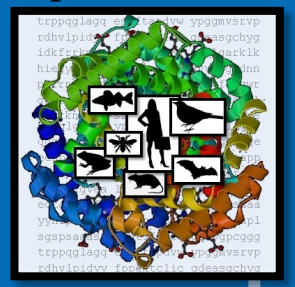
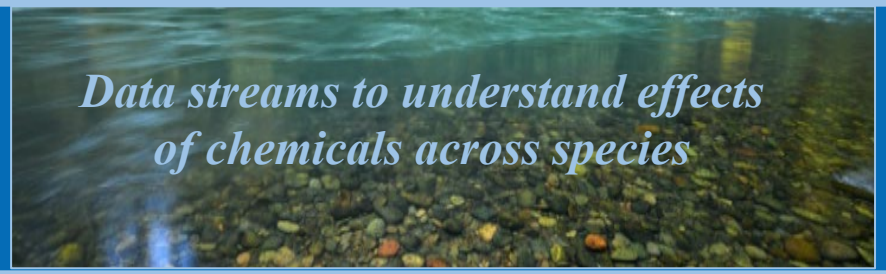
SeqAPASS Version 6.0 – Integrating ecotoxicology data into sequence-based predictions of chemical susceptibility

SeqAPASS v6

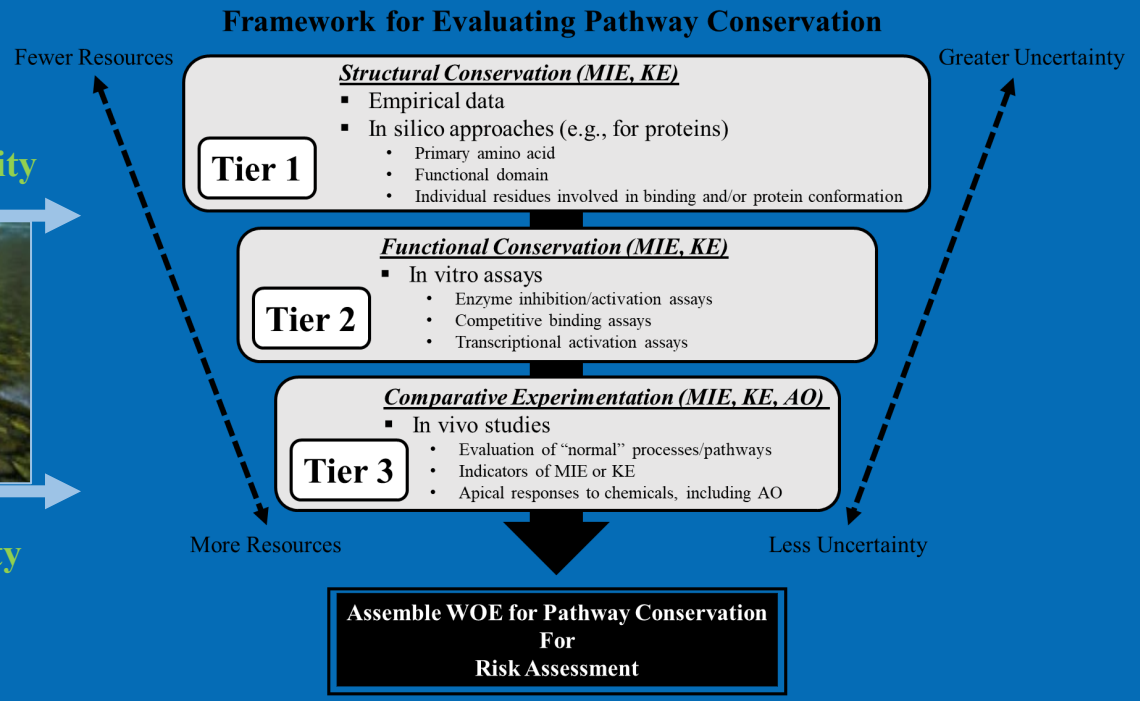


Sequence-based predictions of chemical susceptibility



Empirical toxicity data detailing chemical sensitivity

ECOTOX v5



Presenter: Carlie A. LaLone, Ph.D.

Oct. 28th, 2021

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

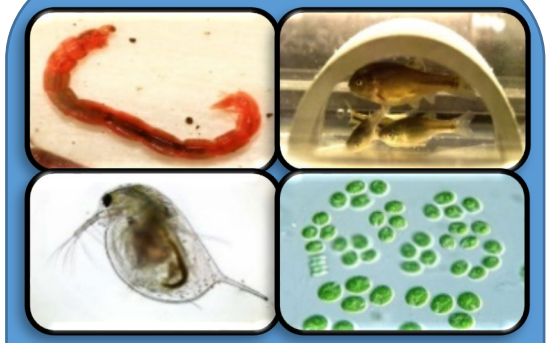
Overview

- Need for cross species extrapolation
- The Sequence Alignment to Predict Across Susceptibility (SeqAPASS) tool
- ECOTOX Knowledgebase
- Bridging sequence with existing empirical
- ECOTOX widget in SeqAPASS
- Case example with transthyretin

Toxicity Testing to Understand Chemical Safety

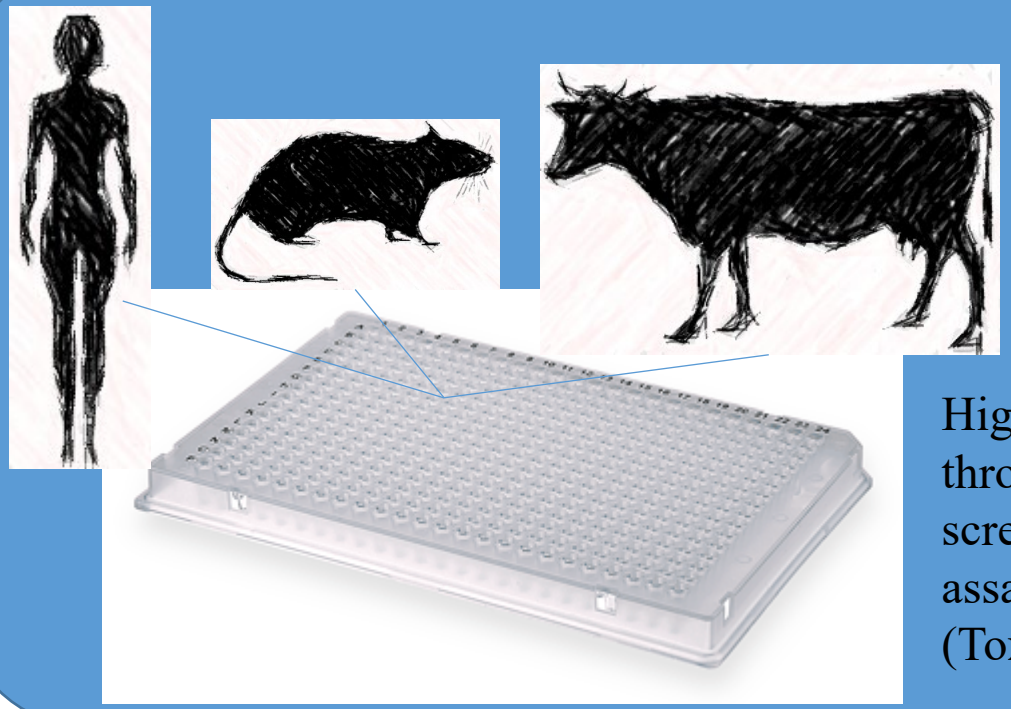
- **<u>US EPA Examples:</u>**
- *Clean Air Act*
- *Clean Water Act*
- *Resource Recovery Act*
- *Endangered Species Act*
- *Food Quality Protection Act*
- *Endocrine Disruptor Screening Program*
- *Federal Insecticide, Fungicide, and Rodenticide Act*
- *Frank R. Lautenberg Chemical Safety for the 21st Century Act*
- *Comprehensive Environmental Response, Compensation, and Liability Act*
- *Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses*

Need for Advances in Species Extrapolation



High throughput transcriptomics

Whole organism toxicity testing



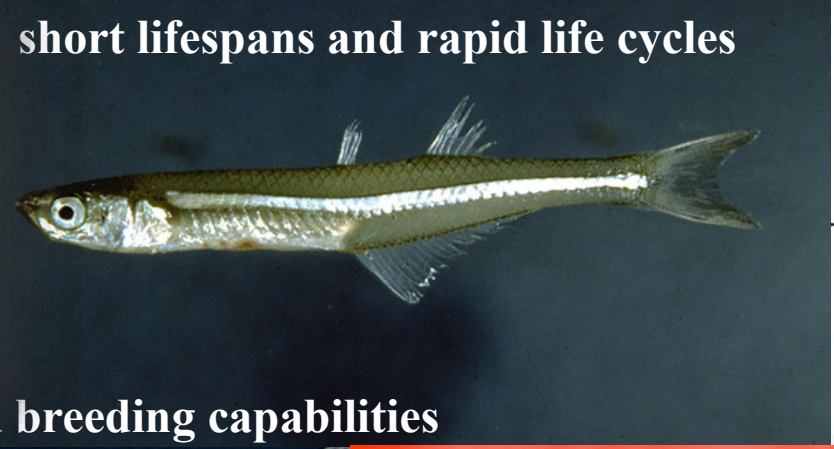
High-throughput screening assays (ToxCast)

Define the taxonomic domain of applicability in AOP development

Use of model organisms as surrogates representing the diversity of species in the environment

cheap and readily available

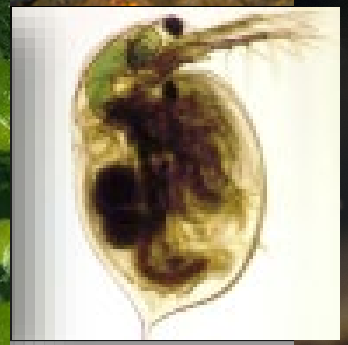
short lifespans and rapid life cycles



easy maintenance and good breeding capabilities

ability to control diet and surroundings

requires least space and time-consuming care



Species Extrapolation

What is it?

- Using existing knowledge about one species to estimate, predict, project, or infer the effect, impact, or trajectory of another species
    - For chemical safety typically dealing with toxicity

Why is it important:

- Limited or no toxicological data for the animal or plant species of interest – reliance on surrogate (model organisms)
    - Impractical to generate new data for all species
- Testing resources are limited
    - International interest to reduce animal use
    - Ever-increasing demand to evaluate more chemicals in a timely and sometimes expedited manner
- Sensitivity of species must be estimated based on scientifically-sound methods of cross-species extrapolation
    - Immense diversity of species in the wild
    - Important challenge for species listed under the Endangered Species Act

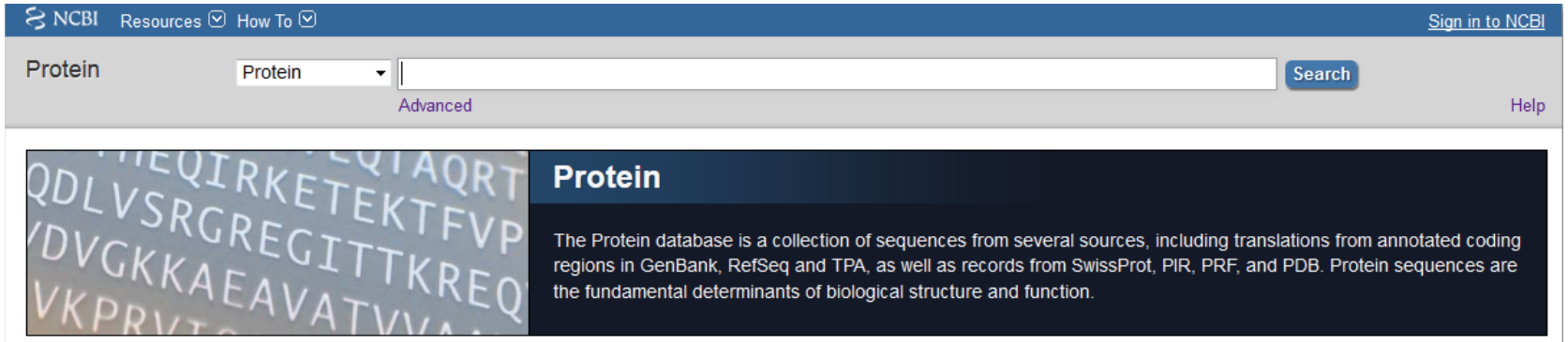
Where could we begin in understanding species similarities and differences?

