatics" OR "3D-stuctures" OR "3D-QSAR")



(fatty acid-binding protein OR liver)AND("site-directed mutagenesis" OR "mol

About 18,400 results (**0.15** sec)

The **Crystal Structure** of the **Liver Fatty Acid-binding Protein** A COMPLEX WITH TWO BOUND OLEATES

J Thompson, N Winter, D Terwey, J Bratt... - Journal of Biological ..., 1997 - ASBMB

The **crystal structure** of the recombinant form of rat **liver fatty acid**-**binding protein** was completed to 2.3 Å and refined to an R factor of 19.0%. The structural solution was obtained by molecular replacement using superimposed polyalanine coordinates of six intracellular ...

☆ 99 Cited by 269 Related articles All 8 versions Web of Science: 206 »

[PDF] jbc.org
Full View

Crystal structure of rat intestinal **fatty-acid-binding protein**: Refinement and analysis of the Escherichia coli-derived **protein** with bound palmitate

JC Sacchettini, JI Gordon, LJ Banaszak - Journal of molecular biology, 1989 - Elsevier

Rat intestinal **fatty**-**acid**-**binding protein** (I-FABP) is a small (15,124 M r) cytoplasmic polypeptide that binds long-chain **fatty** acids in a non-covalent fashion. I-FABP is a member of a family of intracellular binding proteins that are thought to participate in the uptake ...

☆ 99 Cited by 314 Related articles All 6 versions Web of Science: 308 »

Crystal structure and thermodynamic analysis of human brain **fatty acid-binding protein**

GK Balendiran, F Schnütgen, G Scapin... - Journal of Biological ..., 2000 - ASBMB

Expression of brain **fatty acid**-**binding protein** (B-FABP) is spatially and temporally correlated with neuronal differentiation during brain development. Isothermal titration calorimetry demonstrates that recombinant human B-FABP clearly exhibits high affinity for the ...

☆ 99 Cited by 190 Related articles All 7 versions Web of Science: 128

[PDF] jbc.org
Full-Text@EPA Libraries

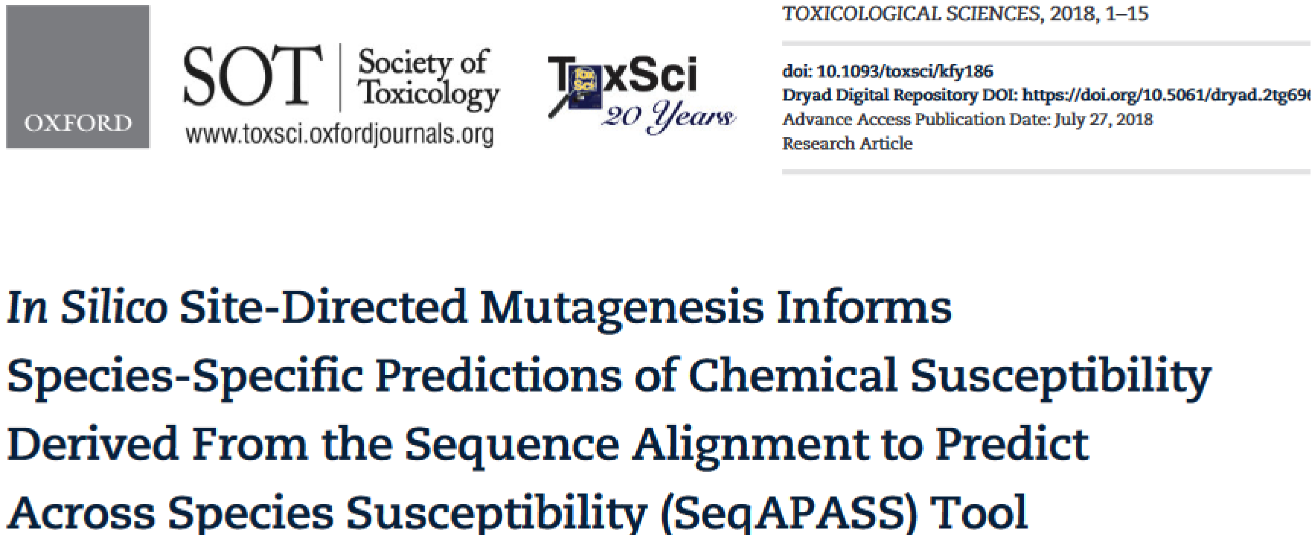
Crystal structure of human serum albumin complexed with **fatty** acid reveals an asymmetric distribution of binding sites

S Curry, H Mandelkow, P Brick, N Franks - Nature Structural & Molecular ..., 1998 - nature.com

... III is in accordance with our structural results, although the difficulty in distinguishing between disorder and occupancy in **crystal structure** determinations makes it ... Although most iLBP proteins bind only a single long-chain **fatty** acid, rat **liver fatty acid binding protein** can hold two ...

