New Frontiers in Toxicology: Predictive Tools for Cross-Species Extrapolation

Sara M. Vliet, Ph.D.

ORISE Postdoctoral Fellow
Great Lakes Toxicology & Ecology Division
Office of Research and Development
U.S Environmental Protection Agency

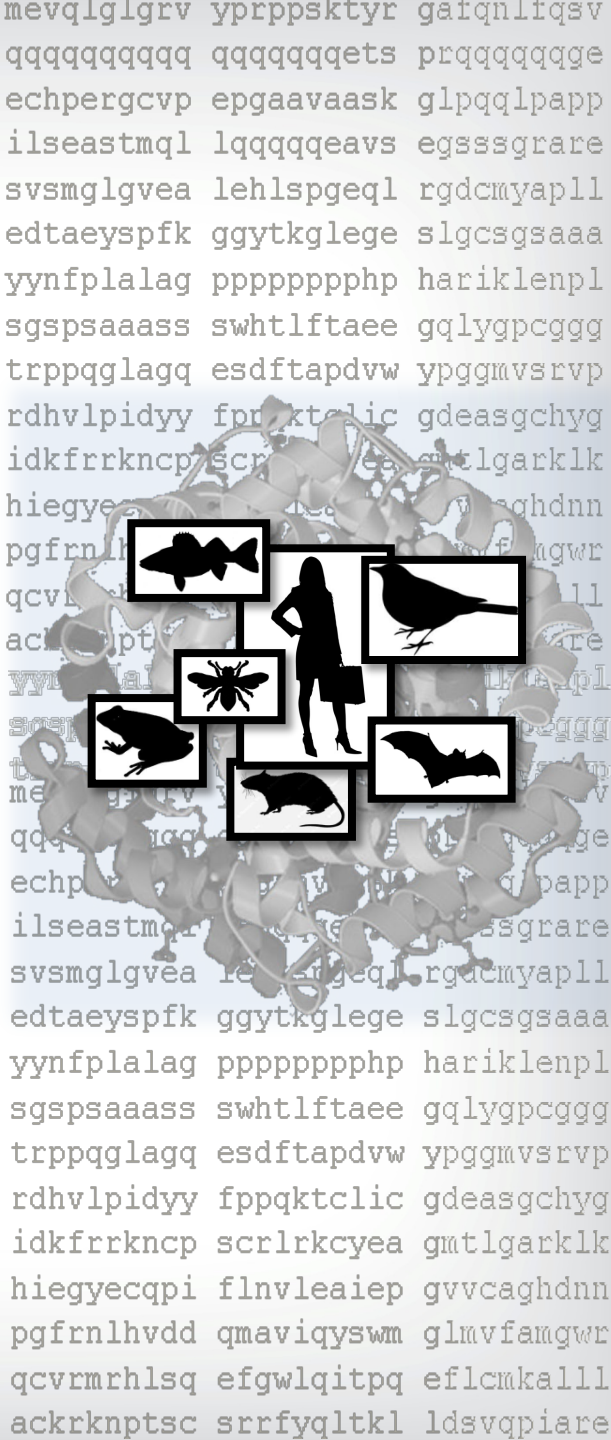
September 2nd, 2020
University of Cincinnati
Department of Environmental and Public Health Sciences Seminar

The views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of the US EPA.

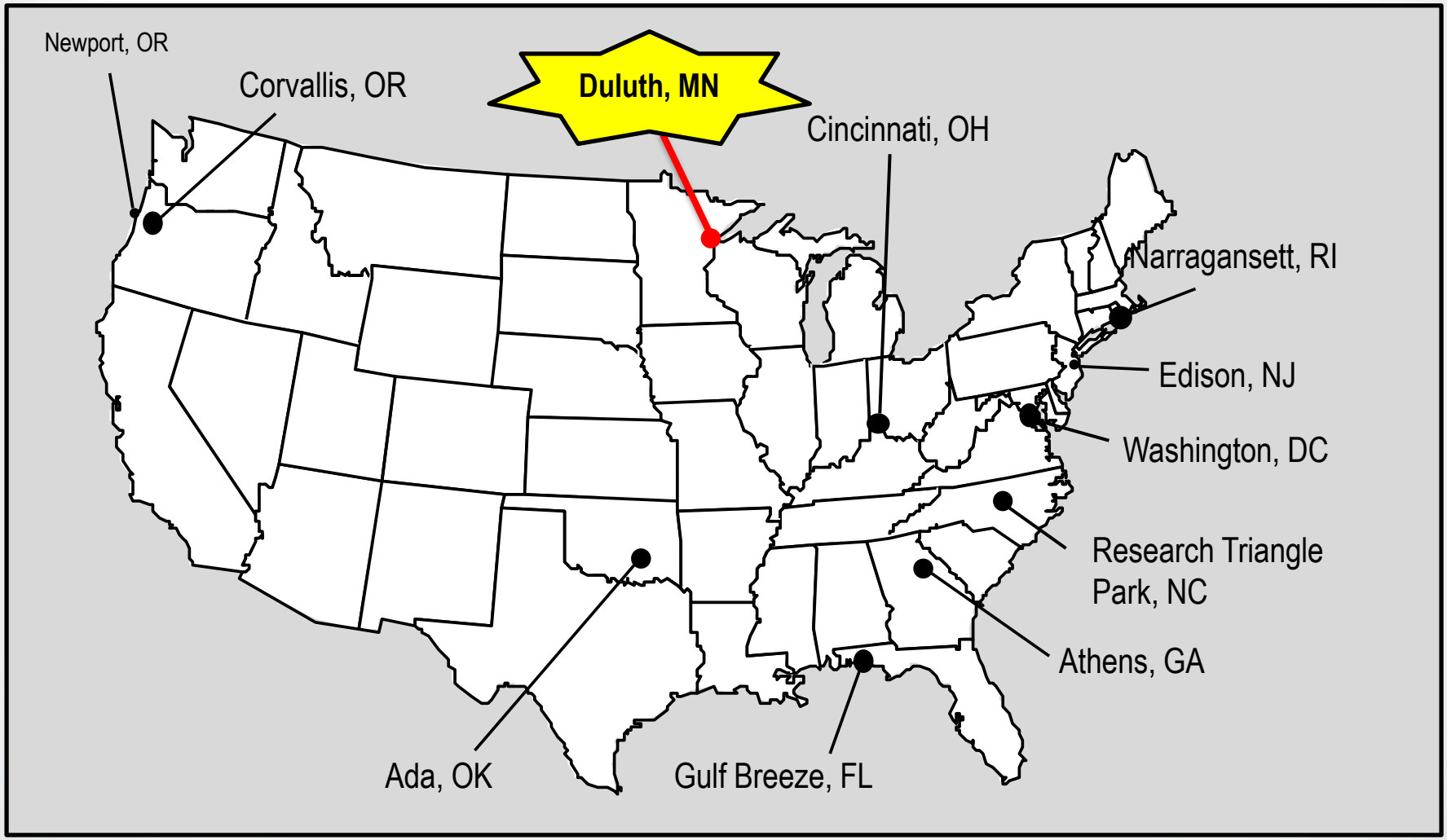
New Frontiers in Toxicology: Predictive Tools for Cross-Species Extrapolation