Look for existing, expanding data that does not require the destruction of live organisms

Sequence and structural data: New tools and technologies have emerged

- Improved sequencing technologies
- Large databases of sequence data

NCBI: 210,703,648 Proteins representing 113,002 Organisms



EPA United States Environmental Protection Agency

Bioinformatics

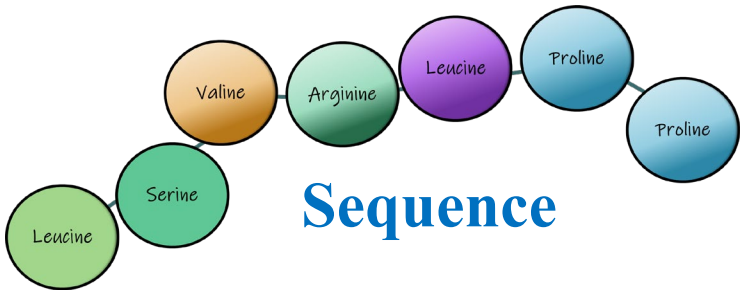
- Combines mathematics, information science, and biology to **answer biological questions**

- Developing methodology and analysis tools **to** **explore large volumes of biological data**
  - Query, extract, store, organize, systematize, annotate, visualize, mine, and interpret complex data
    - Usually **pertains to DNA and amino acid sequences**

Let the computers do the work

EPA United States Environmental Protection Agency

Begin Simple and Advance as the Science Advances



Sequence

Structure

Function

Consider sequence and structural attributes to understand protein conservation across species

EPA United States Environmental Protection Agency

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

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

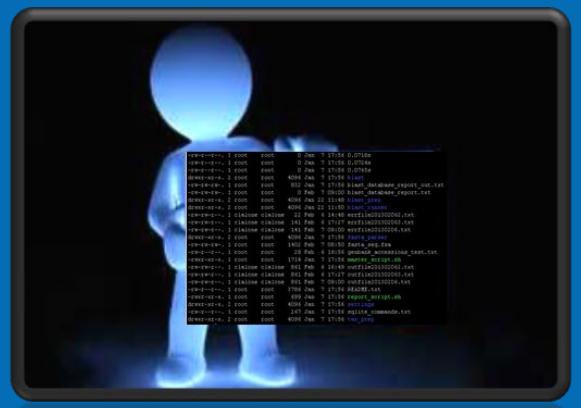
doi: 10.1093/toxsci/kfw119
Advance Access Publication Date: June 30, 2016
Research article

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

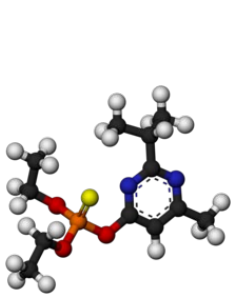
Carlie A. LaLone,*,[1] Daniel L. Villeneuve,* David Lyons,† Henry W. Helgen,‡ Serina L. Robinson,§,[2] Joseph A. Swintek,¶ Travis W. Saari,* and Gerald T. Ankley*

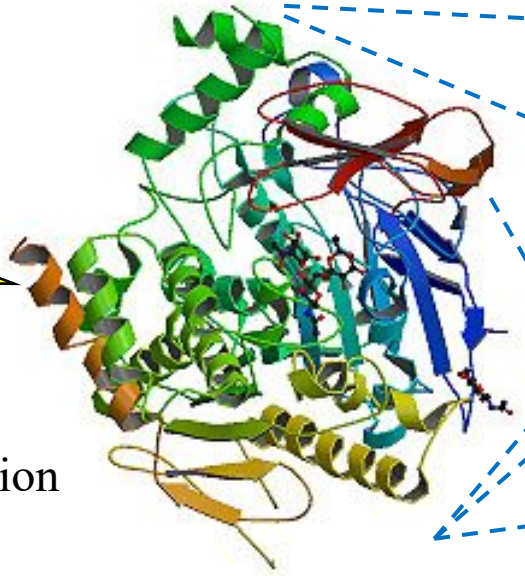
https://seqapass.epa.gov/seqapass/

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)



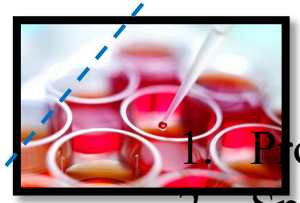
What information is required for a SeqAPASS query?





Chemical-Protein Interaction

Knowledge of a sensitive or targeted species

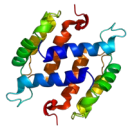


Knowledge of the model organism used in an *in vitro* assay

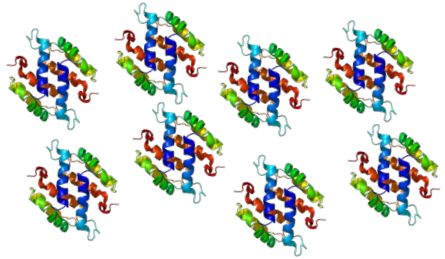
1. Protein
2. Species

Knowledge of the species for which the Key Event was developed

Chemical Molecular Target in Target Species



≅



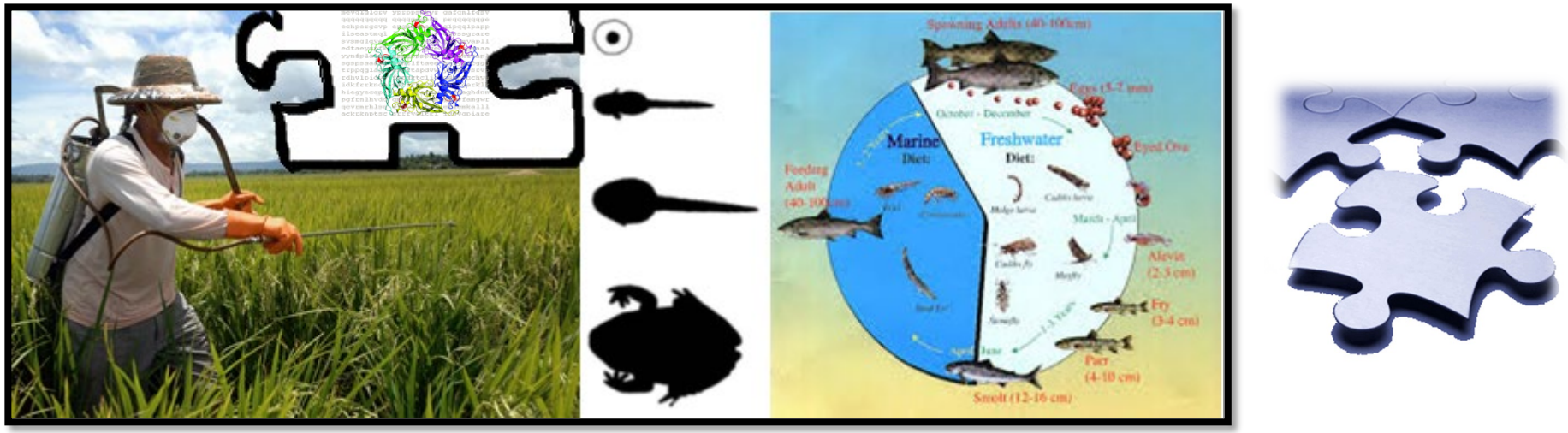
Compare to Millions of Proteins From Thousands of Species

Greater similarity = Greater likelihood that chemical can act on the protein

Line of Evidence: Predict Potential Chemical Susceptibility Across Species

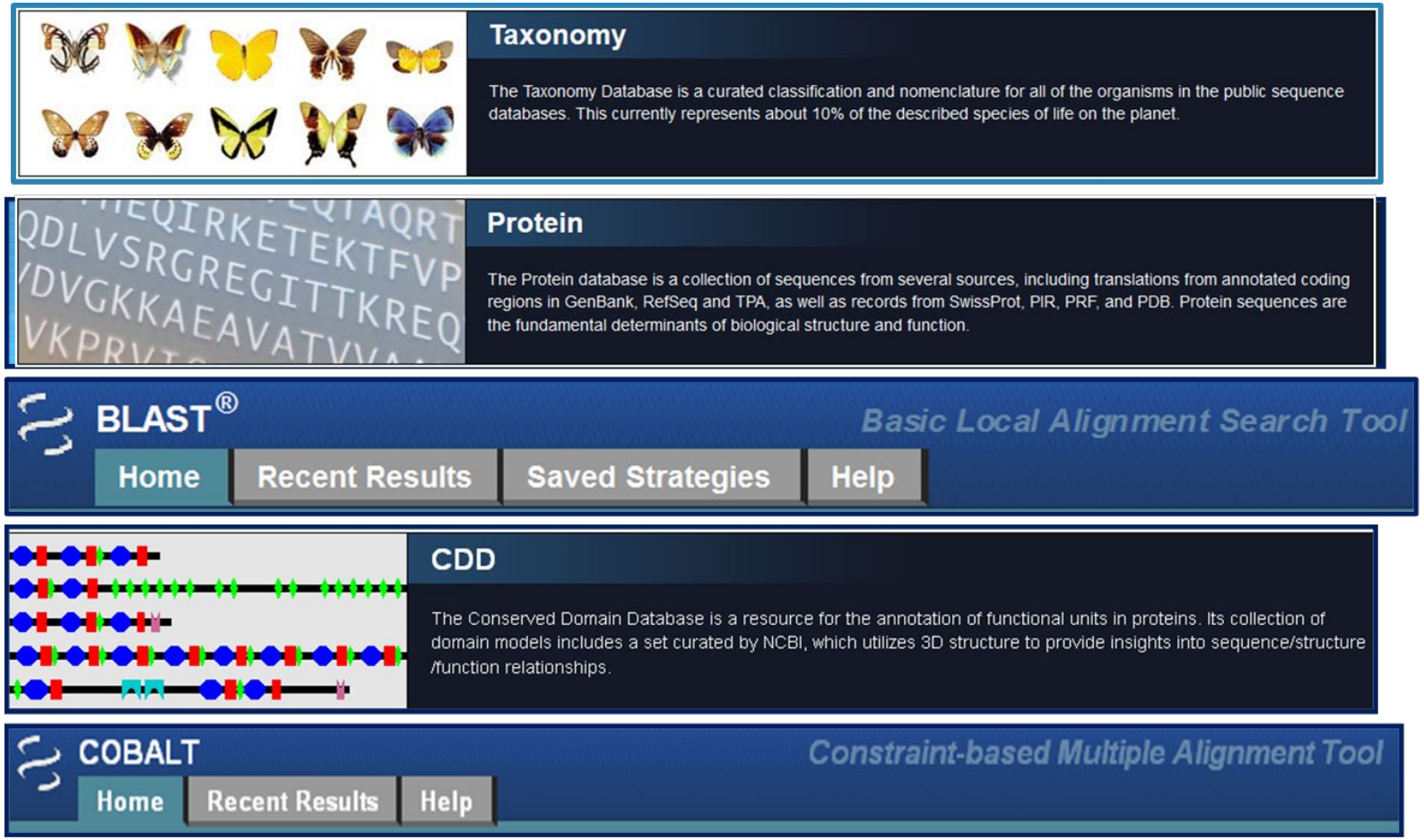
Predict Relative Intrinsic Susceptibility

- Intrinsic susceptibility can be defined as the vulnerability (or lack thereof) of an organism to chemical insult due to its inherent biological composition
  - Receptor/enzyme (protein) available for the chemical to act upon
- Relative: based on comparisons to a query protein
  - Molecular target conservation is but a component of multiple determinants of species susceptibility