☆ 99 Cited by 1277 Related articles All 7 versions Web of Science: 948 »

Focus on Structure: Level 3 Susceptibility Predictions



OXFORD
SOT Society of Toxicology
www.toxsci.oxfordjournals.org
ToxSci 20 Years

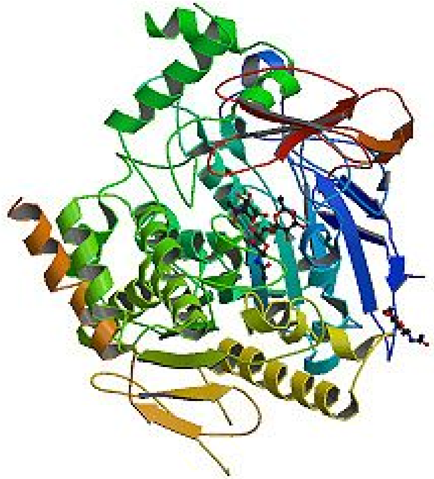
TOXICOLOGICAL SCIENCES, 2018, 1–15

doi: 10.1093/toxsci/kfy186
Dryad Digital Repository DOI: https://doi.org/10.5061/dryad.2tg69
Advance Access Publication Date: July 27, 2018
Research Article

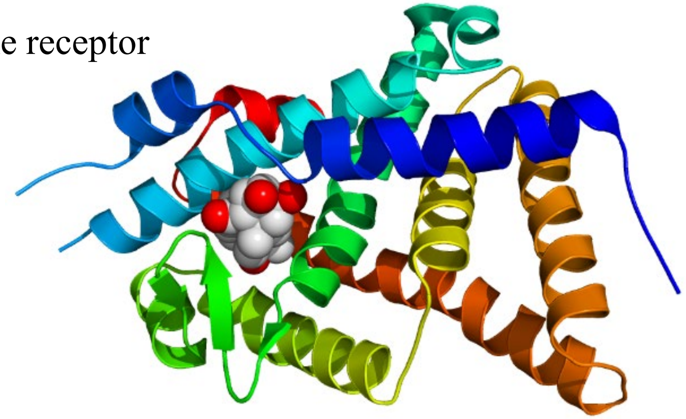
In Silico Site-Directed Mutagenesis Informs Species-Specific Predictions of Chemical Susceptibility Derived From the Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) Tool

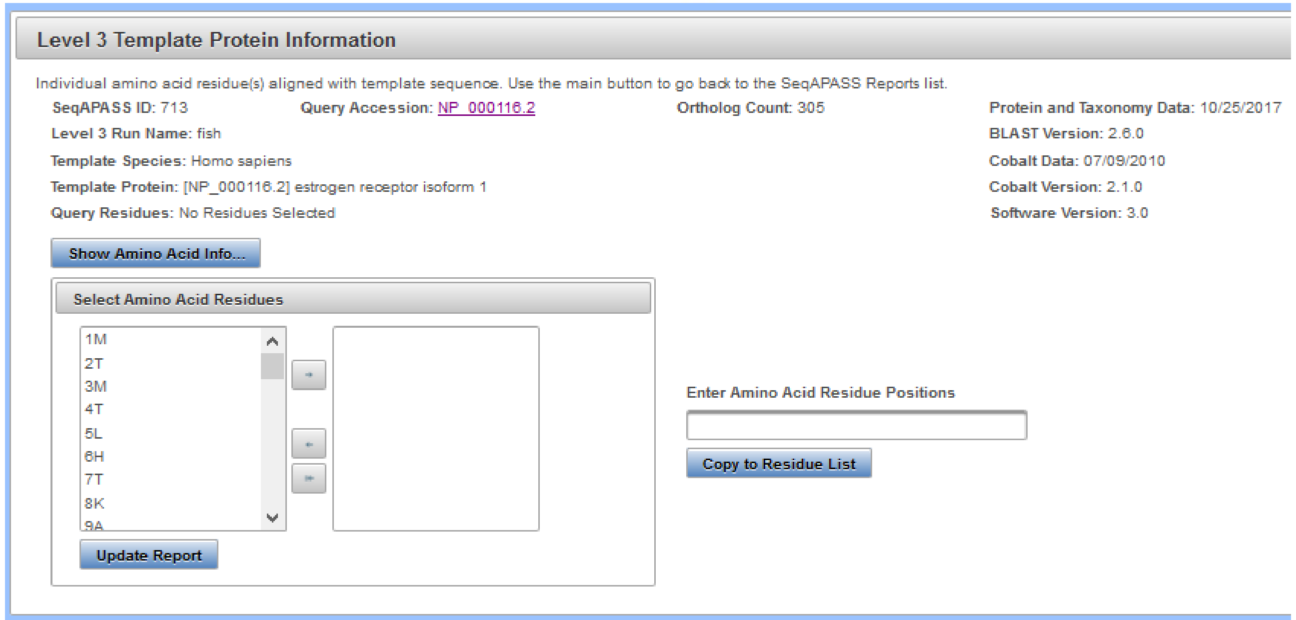
Jon A. Doering,[*,†] Sehan Lee,[‡,§] Kurt Kristiansen,[¶] Linn Evenseth,[¶] Mace G. Barron,[‡] Ingebrigt Sylte,[¶] and Carlie A. LaLone[*,1]