Outline

1. Introduction: EPA and ORISE Missions
2. Toxicity Testing: The Need for Cross-Species Understanding
3. SeqAPASS Tool
4. Androgen Receptor: A Case-Study
5. Working Towards “Validation” via Systematic Review



Office of Research and Development



Great Lakes Toxicology & Ecology Division

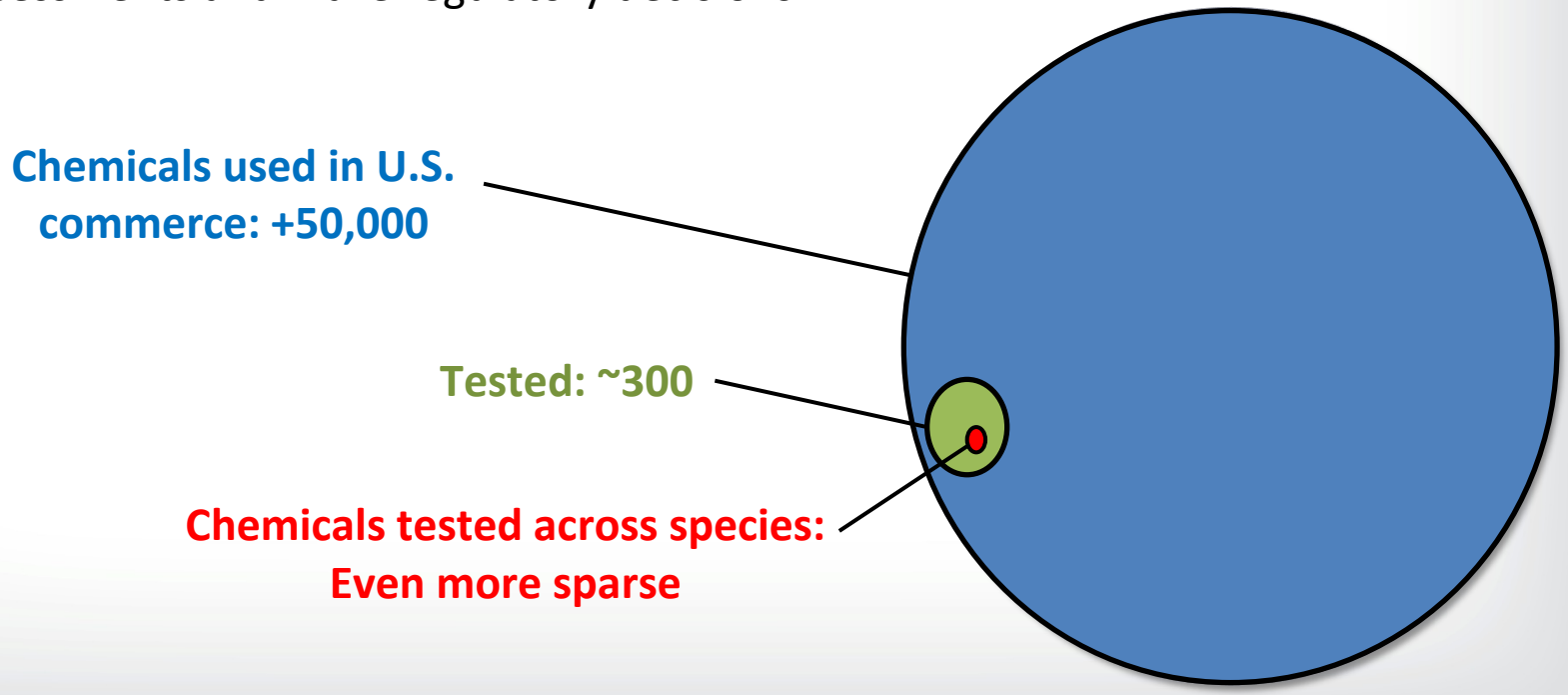
- Freshwater ecology
- Freshwater ecotoxicology
- Predictive and computational toxicology

SCIENCE STILL PROCESSING

- **ORISE Postdoctoral Fellow**
- ORISE: U.S. Department of Energy asset contract dedicated to enabling critical scientific, research, and health initiatives
- At EPA, ORISE Fellows participate in project-specific research, current environmental research, and developmental activities
- Positions focused on <u>Training opportunities, scientific growth, and professional development!</u>

EPA: The Mission

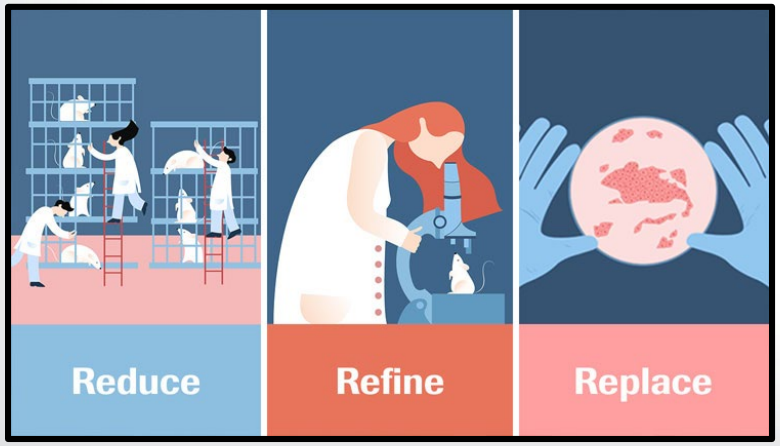
- **Protect Human Health and the Environment**
  - Ensure that chemicals in the marketplace are safe
  - Evaluate the ecological and human health risks associated with toxic chemicals
- **Only a tiny fraction of the compounds around us have been tested for safety**
  - Without information on a chemical's toxicity, it's impossible to conduct informed risk-assessments and make regulatory decisions



Historical Toxicity Testing

- Whole Animal Exposures
- Use of surrogate species
- Observe Toxic Outcome
- E.g. tumor development, mortality, reproduction, malformations, etc.

