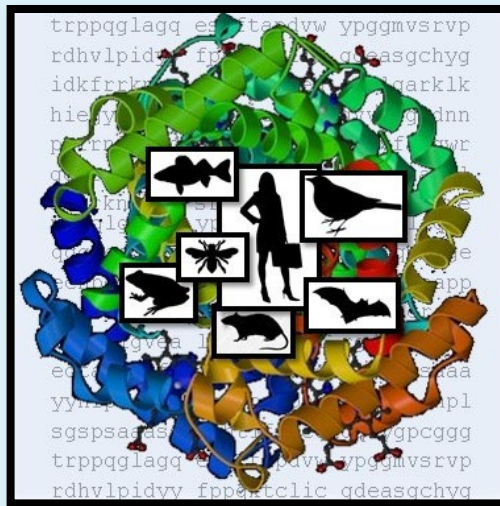


# Overview: Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)



NAMs Training Workshop

RTP, NC

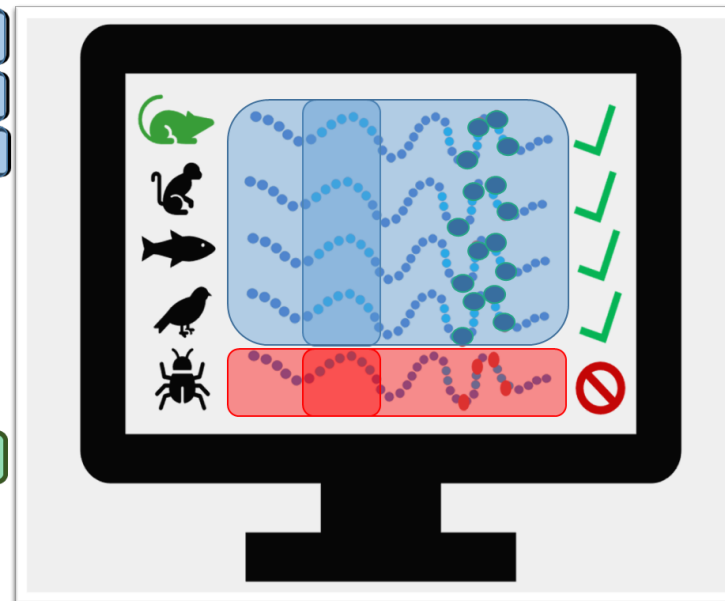
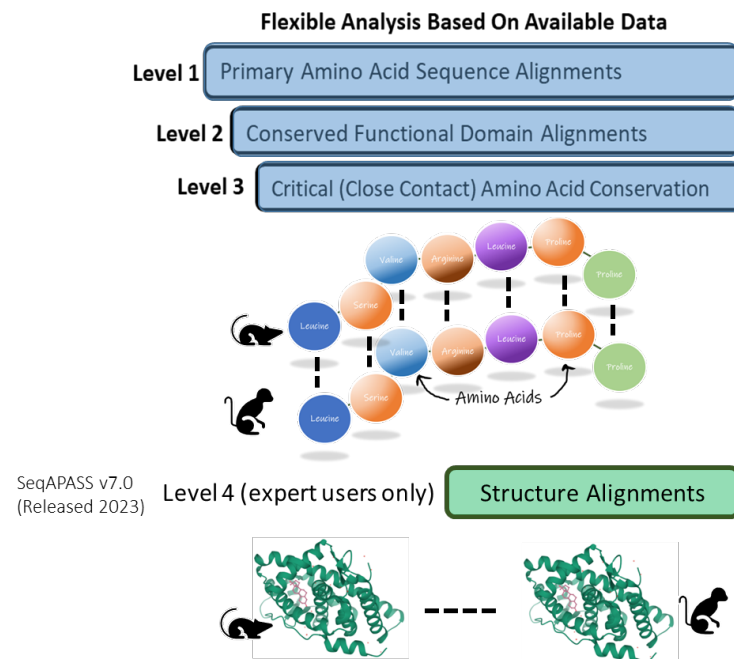
April 24– 25, 2024

Carlie A. LaLone & Marissa Brickley



# Purpose of SeqAPASS

- Species extrapolation of toxicity knowledge
  - Is the biology conserved across species?
  - Is the chemical protein interaction conserved across species?
- Yields a prediction of the likelihood for chemical susceptibility
  - Gather lines of evidence toward protein sequence, structural, and chemical-protein interaction similarity



# Data and Connections

## Guide the User to Sources to Identify Protein Targets:

- DrugBank - <https://www.drugbank.ca>
- VSDB: Veterinary Substances DataBase - <http://sitem.herts.ac.uk/aeru/vsdb/index.htm>
- Therapeutic Target Database - <http://db.idrblab.net/ttd/>
- The Toxin and Toxin-Target Database - <http://www.t3db.ca>
- AOP-Wiki - <https://aopwiki.org>
- CompTox Chemicals Database - <https://comptox.epa.gov/dashboard>