Acetylcholinesterase



Ecdysone receptor



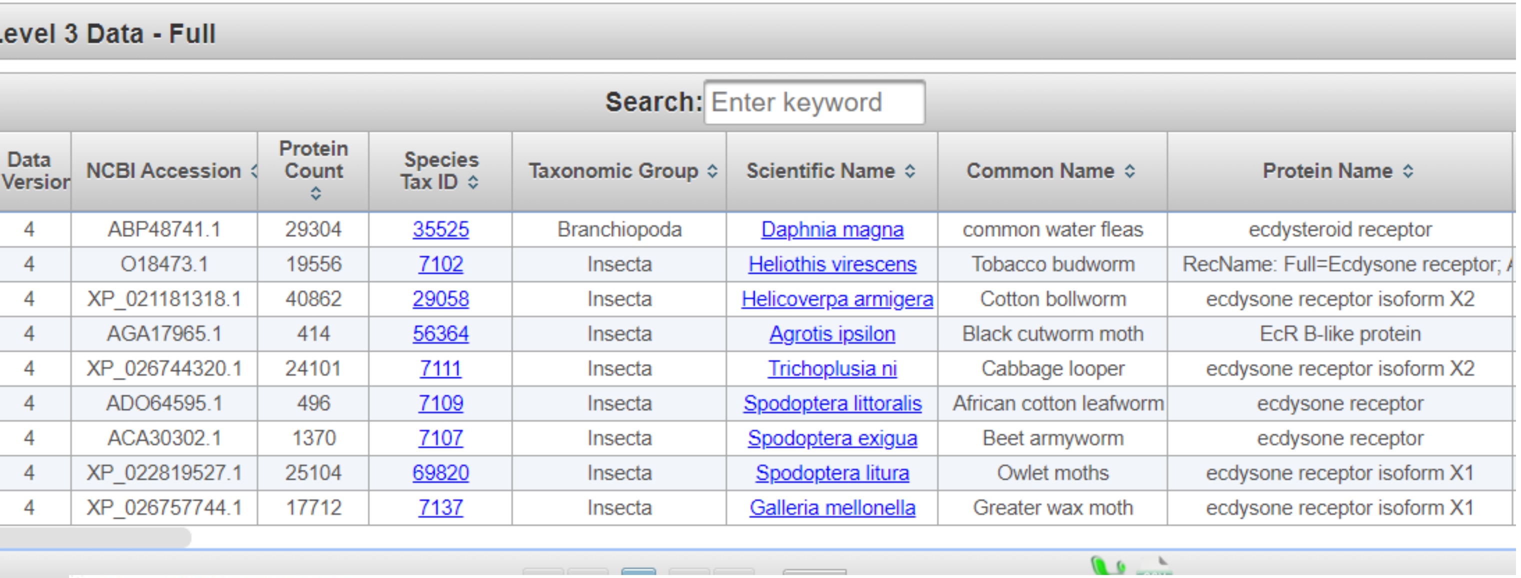


Level 3 Template Protein Information

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

SeqAPASS ID: 713
Query Accession: NP_000116.2
Ortholog Count: 305
Protein and Taxonomy Data: 10/25/2017
Level 3 Run Name: fish
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: 3.0