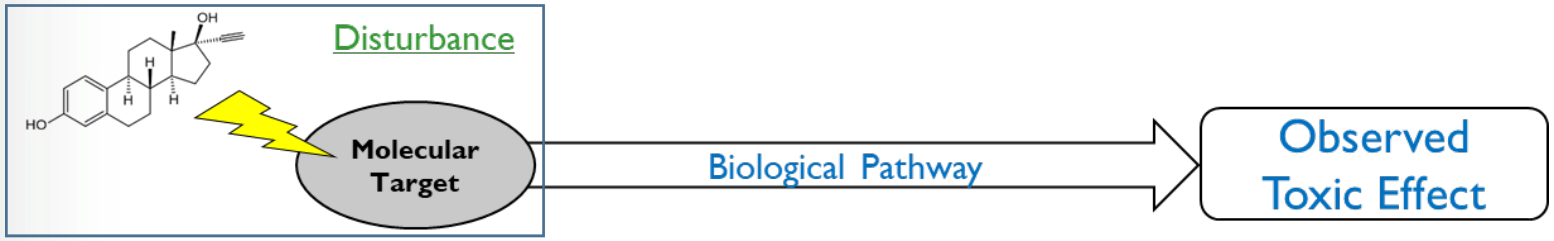
- In Sept. 2019, the current EPA administrator outlined goals to reduce animal use:
  - Reduce requests for, and funding of, mammal studies by 30% by 2025 (**5 years away!**)
  - Eliminate **all** mammal study requests and funding by 2035 (**15 years away!**)



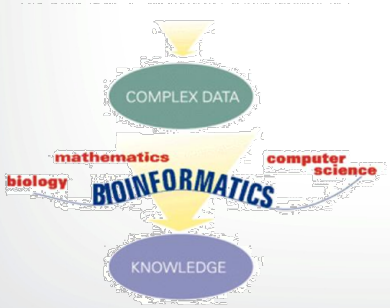
How do we get there?

- **Toxicity Testing in the 21st Century:**
- *In vitro* and *in silico* methods
- Pathway-based approaches
- Focus on disturbance of the **biological pathway**
- Interaction of chemical with molecular target as a predictor of observable toxic effects

New Approach Methods (NAMs)



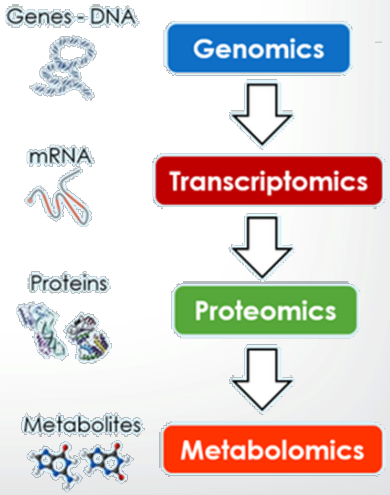
In Silico & Bioinformatics



In Vitro Assays

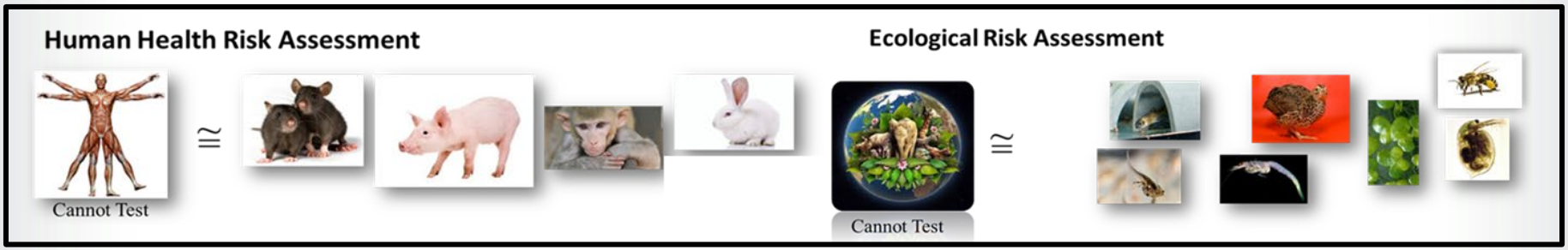
High-Throughput Screening

Omics

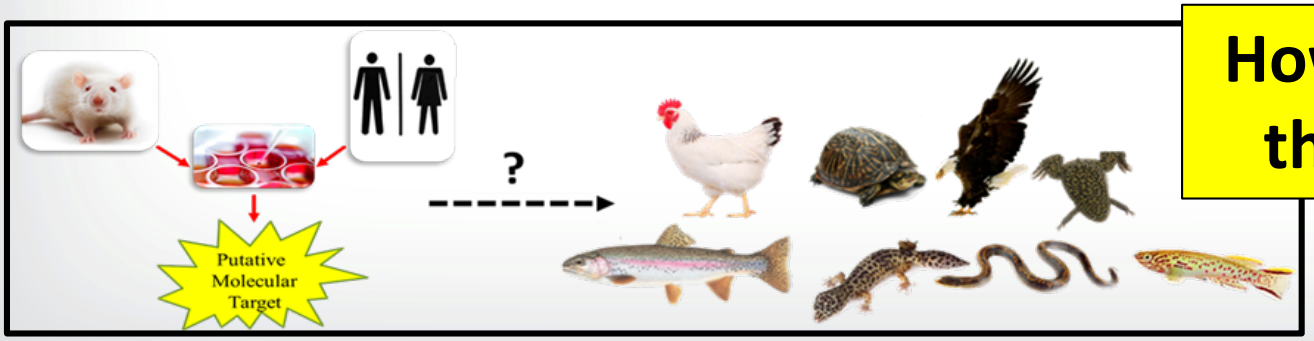


Surrogate Species in Toxicity Testing

- In whole animal testing, the use of representative species is necessary
- **Assumed** that sensitivity of species to a chemical is a function of their relatedness
  - How well do these surrogate models represent the species of concern?