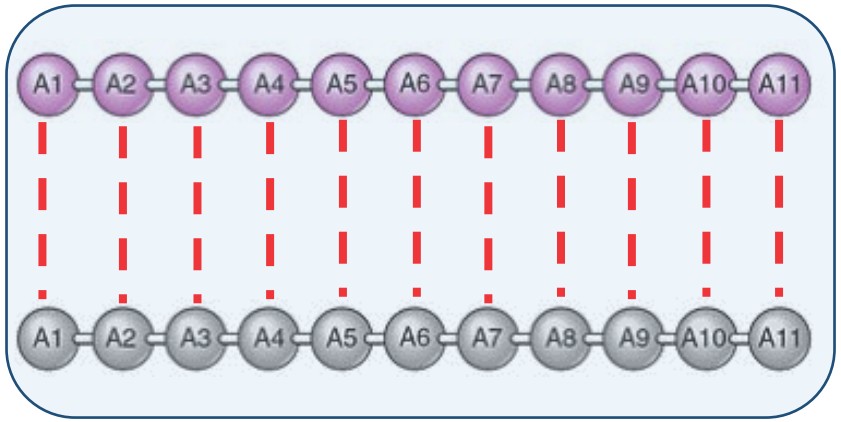
Available Databases and Tools

National Center for Biotechnology Information
Established in 1988: a division of National Library of Medicine at NIH



SeqAPASS: Level 1

Primary Amino Acid Sequence Alignment

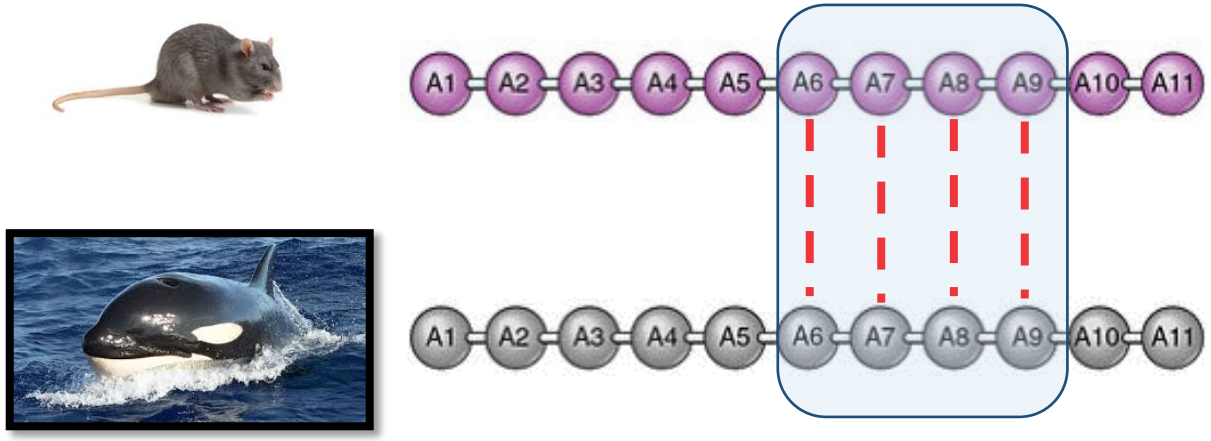


	Bit Score	Percent Similarity
Query	1241.9	100
Hit	1229.5	99.0
Hit	1223.0	98.5
Hit	1111.3	89.5
Hit	862.4	69.4

$$\text{Percent Similarity} = \frac{\text{Hit Bit Score}}{\text{Query Bit Score}} \times 100$$

SeqAPASS: Level 2

Functional Domain Sequence Alignment

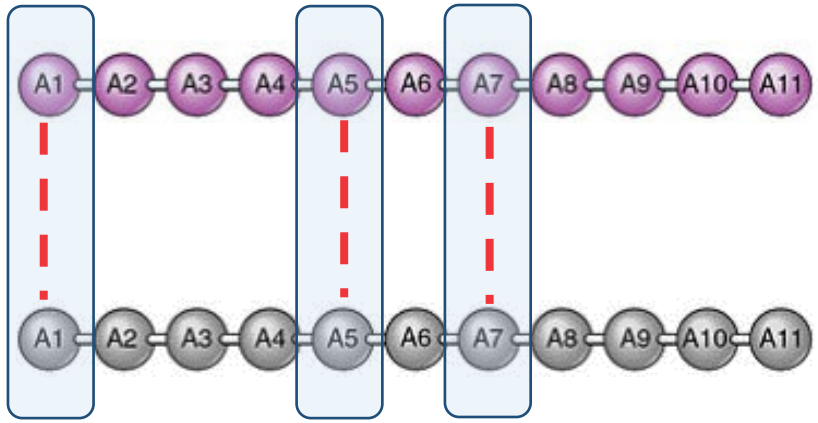


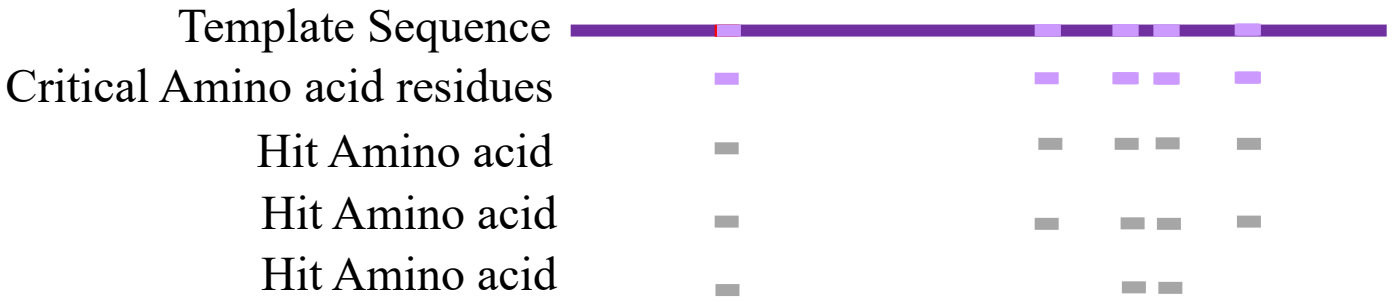
Query Sequence	Primary Amino Acid Sequence	Bit Score	Percent Similarity
Query Sequence domain		482.6	100
Hit domain		471.9	97.8
Hit domain		303.5	62.9
Hit domain		100.1	20.7

$$\text{Percent Similarity} = \frac{\text{Hit Bit Score}}{\text{Query Bit Score}} \times 100$$

SeqAPASS: Level 3

Critical Amino Acid Sequence Alignment



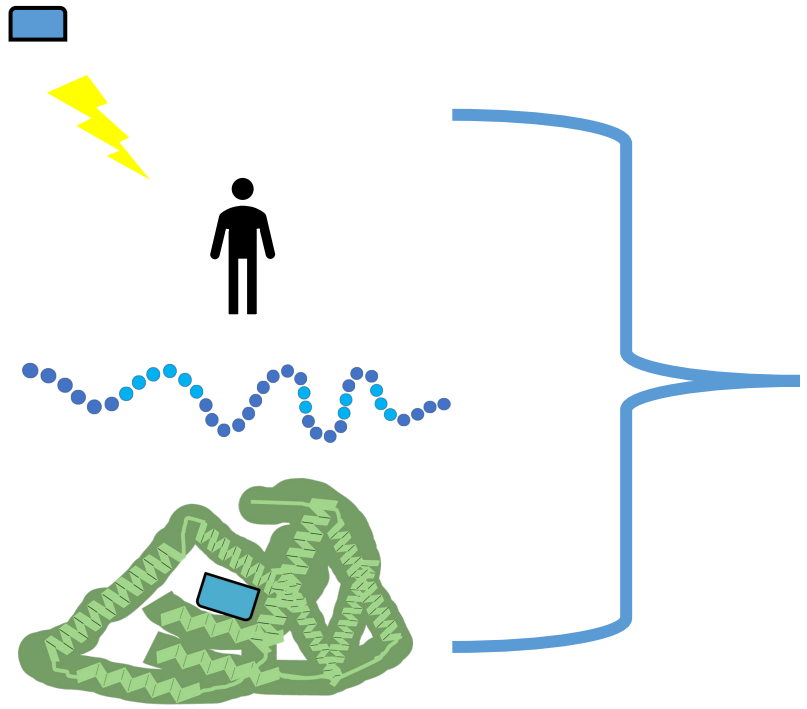


Rules to automate susceptibility prediction:

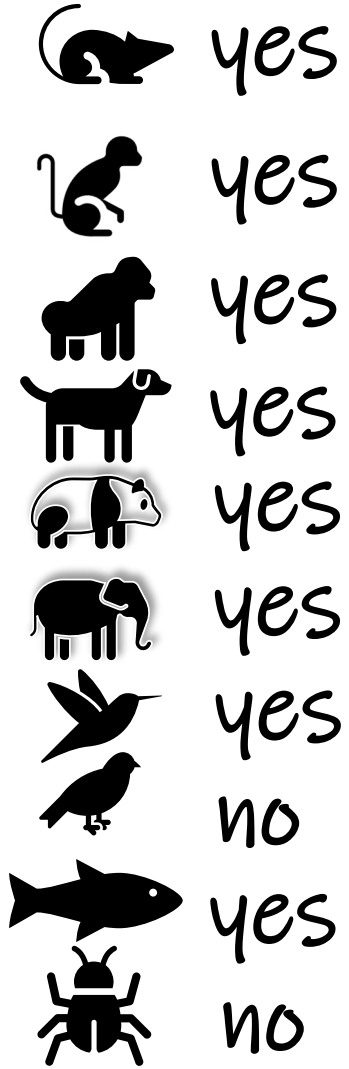
- Same side-chain classification as template
- MW as measure of size 30g/mol or less from template

Exact Match
Partial Match
Not a Match

SeqAPASS Predicts Likelihood of Similar Susceptibility based on Sequence Conservation:



yes

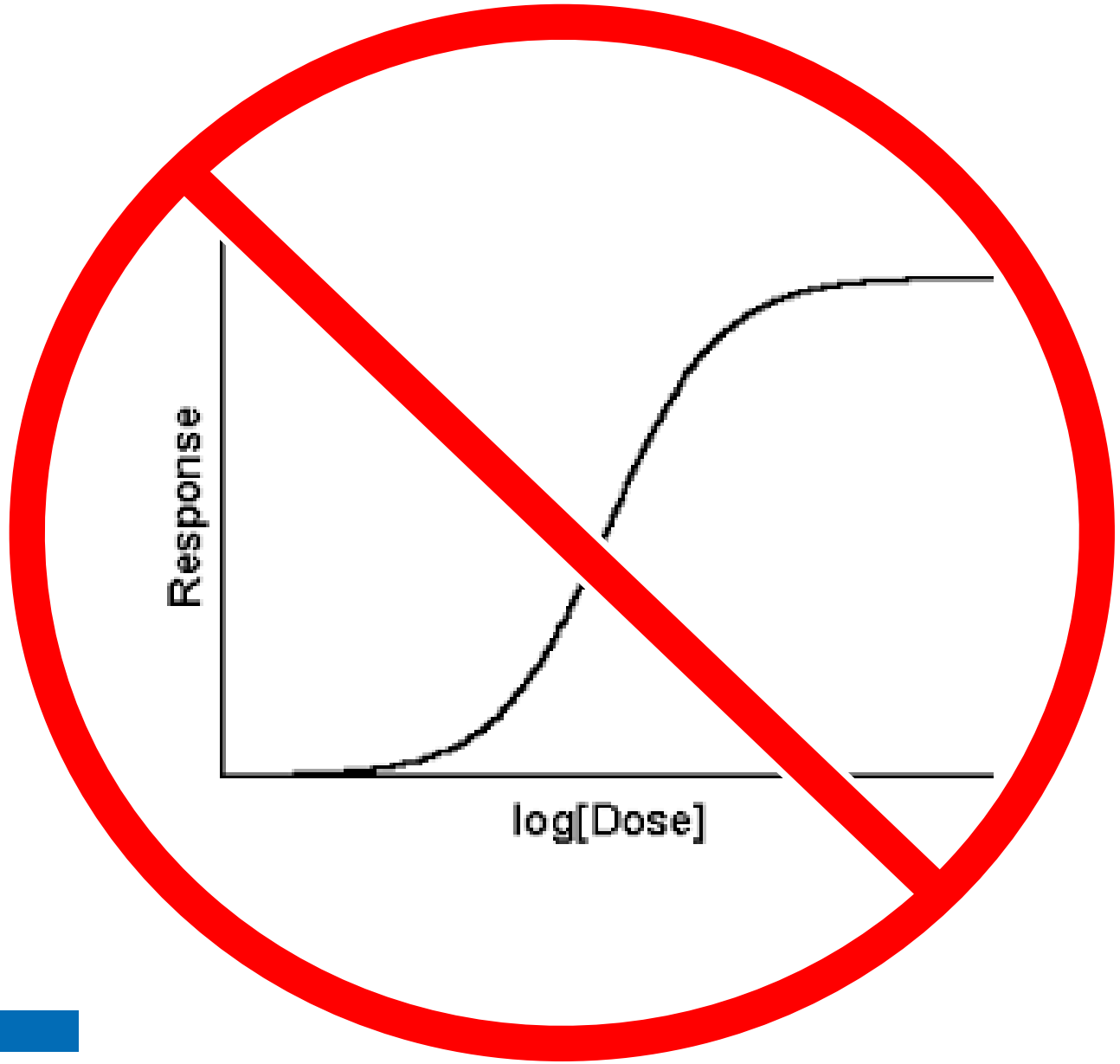


yes
yes
yes
yes
yes
yes
yes
no
yes
no

Line(s) of evidence indicate

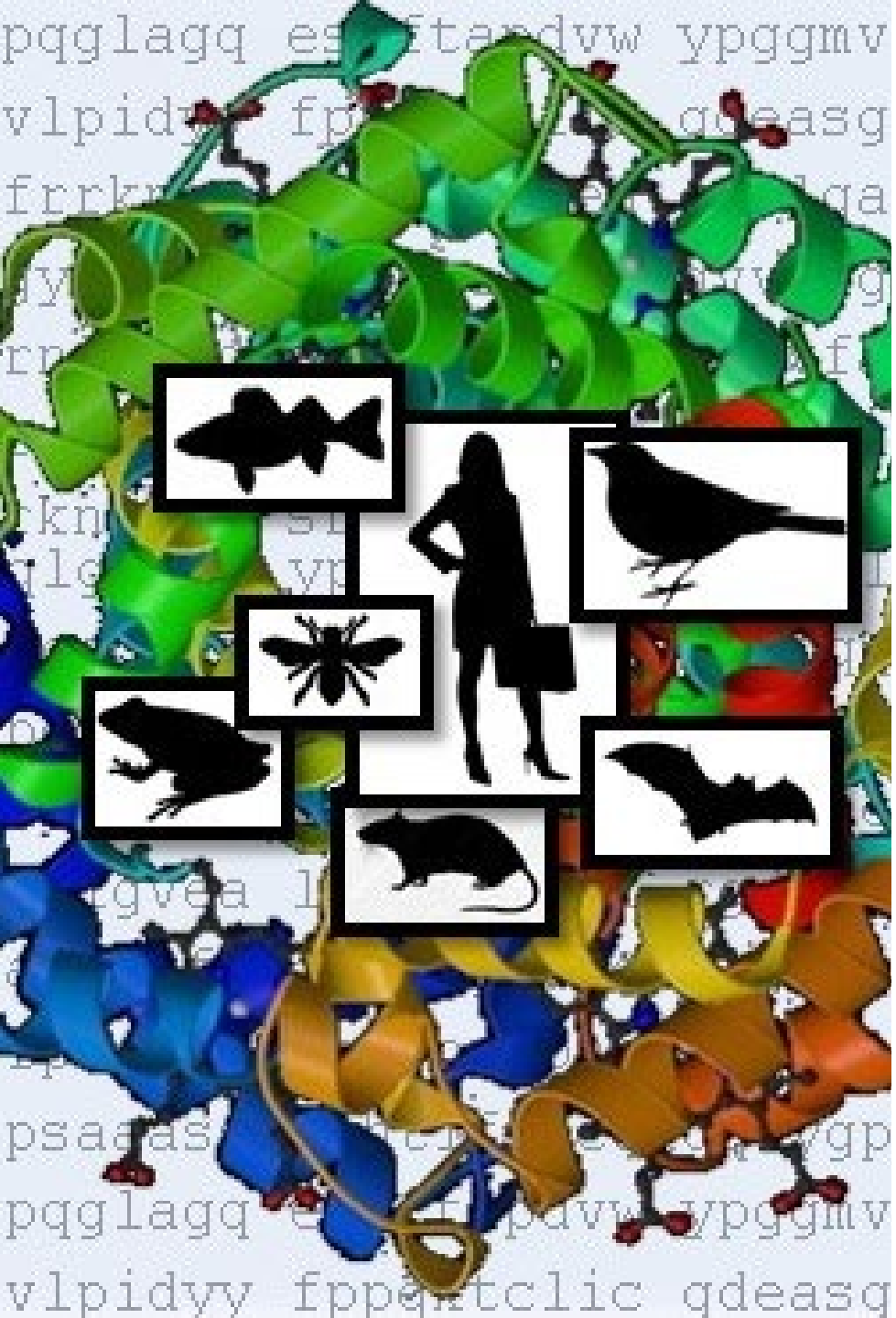
- The protein is conserved
- The protein is NOT conserved

SeqAPASS DOES NOT predict the degree of sensitivity/susceptibility:



Factors that make a species sensitive

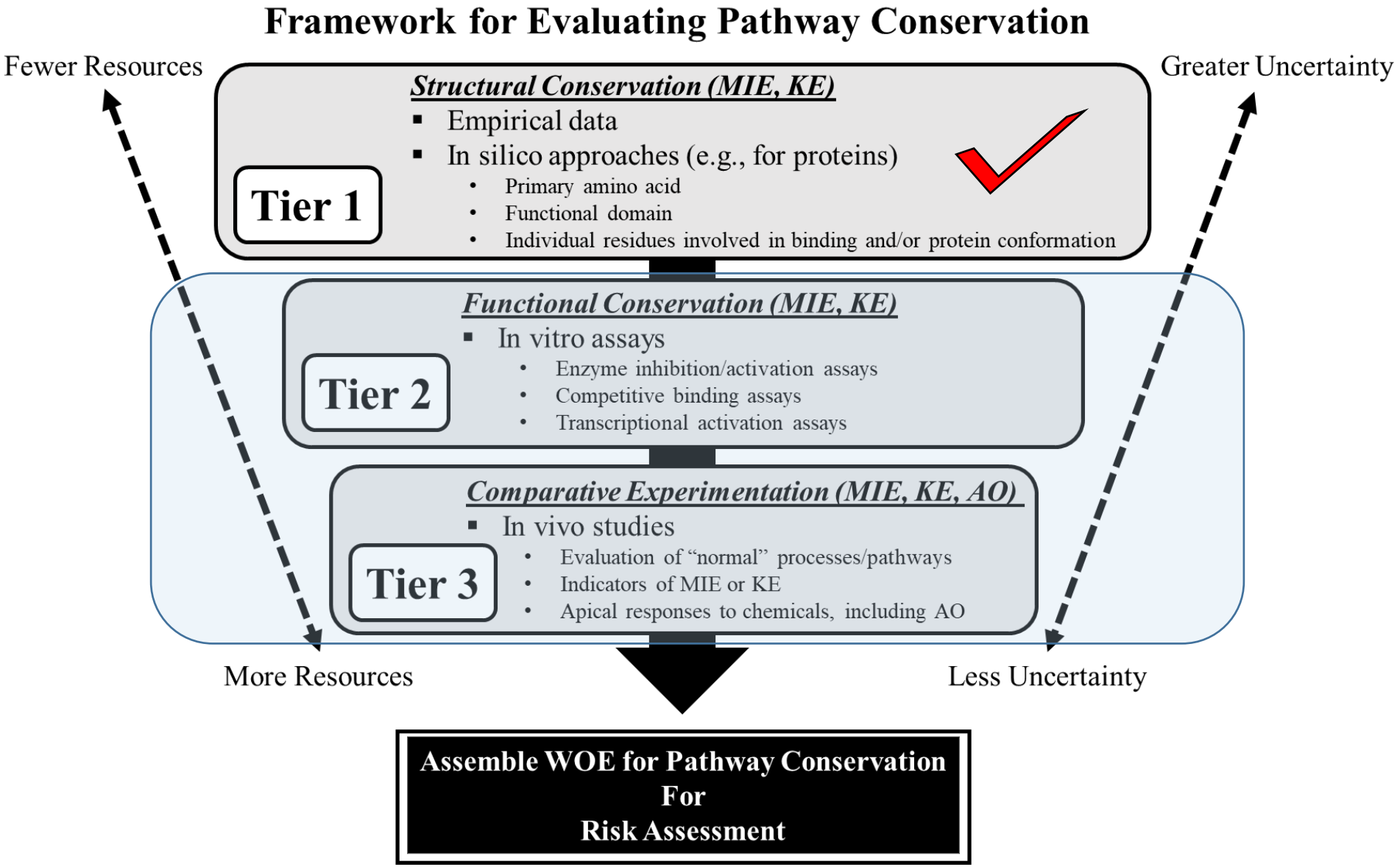
- Exposure
- Dose
- ADME
- Target receptor availability
- Life stage
- Life history
- etc.
- etc.



Strengths of SeqAPASS

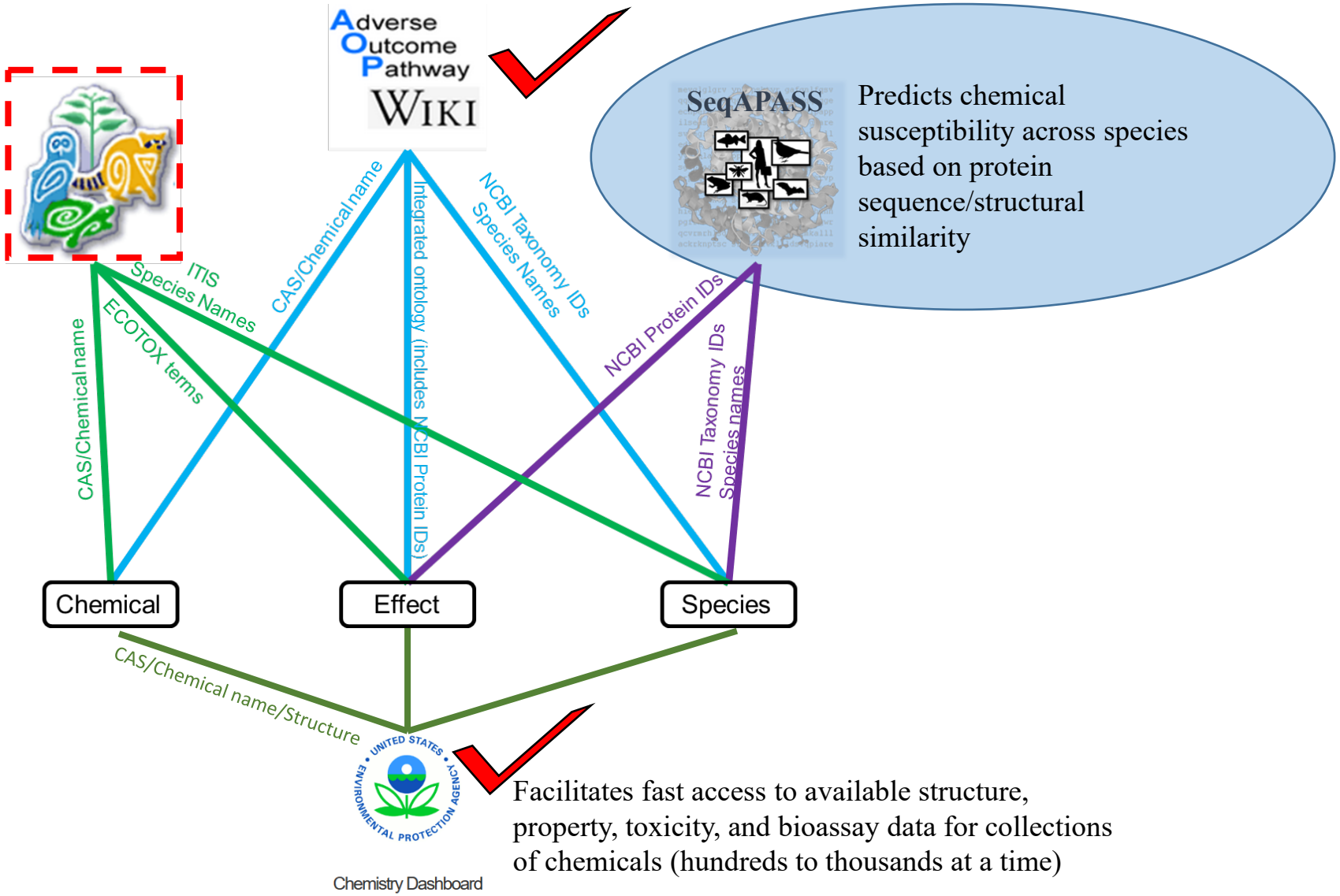
- **Publicly available** to all
- Lines of evidence for conservation for **100s-1000s of species** rapidly
- Takes advantage of **well-established tools and databases**
- Streamlined, consistent, transparent, and published methods
  - **Case examples** to demonstrate applications
- **Guides users** to appropriate input
- **Evolves** as bioinformatics approaches become more user friendly
  - Smart automation or semi-automation