EPA United States Environmental Protection Agency



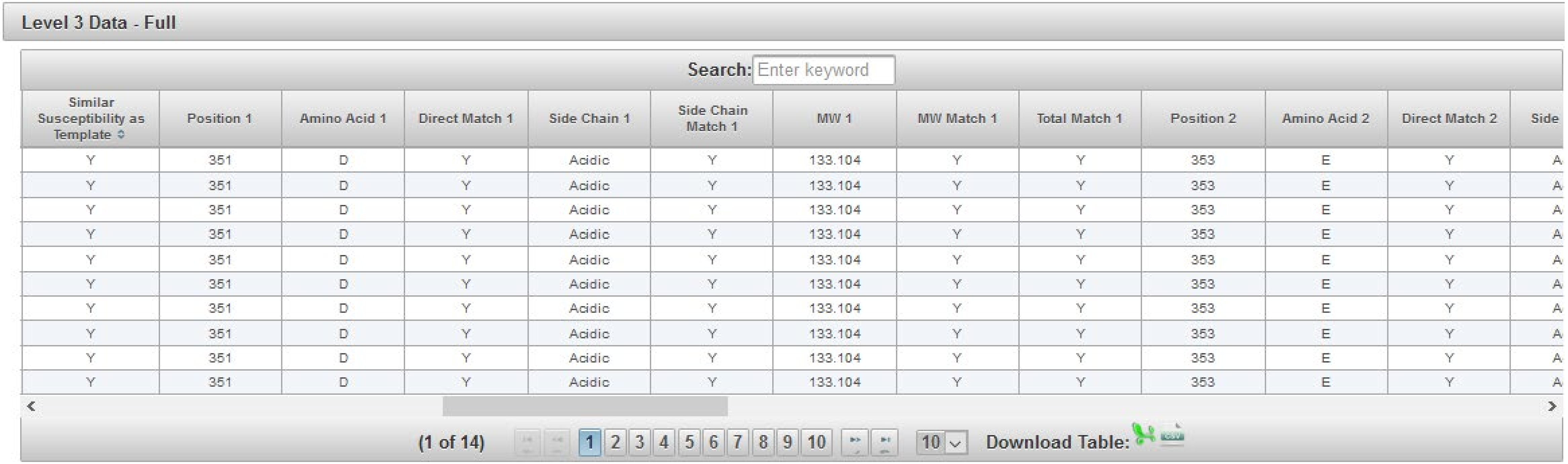
evel 3 Data - Full

Search: Enter keyword

Data Versior	NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Scientific Name	Common Name	Protein Name
4	ABP48741.1	29304	35525	Branchiopoda	Daphnia magna	common water fleas	ecdysteroid receptor
4	O18473.1	19556	7102	Insecta	Heliothis virescens	Tobacco budworm	RecName: Full=Ecdysone receptor;
4	XP_021181318.1	40862	29058	Insecta	Helicoverpa armigera	Cotton bollworm	ecdysone receptor isoform X2
4	AGA17965.1	414	56364	Insecta	Agrotis ipsilon	Black cutworm moth	EcR B-like protein
4	XP_026744320.1	24101	7111	Insecta	Trichoplusia ni	Cabbage looper	ecdysone receptor isoform X2
4	ADO64595.1	496	7109	Insecta	Spodoptera littoralis	African cotton leafworm	ecdysone receptor
4	ACA30302.1	1370	7107	Insecta	Spodoptera exigua	Beet armyworm	ecdysone receptor
4	XP_022819527.1	25104	69820	Insecta	Spodoptera litura	Owlet moths	ecdysone receptor isoform X1
4	XP_026757744.1	17712	7137	Insecta	Galleria mellonella	Greater wax moth	ecdysone receptor isoform X1

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
X	Unknown	Unknown	-100.0
Y	Tyrosine	Aromatic	181.191