- These questions **still apply** to the new approach methods of modern toxicology
- Extrapolation of bioinformatics, in vitro, high-throughput screening, and omics approaches from test species (E.g. mammalian cells, etc.) to other species
- Is the test species used to determine molecular target a good surrogate for the species we want to protect?



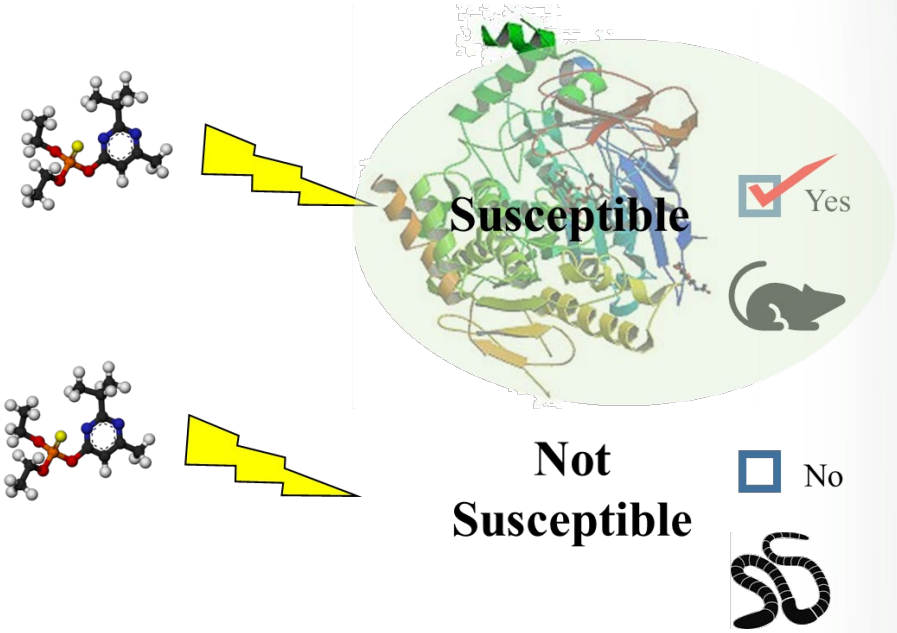
How do we address these questions?

Chemical Sensitivity Across Species

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

Factors that make a species sensitive

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



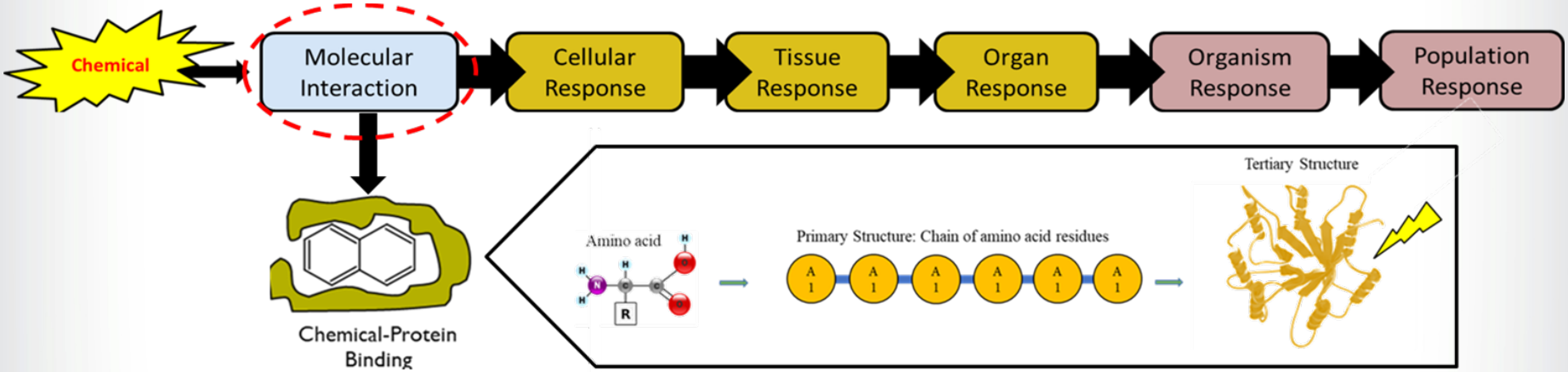
How can we do this?

Use knowledge of protein-chemical interactions
AND
Use tools and technologies for Protein Assessment

Knowledge of protein-chemical interactions

Molecular target: 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.)

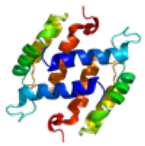


- Evaluation of protein similarity across species can provide a measure of molecular target (protein) conservation and allow predictions of cross-species sensitivity

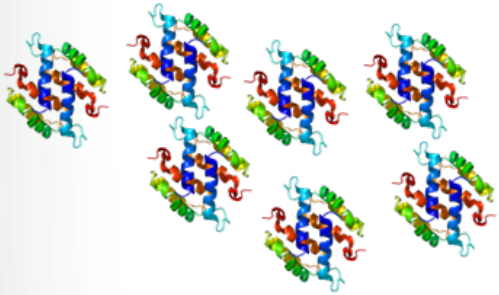
Tools and technologies for Protein Assessment

- Improved sequencing technologies
- Large databases of protein sequence data that are continually growing
- Rapidly access information on **millions** of proteins and **thousands** of species

Chemical Molecular Target in Target Species



≅



Millions of Proteins Representing Thousands Of Species

Does the target exist across species and, if so, how conserved is it?

Taxonomy

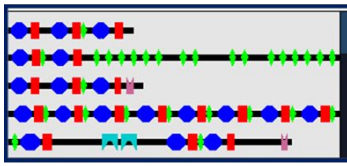
The Taxonomy Database is a curated classification and nomenclature for all of the organisms in the public sequence databases. This currently represents about 10% of the described species of life on the planet.