Bringing Available Knowledge Together



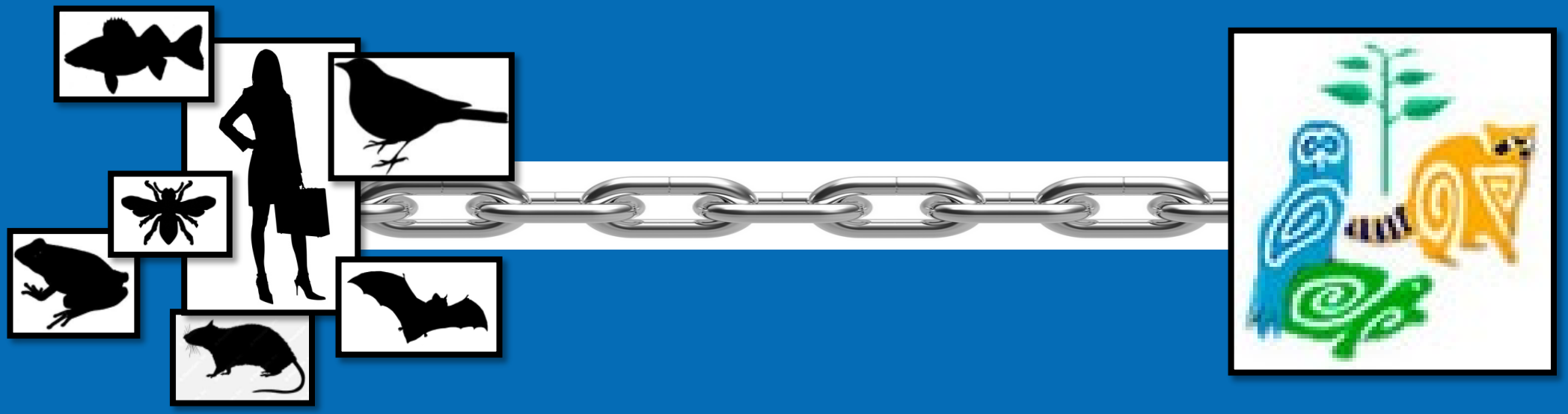
Facilitate rapid access to relevant data

A central repository for all AOPs developed



EPA
United States
Environmental Protection
Agency

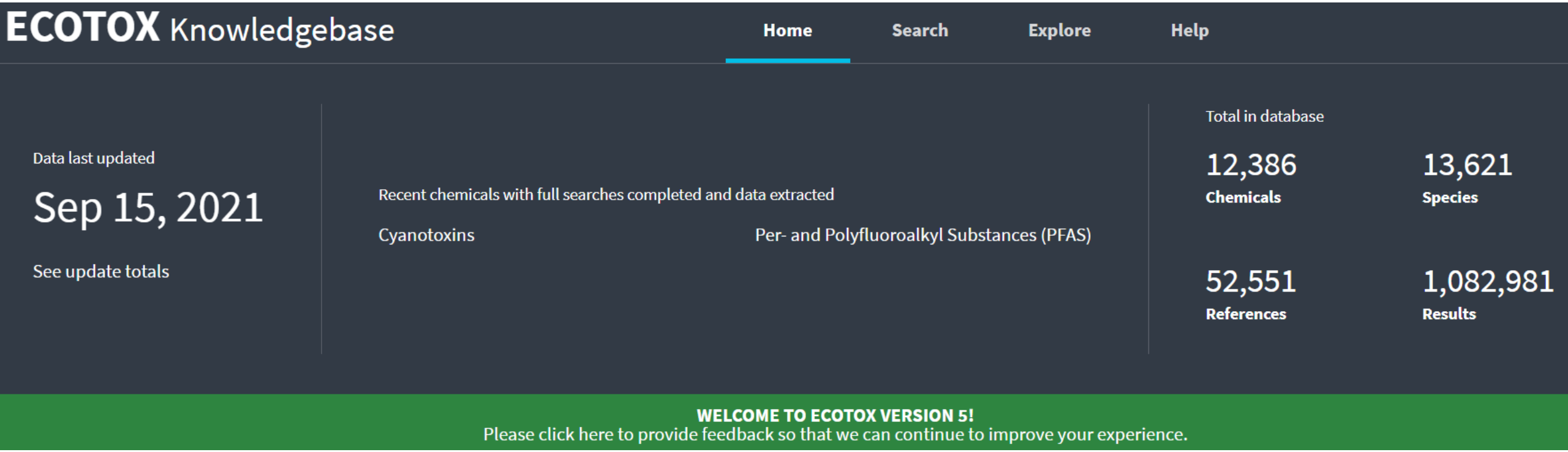
Rapidly Connect Sequence-based predictions of chemical susceptibility to Available Toxicity Data



Evidence for structural and functional conservation

ECOTOXicology (ECOTOX) Knowledgebase

30+ year history:
Originated in the early 1980s
US Environmental Protection Agency Office of Research and Development



Publicly available, curated database providing toxicity data from singe-chemical exposure studies to aquatic life, terrestrial plants, and wildlife
From comprehensive search and review of open and grey literature

- Data extracted from acceptable studies with up to 250 fields
- Updated quarterly

ECOTOX Knowledgebase

or

Applicability of Studies

- **Paper must meet these criteria**
  - **Single chemical exposure**
  - **Ecologically-relevant species**
  - **Must be able to verify CAS registry numbers**
  - **Must be able to verify taxonomic information for test species**
  - **Exposure to live organism, viable tissue or cells**
  - **Report concurrent exposure concentration, dose or application rate**
  - **Report duration of exposure**
  - **Primary source of the data**
  - **Study must be a full article in English**
- **The following studies are excluded**
  - **Air pollution studies related to CO2 and ozone**
  - **Studies on humans, monkeys, bacteria, viruses and yeast**
  - **Review and summary articles**
  - **Terrestrial studies with an inhalation route of exposure**
  - **Non-English publications and abstracts**

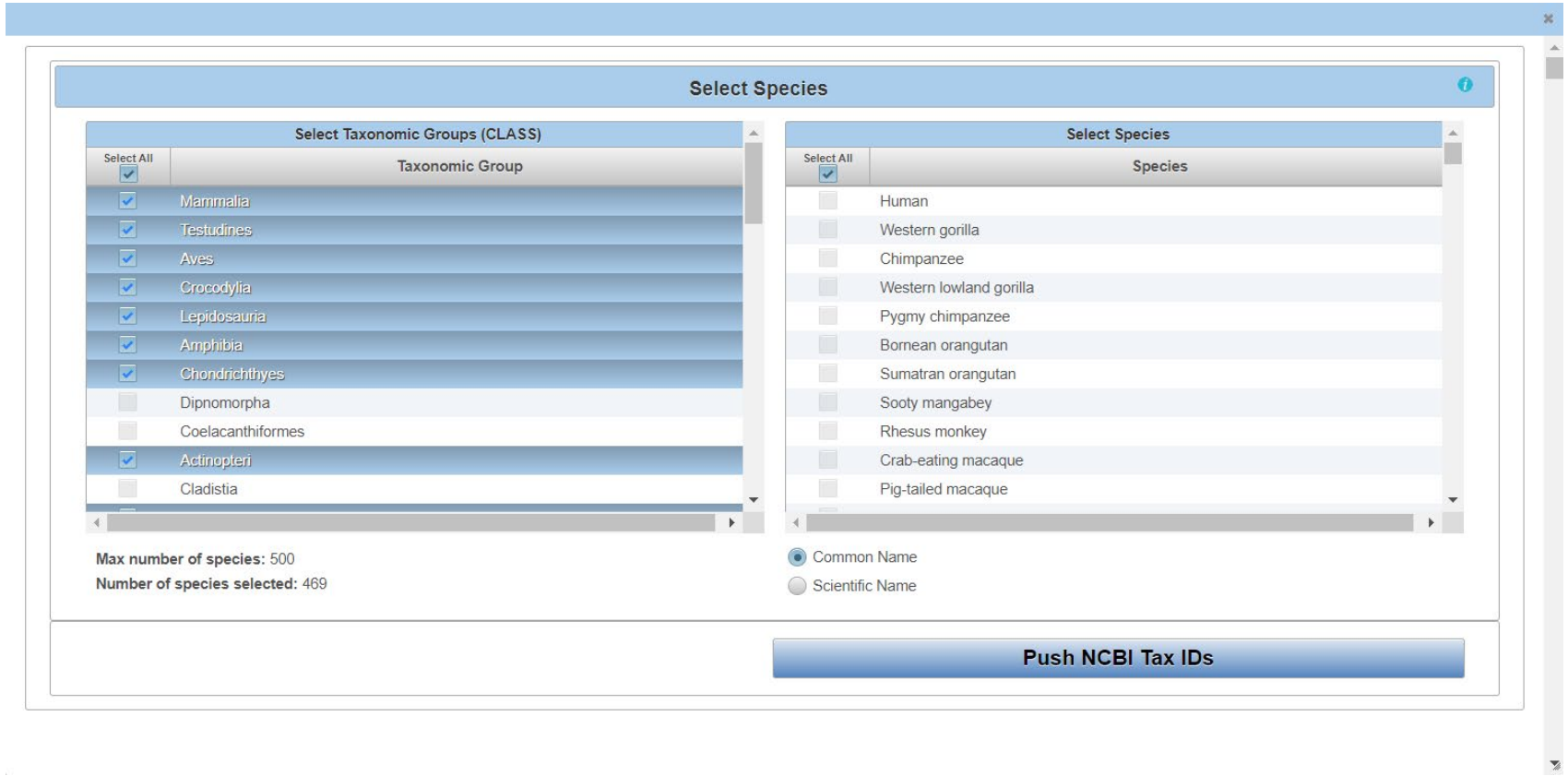
Rapidly connect predictions to empirical data

- 30+ year history of supporting EPA (Originated in the early 80's)
- Comprehensive, publicly available, curated database providing chemical environmental toxicity data on aquatic and terrestrial wildlife, including plants from the open literature
  - Curated data from 52,500+ publications
  - Species, Chemicals and number of records added each year