## Provide Transparency for Source Data and Executables:

- NCBI Taxonomy Database - <http://www.ncbi.nlm.nih.gov/taxonomy>
- NCBI Protein Database - <http://www.ncbi.nlm.nih.gov/protein>
- NCBI Conserved Domain Database - <http://www.ncbi.nlm.nih.gov/Structure/cdd/>
- NCBI COBALT - <http://www.st-va.ncbi.nlm.nih.gov/tools/cobalt/>
- I-TASSER - <https://zhanggroup.org/I-TASSER/>
- TM-align - <https://zhanggroup.org/TM-align/>
- AlphaFold - <https://alphafold.ebi.ac.uk/>
- RCSB PDB - <https://www.rcsb.org/>

## Guide the User to Appropriate Resources for Individual Amino Acid Comparisons:

- Google Scholar - <https://scholar.google.com/>

## Aids for Data Synthesis and Connection:

- ECOTOX Knowledgebase - <https://cfpub.epa.gov/>
- U.S. Fish & Wildlife Environmental Conservation Online System - <https://ecos.fws.gov/ecp/>



CompTox Chemicals Dashboard



	<b>Taxonomy</b> 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.
	<b>Protein</b> 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.
	<b>CDD</b> 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.



COBALT

AlphaFold Protein Structure Database



Google Scholar



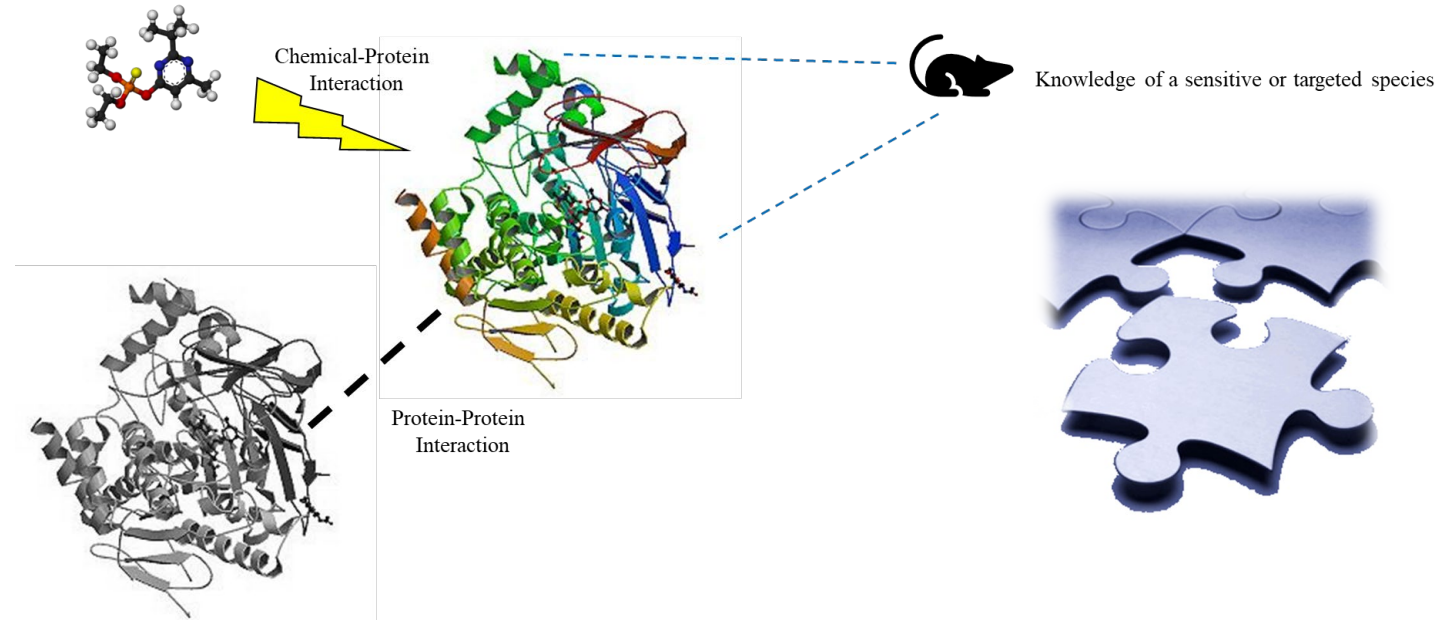
U.S. Fish & Wildlife Service

ECOS Environmental Conservation Online System

Conserving the Nature of America

# Limitations or considerations

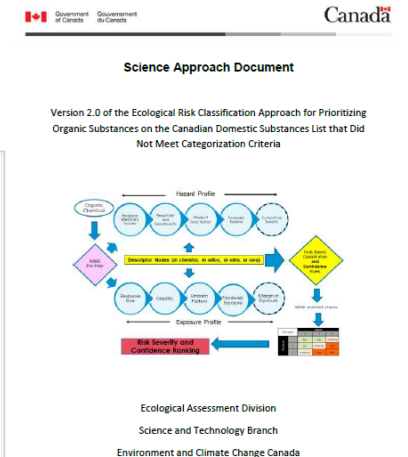
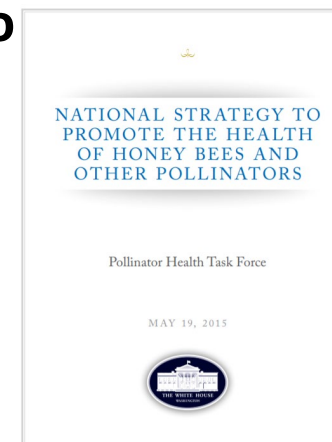
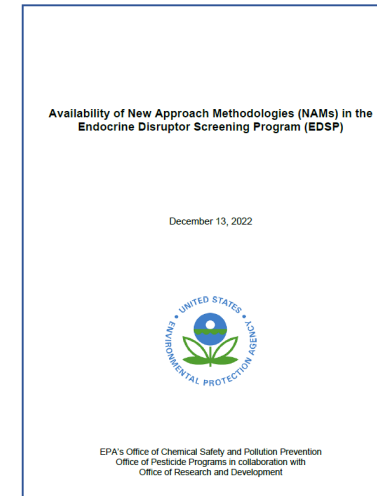
- Domain of applicability is clearly defined for use of data for species extrapolation purposes vs. hypothesis generation



- Limited by the quality and quantity of protein sequence and structure availability, including annotation accuracies
  - Continues to improve and require less resources to attain