Level 3 Data - Full

Search: Enter keyword

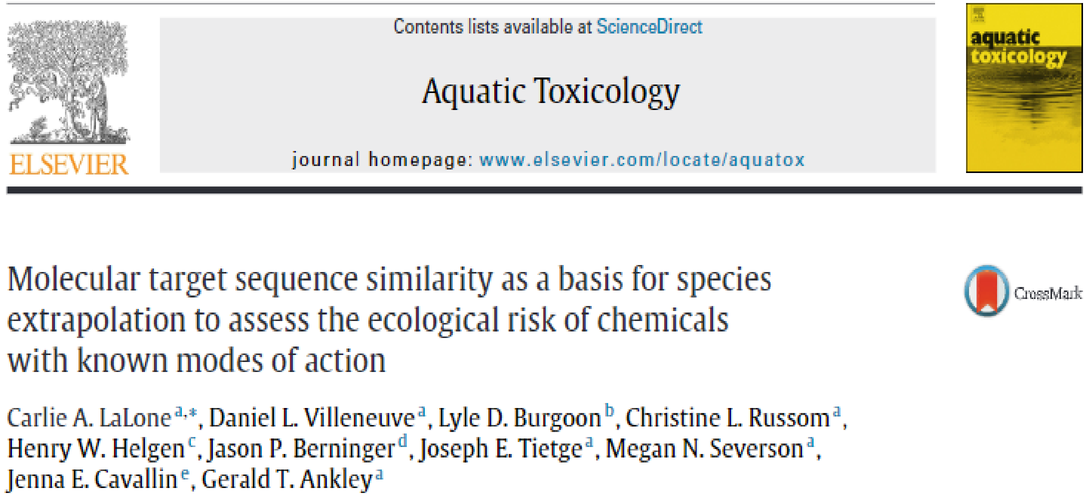
Similar Susceptibility as Template	Position 1	Amino Acid 1	Direct Match 1	Side Chain 1	Side Chain Match 1	MW 1	MW Match 1	Total Match 1	Position 2	Amino Acid 2	Direct Match 2	Side
Y	351	D	Y	Acidic	Y	133.104	Y	Y	353	E	Y	A
Y	351	D	Y	Acidic	Y	133.104	Y	Y	353	E	Y	A
Y	351	D	Y	Acidic	Y	133.104	Y	Y	353	E	Y	A
Y	351	D	Y	Acidic	Y	133.104	Y	Y	353	E	Y	A
Y	351	D	Y	Acidic	Y	133.104	Y	Y	353	E	Y	A
Y	351	D	Y	Acidic	Y	133.104	Y	Y	353	E	Y	A
Y	351	D	Y	Acidic	Y	133.104	Y	Y	353	E	Y	A
Y	351	D	Y	Acidic	Y	133.104	Y	Y	353	E	Y	A
Y	351	D	Y	Acidic	Y	133.104	Y	Y	353	E	Y	A
Y	351	D	Y	Acidic	Y	133.104	Y	Y	353	E	Y	A

Evolution of the SeqAPASS tool

- v1.0 (2016): Develop interface Level 1 & 2 and integrate essential functionality
  - Case studies to support
  - Collect Feedback from Users
- v2.0 (2017): develop Level 3 Susceptibility Predictions
  - Case studies to support
  - Semi automated update of executables and databases (protein, taxonomy, conserved domains)
  - Collect Feedback from Users
- v3.0 (2018): Develop visualization (Level 1 & 2), automate Level 3 Susceptibility Predictions
  - Case studies to support
  - Update executables and databases (protein, taxonomy, conserved domains)
  - Collect Feedback from Users
  - User Guide
  - Training materials
- v4.0 (2019): Improve visualization, user guidance, summary tables, interoperability
  - Case studies to support
  - Update executables and databases
  - Collect Feedback from Users
  - Update/publish user guide
  - Publish SOP for testing

Application of SeqAPASS

MIE → KE → KE → KE → AO


Contents lists available at ScienceDirect

Aquatic Toxicology

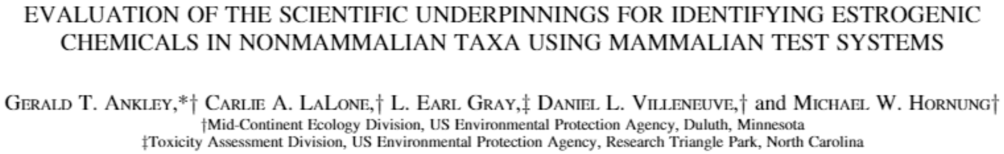
journal homepage: www.elsevier.com/locate/aquatox

Molecular target sequence similarity as a basis for species extrapolation to assess the ecological risk of chemicals with known modes of action

Carlie A. LaLone[a,*], Daniel L. Villeneuve[a], Lyle D. Burgoon[b], Christine L. Russom[a], Henry W. Helgen[c], Jason P. Berninger[d], Joseph E. Tietge[a], Megan N. Severson[a], Jenna E. Cavallin[e], Gerald T. Ankley[a]