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.

BLAST® *Basic Local Alignment Search Tool*

Home | Recent Results | Saved Strategies | Help



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.

COBALT *Constraint-based Multiple Alignment Tool*

Home | Recent Results | Help

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)

- Online tool for understanding species susceptibility
- Facilitates rapid and quantitative assessment of protein similarity across species
- Provides a foundation for predicting the taxonomic domain of applicability
- Publicly available, free accounts provided to **anyone** who requests one!
- Developed with both researchers and risk assessors in mind

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

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

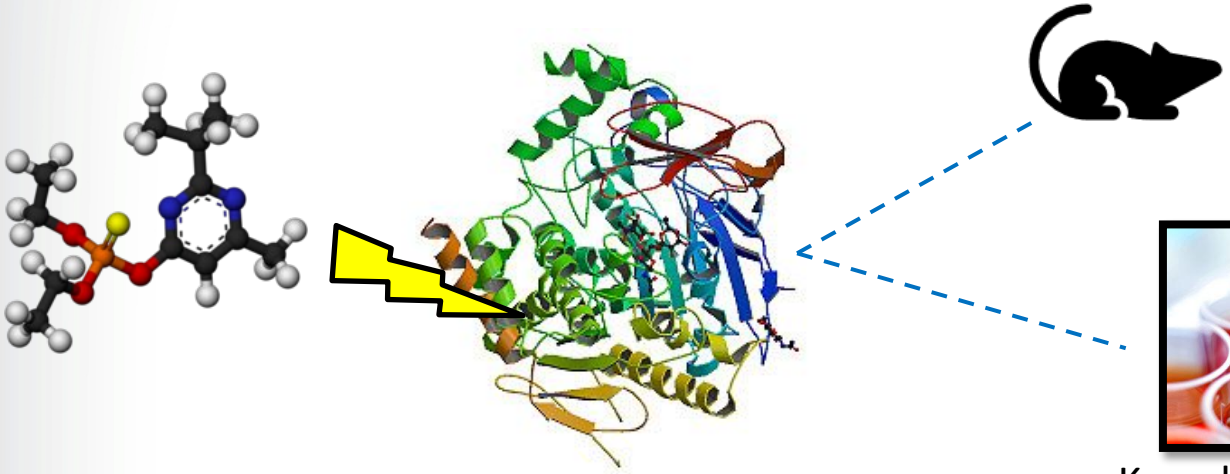
TOXICOLOGICAL SCIENCES, 153(2), 2016, 228–245

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*

What information is required for a SeqAPASS Analysis?



Chemical-Protein Interaction

Knowledge of a sensitive or targeted species

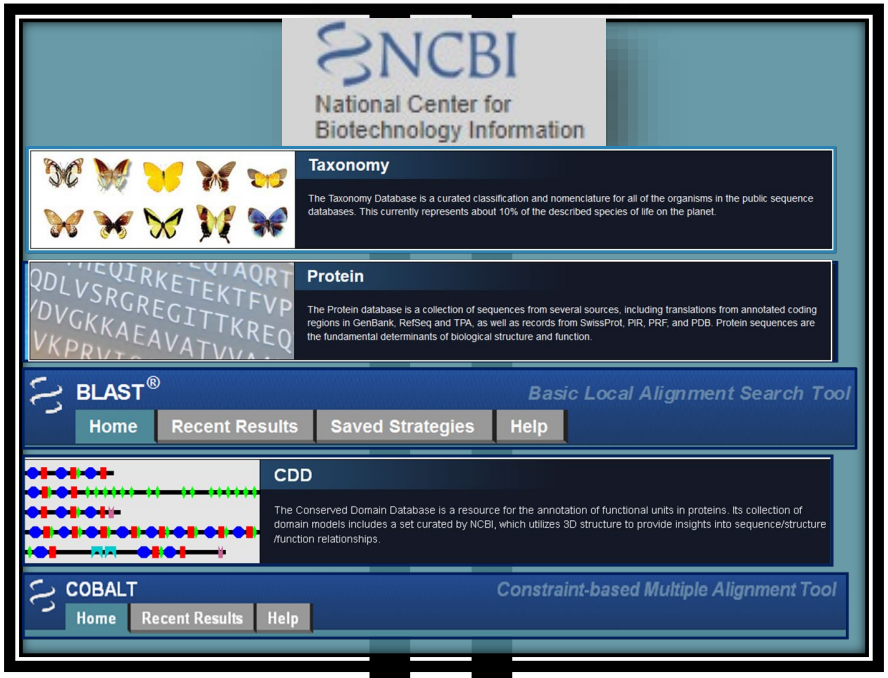
Knowledge of the model organism used in an *in vitro* assay

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

How does it work?

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

SeqAPASS

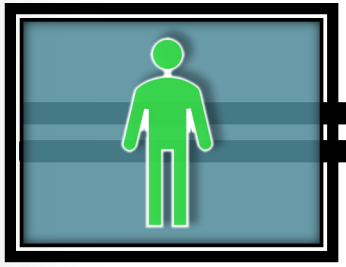

NCBI
National Center for
Biotechnology Information
Taxonomy
The Taxonomy Database is a curated classification and nomenclature for all of the organisms in the public sequence databases. This currently represents about 10% of the described species of life on the planet.
Protein
The Protein database is 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.
BLAST®
Basic Local Alignment Search Tool
Home
Recent Results
Saved Strategies
Help
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.
COBALT
Constraint-based Multiple Alignment Tool
Home
Recent Results
Help