Rapidly connect predictions to empirical data

Choose Taxa/Species

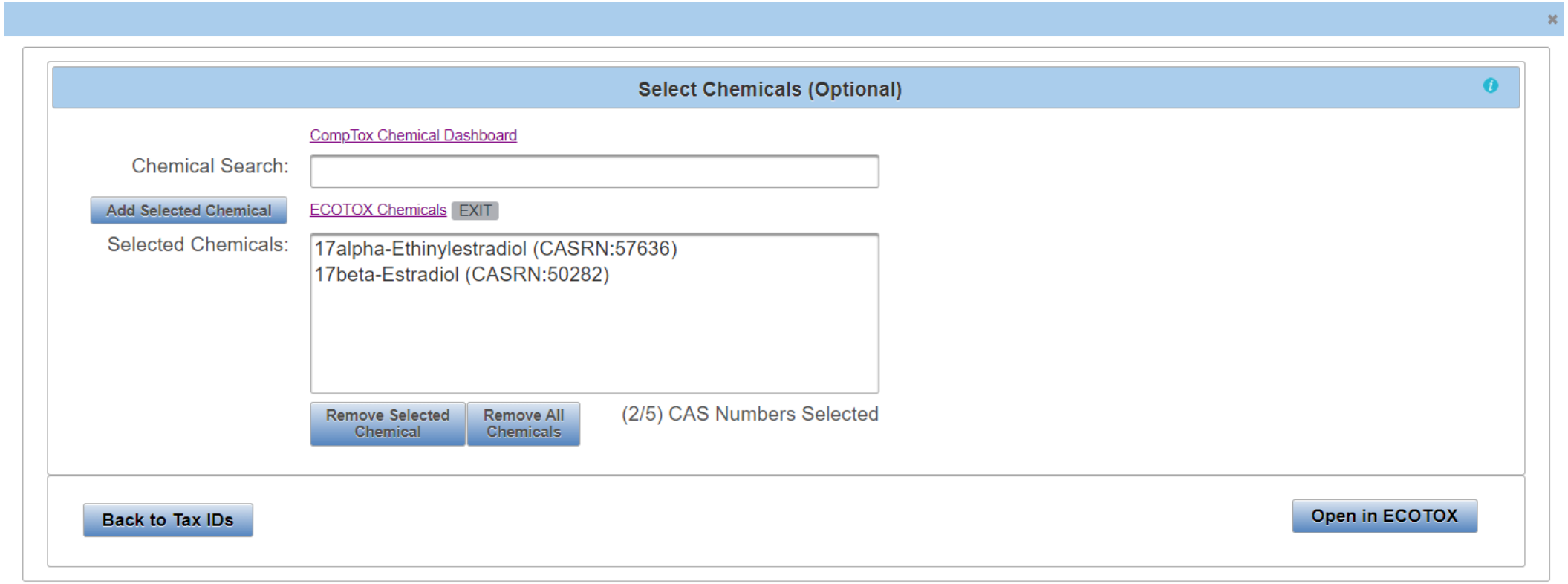
Note: Must be less than 500 species to push to ECOTOX



Rapidly connect predictions to empirical data

Select Chemicals

Note: Must be less than 5 chemicals to push to ECOTOX



EPA United States Environmental Protection Agency

Rapidly connect predictions to empirical data

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 (1)
All

Chemicals (2)
2 Selected

Class (14)
All

Order (32)
All

89 Species

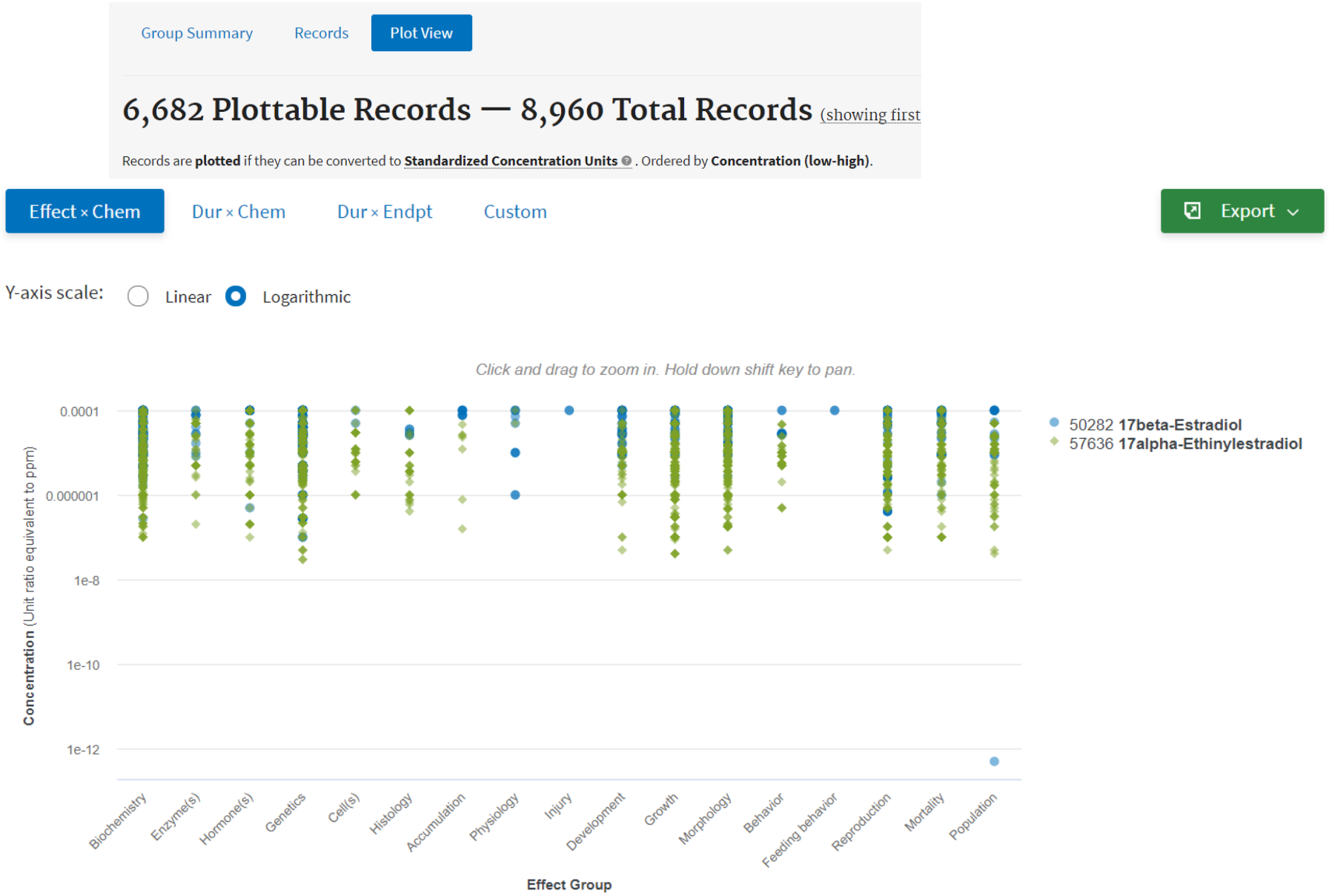
Export CSV

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

*Showing **all** 89 species from **Acanthogobius flavimanus** to **Zoarces viviparus***

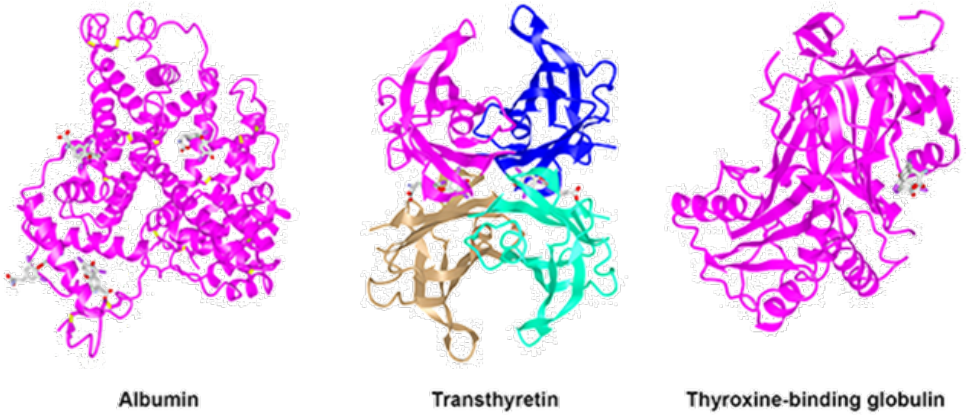
SCIENTIFIC NAME	COMMON NAME	RECORDS	PUBLICATIONS	YEAR MIN	YEAR MAX
type to filter...	...	...	...	...	...
Acanthogobius flavimanus	Yellowfin Goby	17	3	2003	2004
Alligator mississippiensis	American Alligator	28	3	1997	1999
Anguilla anguilla	Common Eel	17	6	1991	2004
Anguilla japonica	Japanese Eel	28	1	2012	2012
Balanus amphitrite	Striped Barnacle	7	1	1998	1998
Betta splendens	Siamese Fighting Fish	19	2	2006	2010

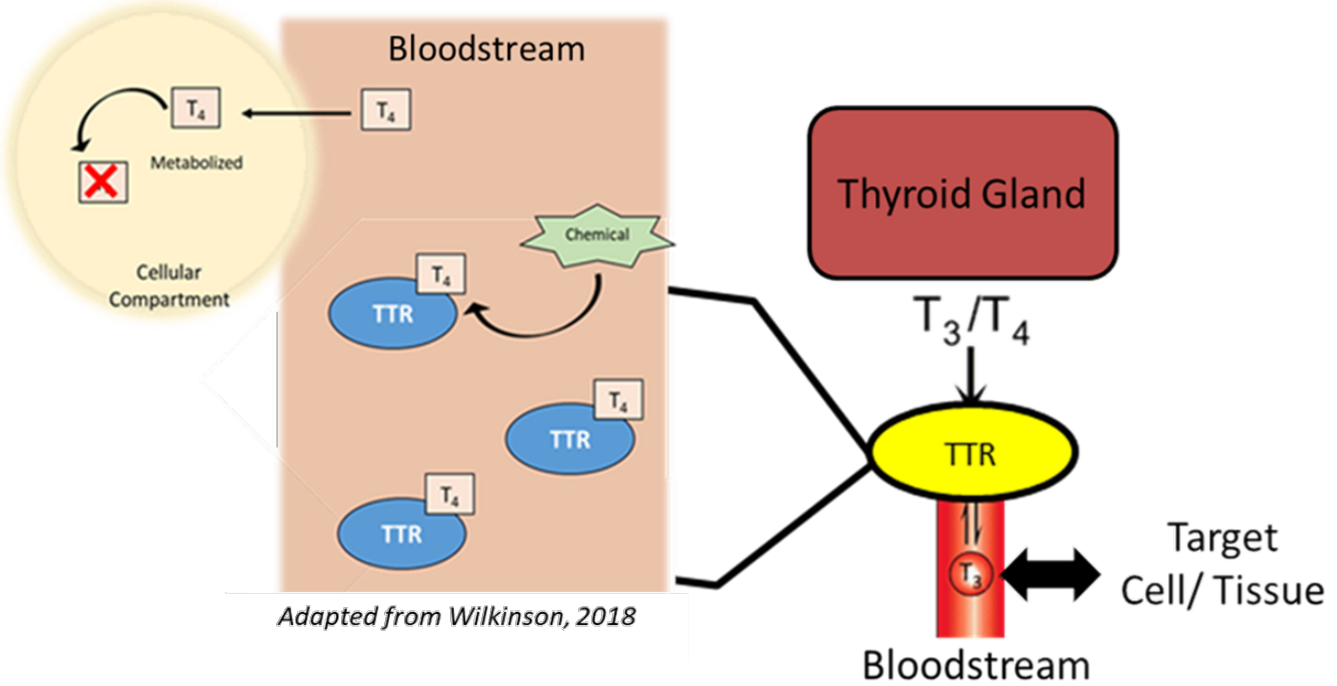
Rapidly connect predictions to empirical data