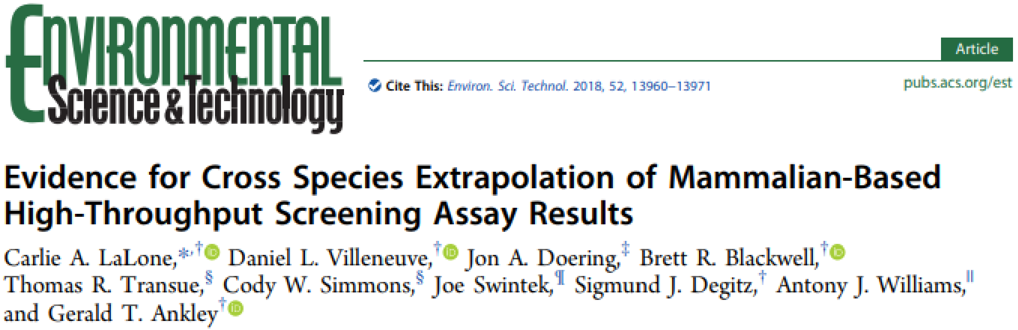
Case studies:	17alpha-ethinyl estradiol	human	estrogen receptor
	Permethrin	mosquito	voltage-gated sodium channel
	17β-trenbolone	bovine	androgen receptor


Environmental Toxicology and Chemistry, Vol. 35, No. 11, pp. 2806–2816, 2016
Published 2016 SETAC
Printed in the USA


EVALUATION OF THE SCIENTIFIC UNDERPINNINGS FOR IDENTIFYING ESTROGENIC CHEMICALS IN NONMAMMALIAN TAXA USING MAMMALIAN TEST SYSTEMS

GERALD T. ANKLEY,*† CARLIE A. LALONE,† L. EARL GRAY,‡ DANIEL L. VILLENEUVE,† and MICHAEL W. HORNUNG†
†Mid-Continent Ecology Division, US Environmental Protection Agency, Duluth, Minnesota
‡Toxicity Assessment Division, US Environmental Protection Agency, Research Triangle Park, North Carolina

Case study:	Human estrogen receptor alpha


Environmental Science & Technology

Article

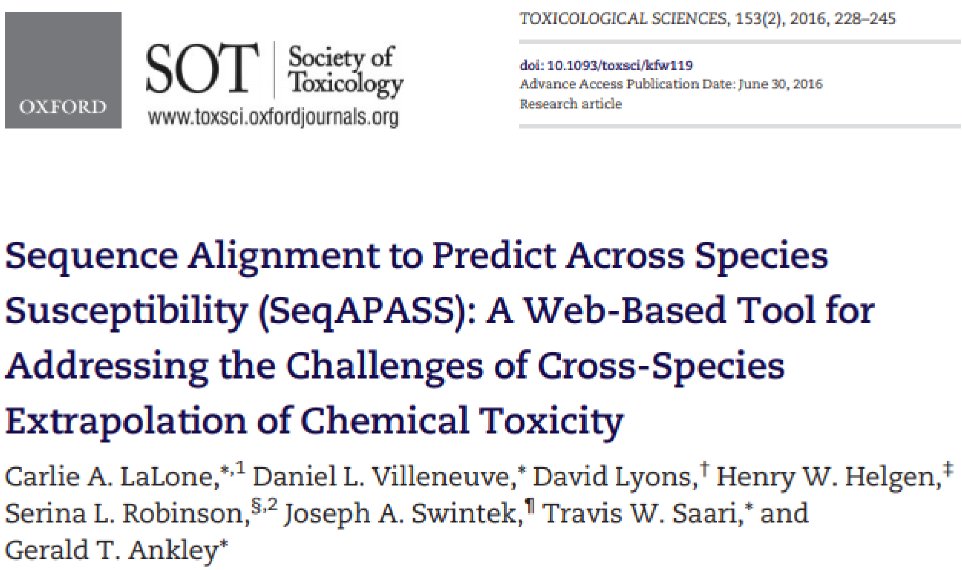
Cite This: Environ. Sci. Technol. 2018, 52, 13960–13971

pubs.acs.org/est

Evidence for Cross Species Extrapolation of Mammalian-Based High-Throughput Screening Assay Results

Carlie A. LaLone,*† Daniel L. Villeneuve,† Jon A. Doering,‡ Brett R. Blackwell,† Thomas R. Transue,§ Cody W. Simmons,§ Joe Swintek,¶ Sigmund J. Degitz,† Antony J. Williams,‖ and Gerald T. Ankley†