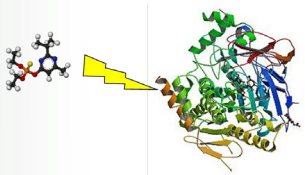
SeqAPASS

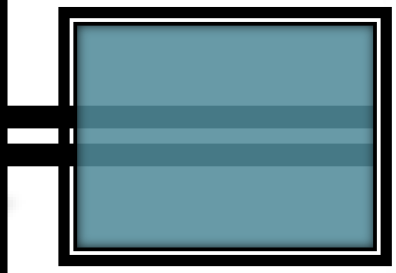
SeqAPASS Level 1

SeqAPASS Level 1

Human Protein Target

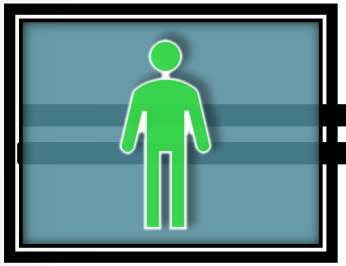


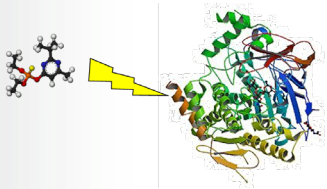


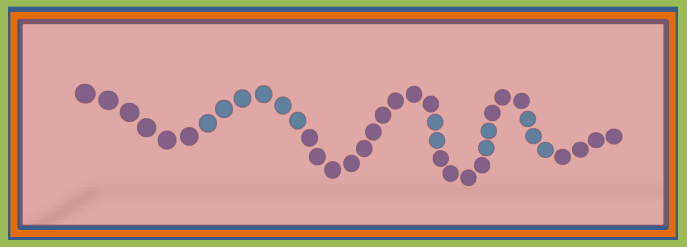


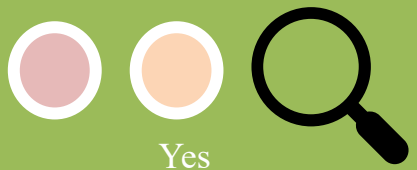
SeqAPASS Level 1

Human Protein Target





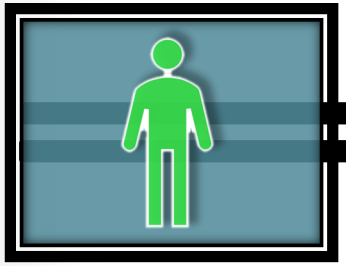


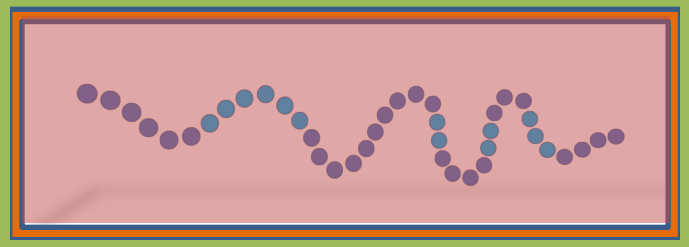


Line of Evidence:
Primary amino acid sequence
Conserved

SeqAPASS Level 1

Human Protein Target

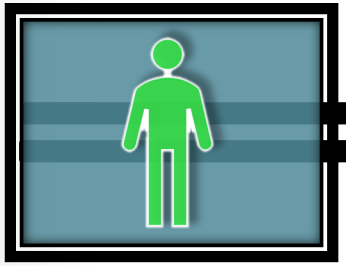




Line of Evidence:
Primary amino acid sequence
Conserved

SeqAPASS Level 1

Human Protein Target

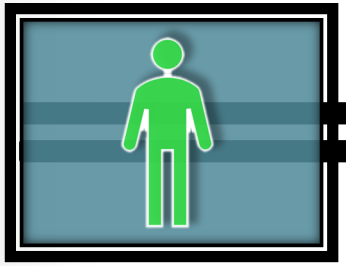


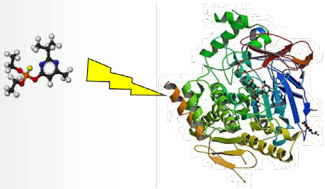
No

Line of Evidence:
Primary amino acid sequence
Conserved

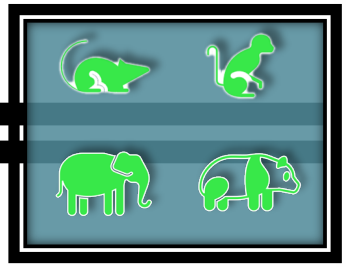
SeqAPASS Level 1

Human Protein Target

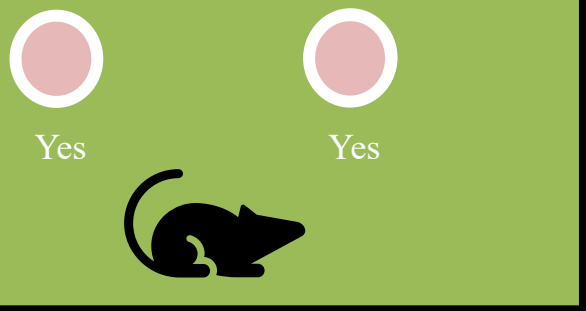


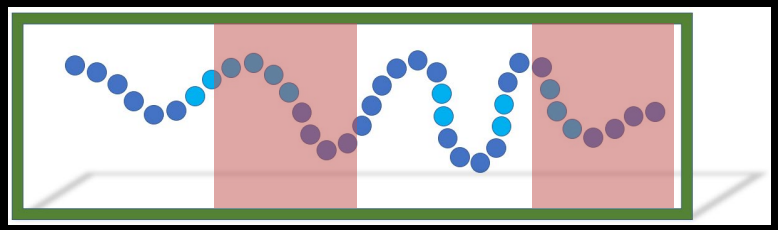


Line of Evidence:
Primary amino acid sequence
Conserved

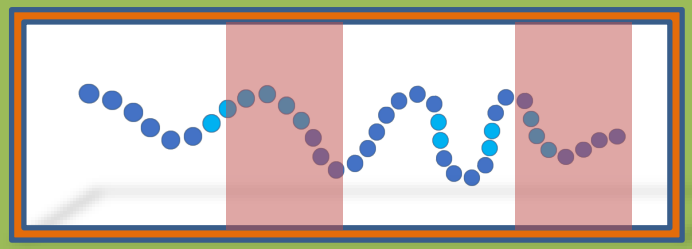


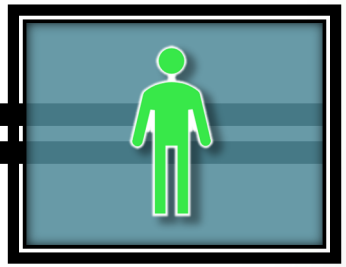
Level 2





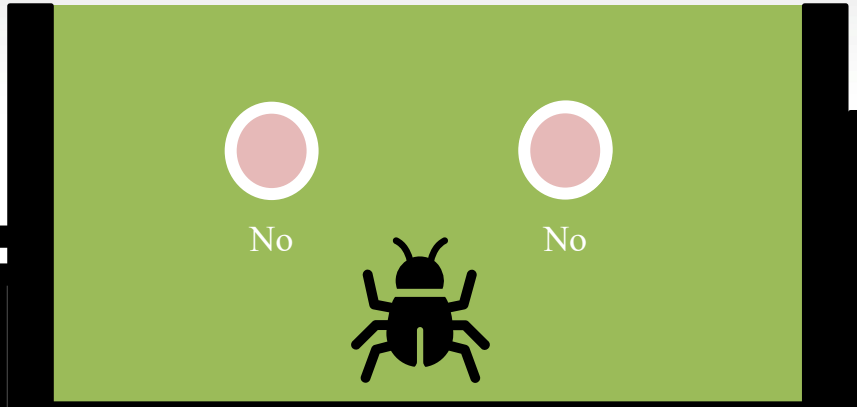
Line of Evidence:
Domain
Conserved

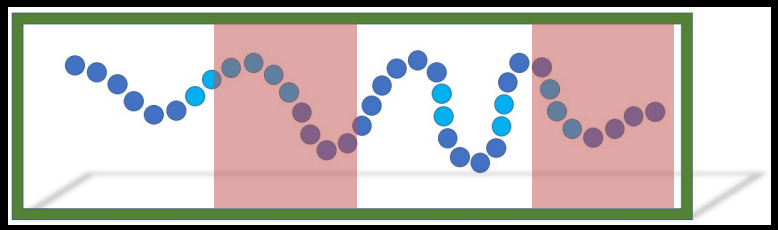




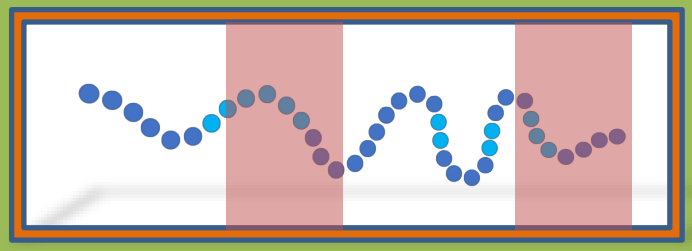
Human Functional Domain(s)

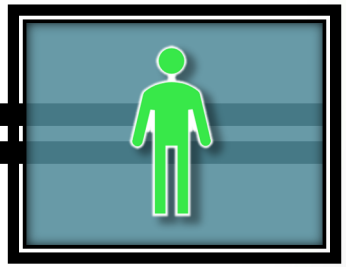
Level 2





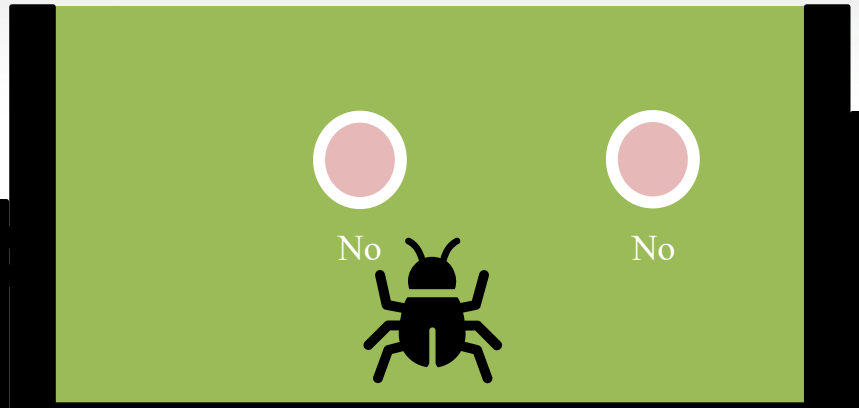
Line of Evidence:
Domain
Conserved

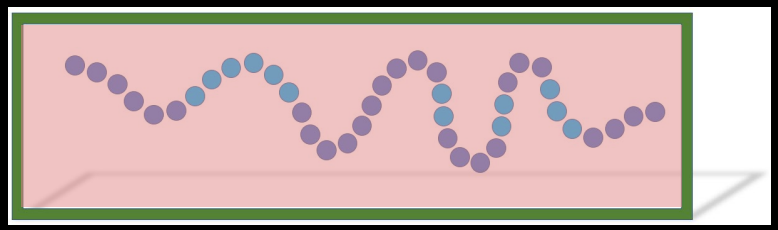




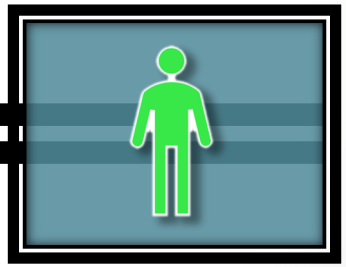
Human Functional Domain(s)