Taxonomic Domain for Adverse Outcome Pathways: Transthyretin

- **Thyroid hormone distribution** proteins, albumin, **transthyretin (TTR),** and thyroxine-binding globulin, are responsible for transporting thyroid hormone (TH) from the thyroid gland **throughout the body to target cells**
- Environmental **chemicals can competitively bind to TTR**, alter hormone homeostasis, and disrupt the thyroid axis
- In silico, in vitro, and in vivo evidence suggests TTR-binding may be an MIE of relevance for many environmental chemicals including some PFAS compounds





Adapted from Wilkinson, 2018

Transthyretin Adverse Outcome Pathways

- Two putative adverse outcome pathways have been developed for competitive binding to transthyretin leading to decreased serum thyroid hormone levels → **How well can we extrapolate these pathways across species?**

Sequence	Type	Event ID	Title	Short name
1	MIE	957	Binding, Transthyretin in serum	Binding, Transthyretin in serum
2	KE	958	Displacement, Serum thyroxine (T4) from transthyretin	Displacement, Serum thyroxine (T4) from transthyretin
3	KE	959	Increased, Free serum thyroxine (T4)	Increased, Free serum thyroxine (T4)
4	KE	960	Increased, Uptake of thyroxine into tissue	Increased, Uptake of thyroxine into tissue
5	KE	961	Increased, Clearance of thyroxine from serum	Increased, Clearance of thyroxine from serum
6	KE	281	Thyroxine (T4) in serum, Decreased	T4 in serum, Decreased
7	KE	280	Thyroxine (T4) in neuronal tissue, Decreased	T4 in neuronal tissue, Decreased
8	KE	756	Hippocampal gene expression, Altered	Hippocampal gene expression, Altered
9	KE	757	Hippocampal anatomy, Altered	Hippocampal anatomy, Altered
10	KE	758	Hippocampal Physiology, Altered	Hippocampal Physiology, Altered
11	AO	402	Cognitive Function, Decreased	Cognitive Function, Decreased

AOP 152: Interference with thyroid serum binding protein transthyretin and subsequent adverse human neurodevelopmental toxicity

Sequence	Type	Event ID	Title	Short name
1	MIE	957	Binding, Transthyretin in serum	Binding, Transthyretin in serum
2	KE	1830	Displacement, Serum thyroxine (T4) from carrier protein	Displacement, Serum thyroxine (T4) from carrier protein
3	KE	959	Increased, Free serum thyroxine (T4)	Increased, Free serum thyroxine (T4)
4	KE	1158	Increased, Hepatic thyroid hormone uptake/transport	Increased, Hepatic thyroid hormone uptake/transport
5	KE	401	Increase, Biliary excretion TH glucuronide	Increase, Biliary excretion TH glucuronide
6	KE	281	Thyroxine (T4) in serum, Decreased	T4 in serum, Decreased
7	AO	1101	Altered, Amphibian metamorphosis	Altered, Amphibian metamorphosis

366: Competitive binding to thyroid hormone carrier protein

Goal: Assess transthyretin (TTR) as a protein target for environmental chemicals and define sensitive species/ populations

Transthyretin (TTR) & Cross Species Considerations

Thyroid Binding Globulin (TBG)	High affinity for TH
Transthyretin (TTR)	Moderate affinity for TH
Albumin	Low affinity for TH

Taxa	THDP Present Development	THDP Present Adult	TTR TH Affinity
Mammals*	TBG, TTR, Albumin	TBG, TTR, Albumin	T4 > T3
Fish	TTR, Albumin	Albumin	T3 > T4
Amphibian	TTR, Albumin	Albumin	T3 > T4
Reptile	TTR, Albumin	Albumin	T3 > T4
Bird	TTR, Albumin	Albumin	T3 > T4

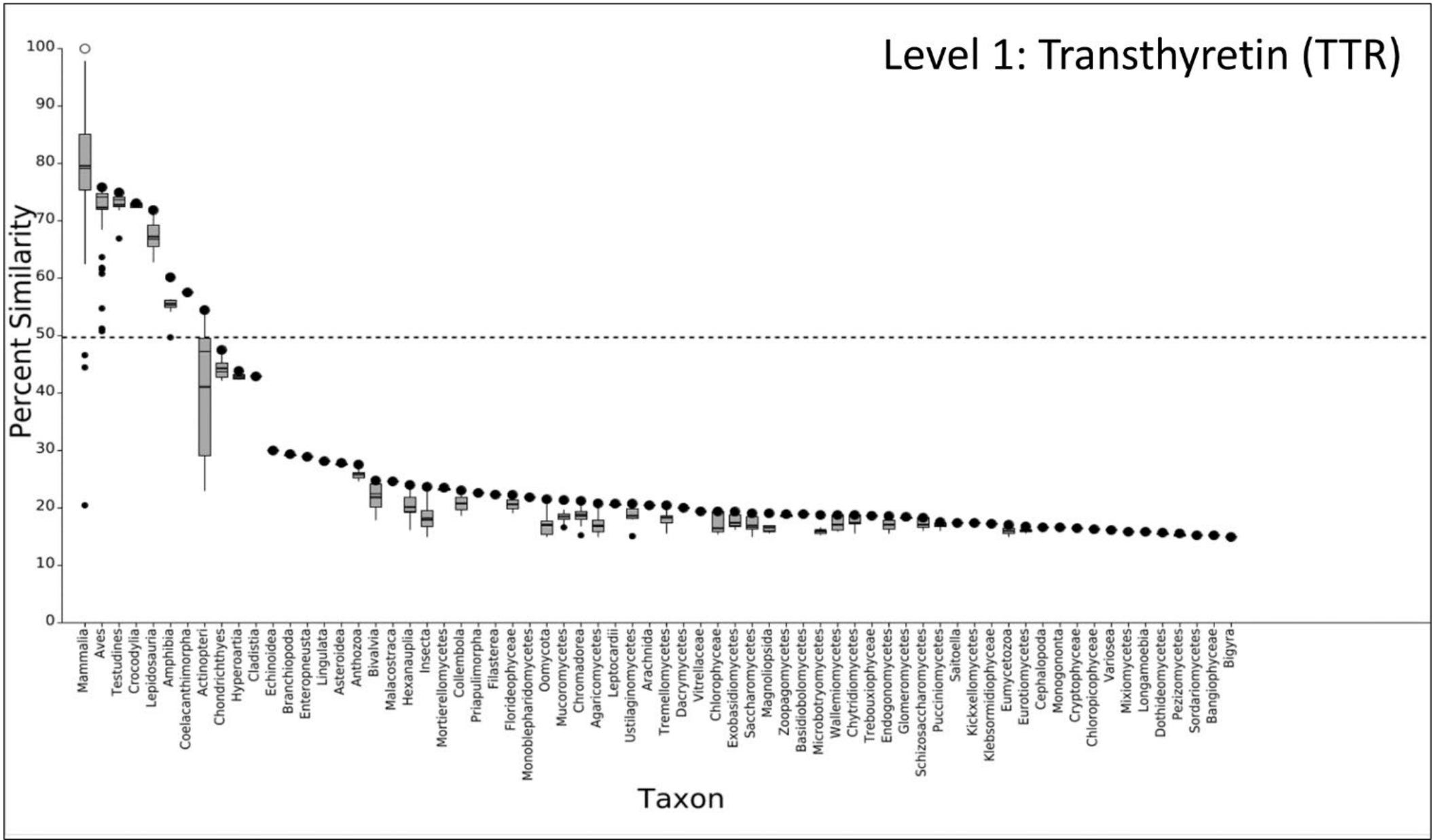
(Schreiber 2002, Rabah et al. 2019, McLean et al. 2017))

- In non-mammalian vertebrates, TTR is important only during developmental stages
- TTR is **less-important** in mammals due to the presence of other distribution proteins
- Structural differences exist between mammals and non-mammalian vertebrates resulting in different TH affinity

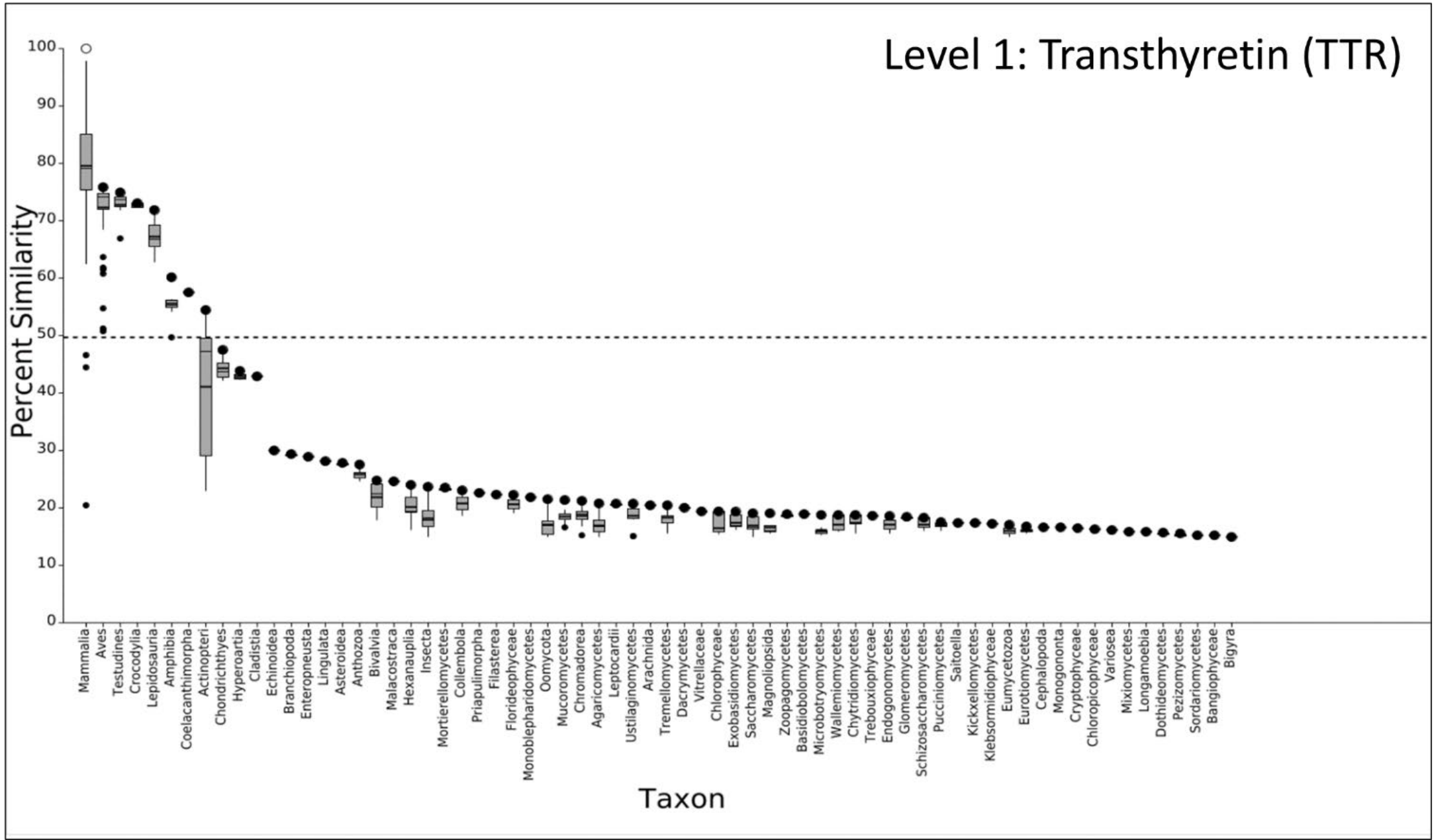
Early life-stages of **fish**, **amphibians**, and **reptiles** are the population potentially **most sensitive** to TTR-binding compounds

EPA United States Environmental Protection Agency

SeqAPASS Level 1



SeqAPASS Level 2

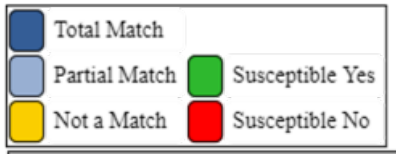


SeqAPASS Level 2

A

Taxonomic Group	Number of Species	Majority Similar Susceptibility?	Number Y	Number N
Mammals	135	Y	129	6
Birds	79	Y	79	0
Fish	53	Y	49	4
Reptiles	23	Y	23	0
Amphibians	4	Y	3	1