Case studies:	All mammalian-based ToxCast Targets (484)


OXFORD

SOT Society of Toxicology
www.toxsci.oxfordjournals.org

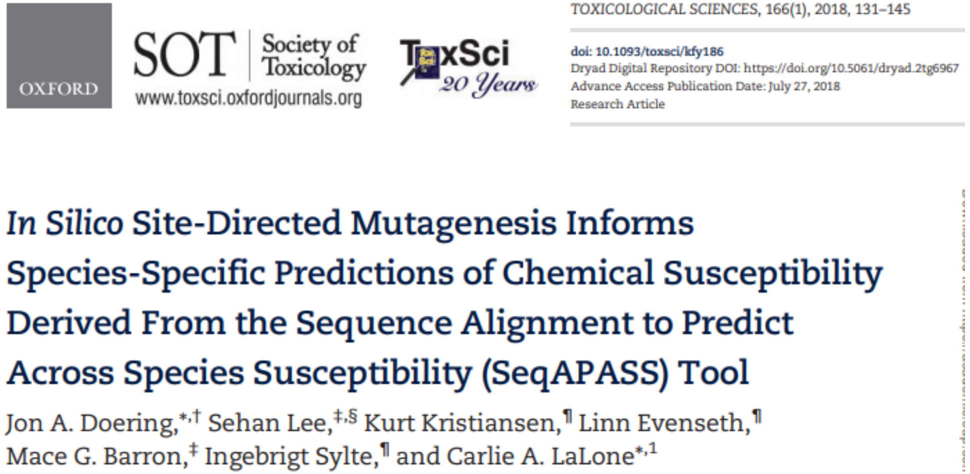
TOXICOLOGICAL SCIENCES, 153(2), 2016, 228–245

doi: 10.1093/toxsci/kfw119
Advance Access Publication Date: June 30, 2016
Research article

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): A Web-Based Tool for Addressing the Challenges of Cross-Species Extrapolation of Chemical Toxicity

Carlie A. LaLone,*,1 Daniel L. Villeneuve,* David Lyons,† Henry W. Helgen,‡ Serina L. Robinson,§,2 Joseph A. Swintek,¶ Travis W. Saari,* and Gerald T. Ankley*

Case studies:	Diacylhydrazines and bisacylhydrazine	tobacco budworm	ecdysone receptor
	Neonicotinoids	honey bee	nicotinic acetylcholine receptor
	Strobilurin fungicides	corn rust	cytochrome b


OXFORD

SOT Society of Toxicology
www.toxsci.oxfordjournals.org

ToxSci 20 Years

TOXICOLOGICAL SCIENCES, 166(1), 2018, 131–145

doi: 10.1093/toxsci/kfy186
Dryad Digital Repository DOI: https://doi.org/10.5061/dryad.2tg6967
Advance Access Publication Date: July 27, 2018
Research Article

In Silico Site-Directed Mutagenesis Informs Species-Specific Predictions of Chemical Susceptibility Derived From the Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) Tool

Jon A. Doering,*,† Sehan Lee,‡,§ Kurt Kristiansen,¶ Linn Evenseth,¶ Mace G. Barron,‡ Ingebrigt Sylte,¶ and Carlie A. LaLone*,1

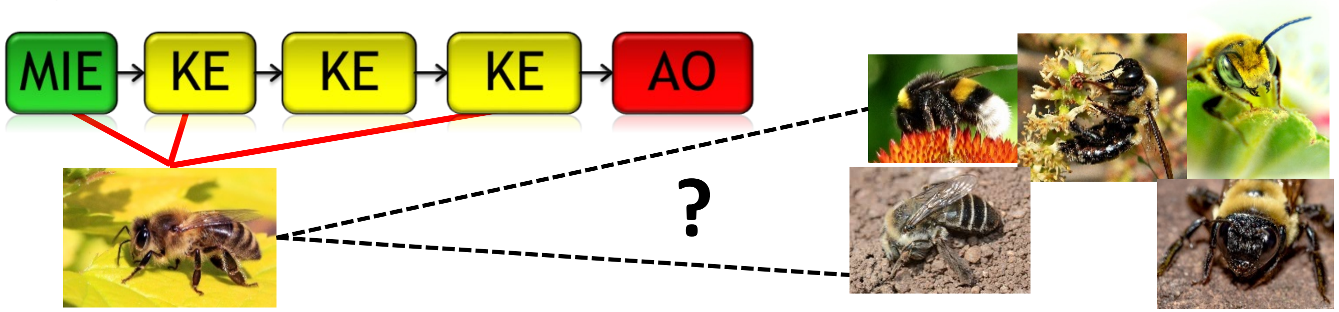
Case studies:	Mouse acetylcholinesterase
	Tobacco budworm ecdysone receptor