  - Large genome sequencing efforts representing the diversity of species
  - Refined and improved annotation pipelines
  - Structural prediction database and approaches enhanced with AI and other bioinformatics technology

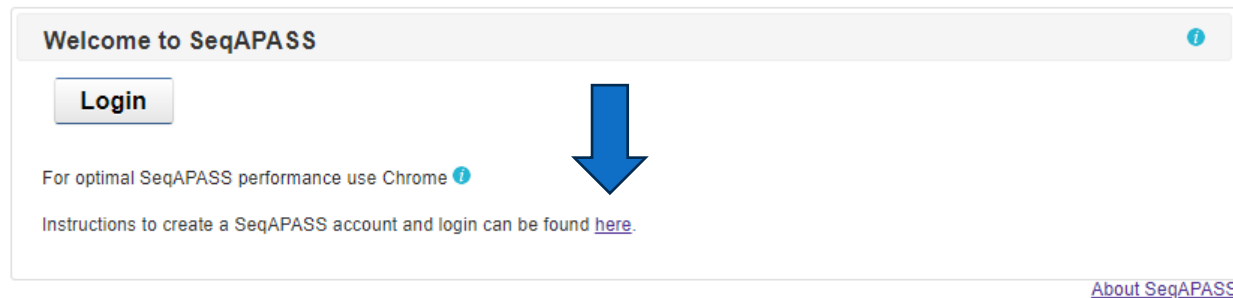
# Example of Tool/Data Use

- **Predict relative intrinsic susceptibility**
  - Pesticides (**OPP FIFRA**)
  - Endangered Species Act (**OPP and Regions ESA**)
  - Derivation of Aquatic Life Criteria (**OW ALC**)
- **Predict chemical bioaccumulation across species**
  - Chemicals of concern: PFAS
- **Extrapolate high throughput screening data (**OPP EDSP**)**
  - Chemicals that target human estrogen receptor alpha, androgen receptor, steroidogenic enzymes, thyroid axis proteins
  - All ToxCast Assay targets
- **Extrapolate adverse outcome pathway knowledge across species**
  - Define the taxonomic relevance: Apis vs Non-Apis bees (**OPP FIFRA**)



# Accessing SeqAPASSv7.1 Released January 2024

- **Free** and **publicly accessible Web-based tool**
- User has a personal account



## Guide to Login for SeqAPASS Users

Steps for acquiring a SeqAPASS account for internal (US EPA) and external users for version 6.0 and beyond.

### Registering for SeqAPASS

#### New 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.

#### New External users

1. Please view directions for creating an account: [Creating New SeqAPASS Account \(pdf\)](#) (668.4 KB)

- <https://seqapass.epa.gov/seqapass/>

# SeqAPASS Points of Contact



Product Owner  
Carlie LaLone  
LaLone.Carlie@epa.gov



PhD Candidate, Pathways Trainee  
Marissa Brickley (Jensen)  
Jensen.Marissa@epa.gov



Lead Developer  
Cody Simmons  
Simmons.Cody@epa.gov



ORISE Trainee  
Ryan Staub  
Staub.Ryan@epa.gov