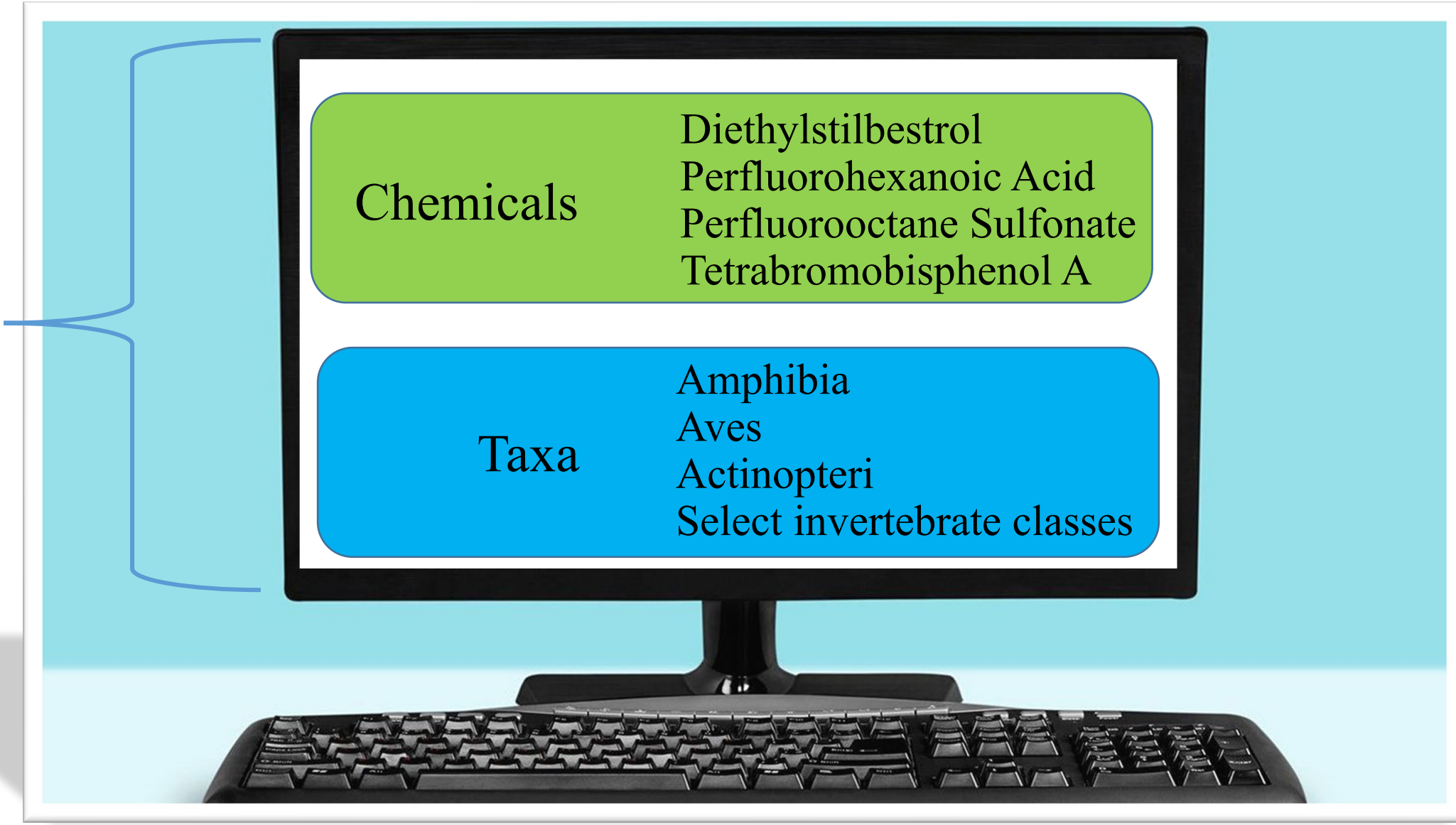
B



Common Name	Similar Susceptibility	Amino Acid 1	Amino Acid 2	Amino Acid 3	Amino Acid 4	Amino Acid 5	Amino Acid 6
Human	Y	35K	128A	129A	130L	137S	139T
Sperm whale	N	35K	128M	129A	130L	137S	139T
Florida manatee	N	35K	128T	129A	130L	137S	139T
Vaquita	N	35K	127M	128A	129L	136S	138T
Yangtze finless porpoise	N	35K	127M	128A	129L	136S	138T
Narwhal	N	35K	126M	127A	128L	135S	137T
Platypus	N	37K	130A	131T	132L	139S	141T
Pike-perch	N	39K	132A	133M	134L	141S	143T
Blunt-snouted clingfish	N	32H	125P	126L	127L	134S	136Y
Black rockcod	N	10H	103P	104L	105L	112S	114Y
Turquoise killifish	N	11H	104P	105L	106L	113S	115Y
Gabon caecilian	N	40K	133A	134L	135F	142I	144T
Chinese bamboo-partridge	N	12K	105T	106T	107V	114S	116T

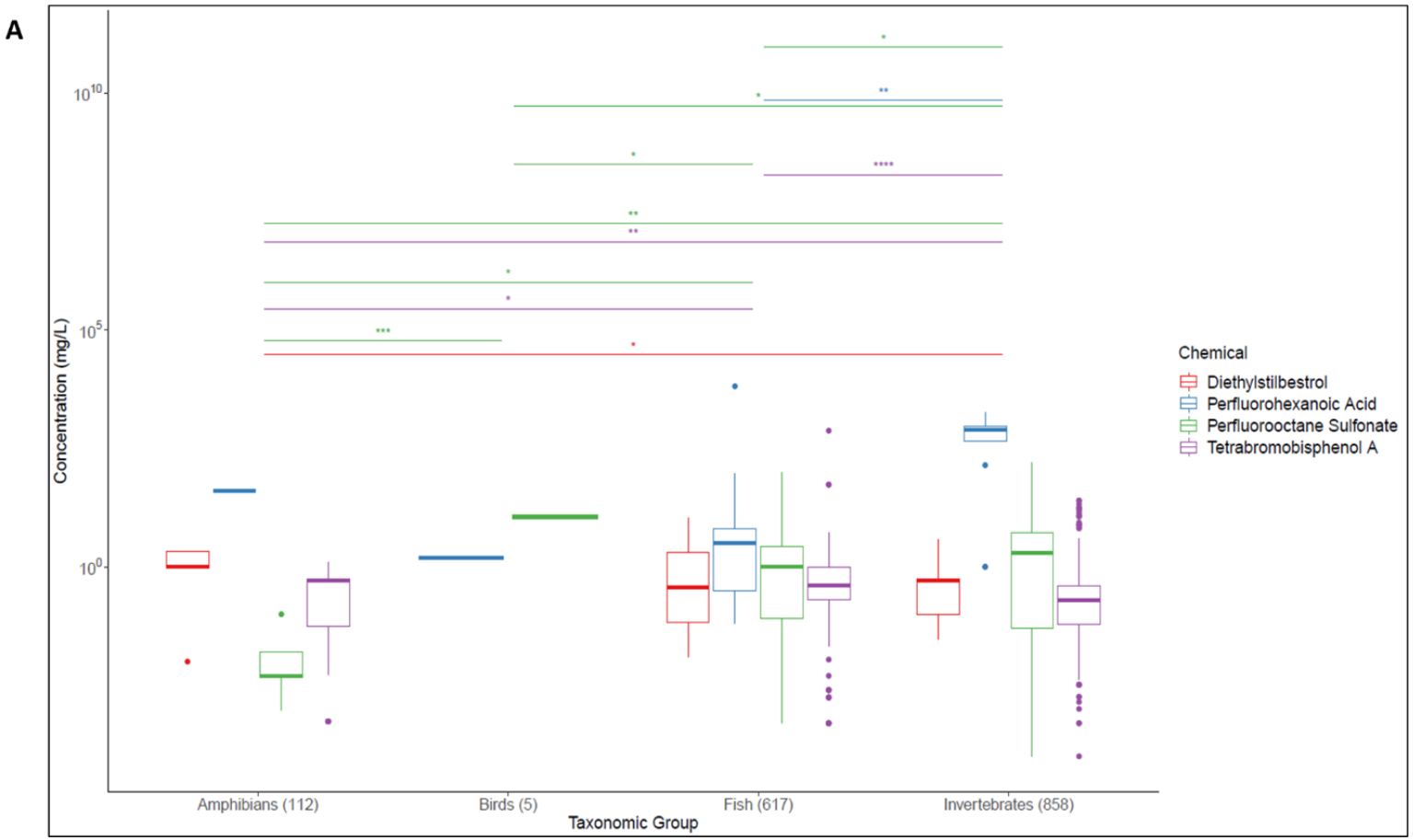
Link resources to gather all lines of evidence for pathway conservation across species

Info for ECOTOX Widget in SeqAPASS

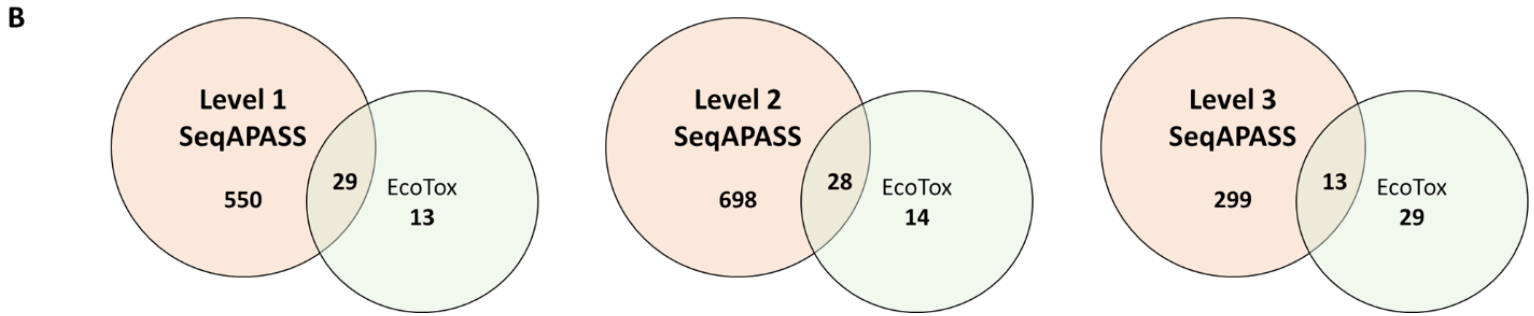


Combine cross species knowledge

Mean effect concentrations across taxonomic groups with data available in the ECOTOX Knowledgebase for select chemicals known to bind to the human transthyretin (TTR) protein



Species overlap



EPA United States Environmental Protection Agency

Combine cross species knowledge

	SeqAPASS L1			SeqAPASS L2			SeqAPASS L3			EcoTox*
	Yes	No	Total	Yes	No	Total	Yes	No	Total	Total
Amphibian	7	0	7	8	0	8	3	1	4	8
Bird	111	0	111	113	0	113	79	0	79	1
Fish	111	5	116	115	7	122	49	4	53	16
Reptile	26	0	26	27	0	27	23	0	23	0

**Total represents data for unique species across DES, PFOS, PFHxS, and TBBPA*

Acknowledgements

U.S. EPA, ORD

Marissa Jensen (University of Minnesota Duluth)
Sally Mayasich (University of Wisconsin)
Donovan Blatz (ORISE)
Monique Hazemi (ORISE)
Sara Vliet (US EPA)
Jon Doering (U of Lethbridge)
Colin Finnegan (Iowa State University)

GDIT

Thomas Transue
Cody Simmons
Audrey Wilkinson
Wilson Menendez

SeqAPASS v6.0 (Released Sept. 2021)

LaLone.Carlie@epa.gov

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

EPA
United States
Environmental Protection
Agency

SeqAPASS Live Demo

Interoperability to aid users