SeqAPASS Advances FY19-22

SeqAPASS development/training and outreach focus
- Public version releases
- Published user guides
- Standard Operating Procedures - Testing
- Advanced Molecular modeling/docking

Laboratory confirmation of predictions
- Site-directed mutagenesis

SeqAPASS application
- PFAS bioaccumulation/toxicity targets
- High priority pesticides
- -Apis vs. non-apis bee

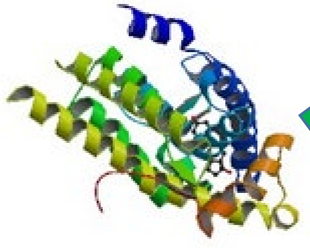


Advanced Molecular Modeling to Inform Species-Specific Predictions of Chemical Susceptibility

Query Species/Sequence

RCSB PDB
PROTEIN DATA BANK
155618 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

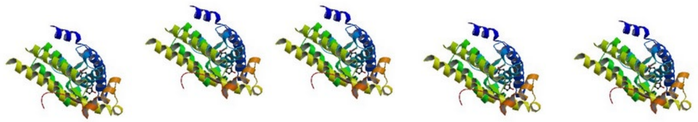
Develop library of quality structures



Hit Species/Sequence

Build Homology Models

Line of evidence: Can homology model be built?



Use homology models to Conduct in silico docking studies

Line of evidence: Can chemical dock?

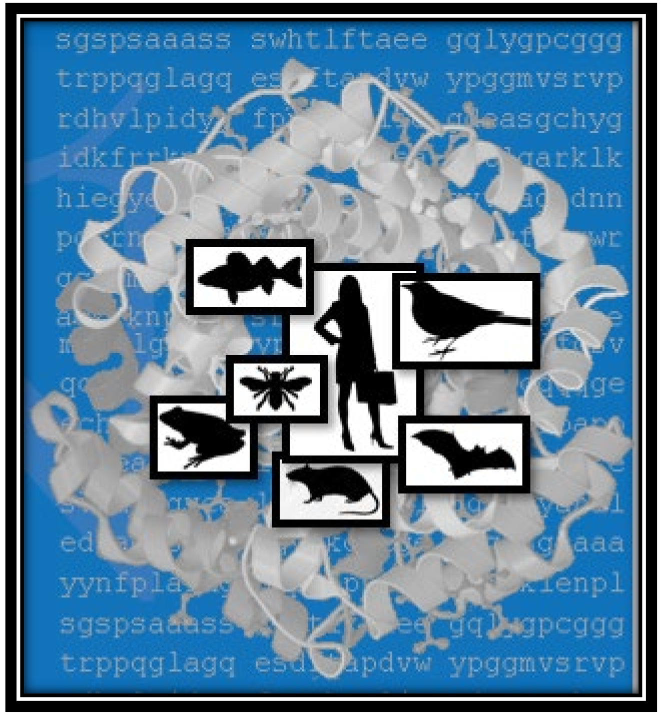
Generate Predicted Binding Affinity values

Take Home Message

Regardless of your experience level with such comparisons/data/results

YOU can use SeqAPASS

After all, it was created for you!!!!



LaLone.Carlie@epa.gov; Blatz.Donovan@epa.gov; Vliet.Sara@epa.gov

SETAC Europe 30th Annual Meeting

3–7 May 2020 | Dublin, Ireland

Training Course

TC09 - Addressing Challenges in Cross-Species Extrapolation of Chemical Toxicity Information: Application of the SeqAPASS Tool

Acknowledgements

U.S. EPA, ORD
- Gerald Ankley
- Daniel Villeneuve
- Donovan Blatz (ORISE)
- Jon Doering (U of Lethbrigde)
- Colin Finnegan (Iowa State University)

GDIT
- Thomas Transue
- Cody Simmons
- Audrey Wilkinson

Badger Technical Services
- Joe Swintek

SeqAPASS v4.0


Choose Search Type: By Species

LaLone.Carlie@epa.gov

https://seqapass.epa.gov/seqapass/