Level 2



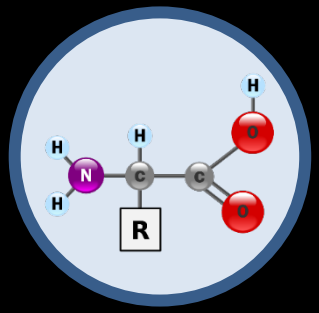


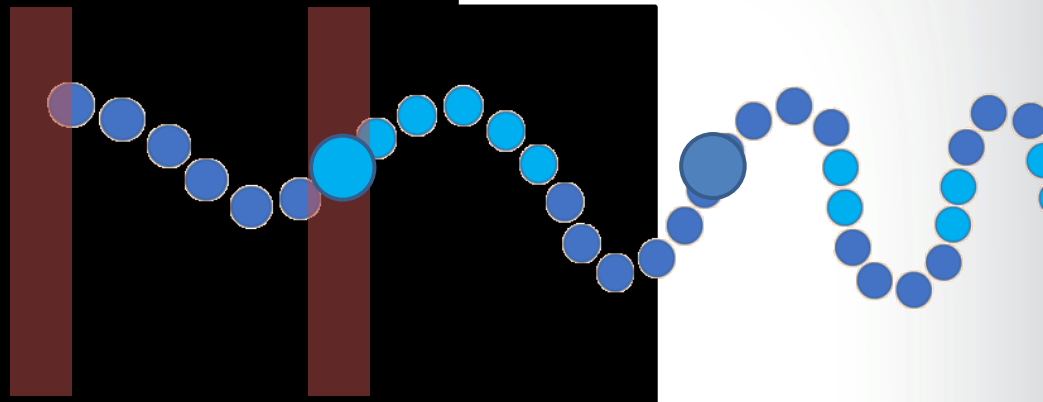
Line of Evidence:
Domain
NOT Conserved

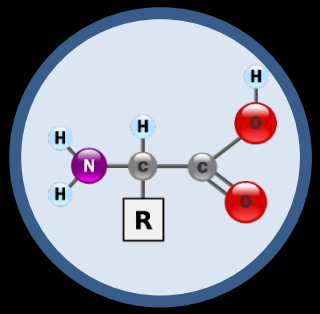


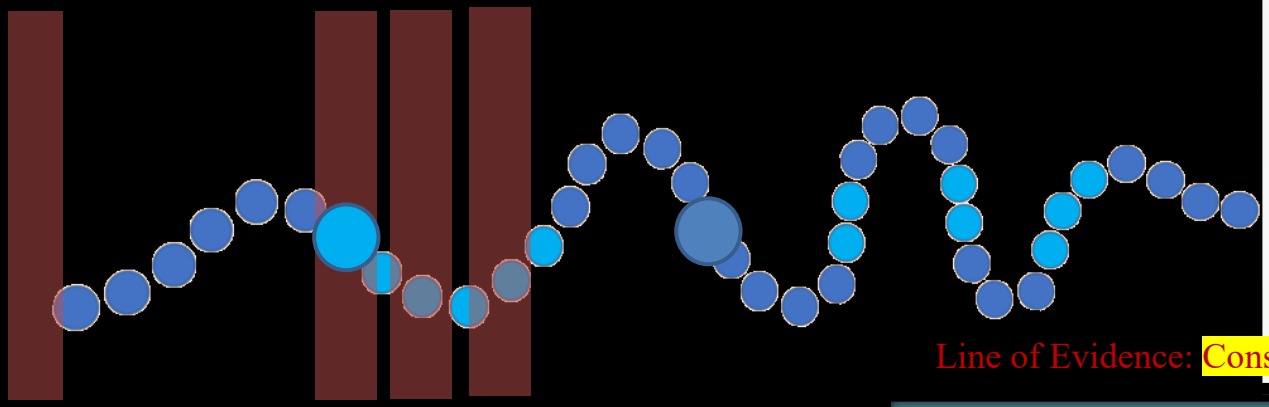
Human Functional Domain(s)

Human Critical Amino Acids



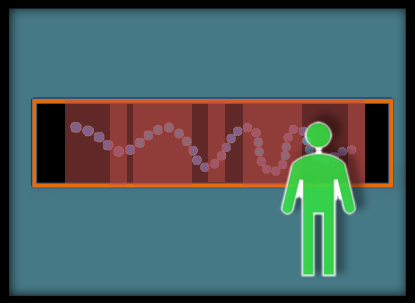






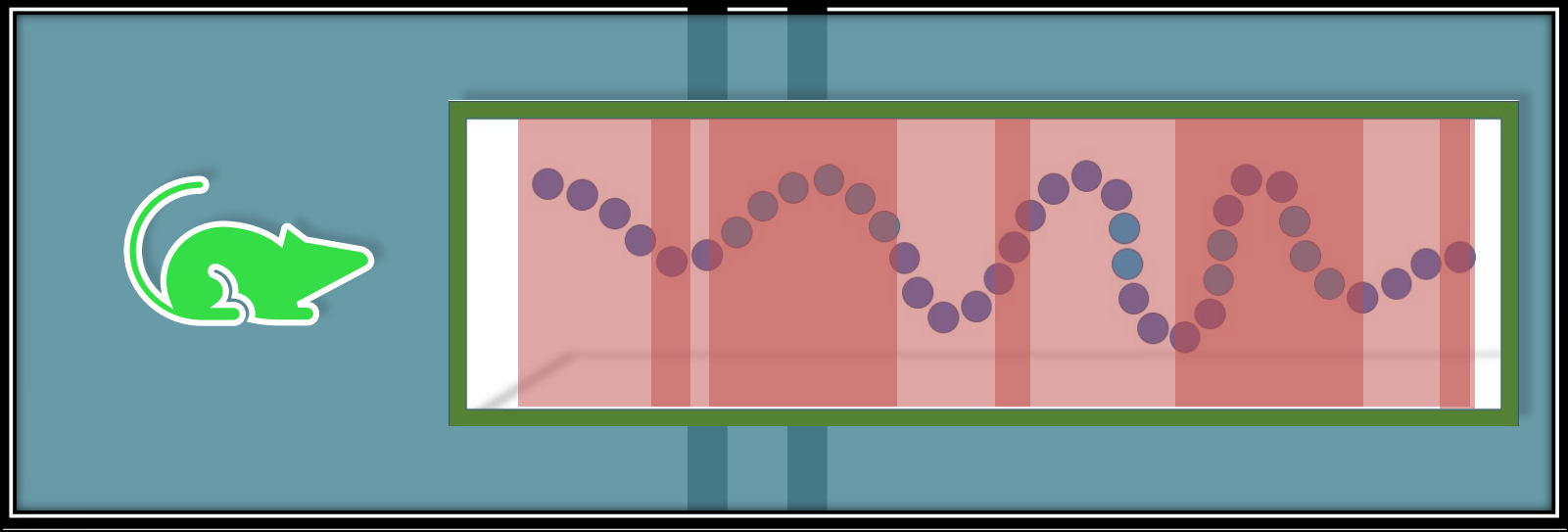
Line of Evidence: Conserved

Level 3



SeqAPASS

Summary



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

Applications of the SeqAPASS Tool
What kind of questions can this help us answer?

- Extrapolate high throughput screening data
- Extrapolate biological pathway knowledge across species
- Predict relative intrinsic susceptibility
- Generate research hypotheses
- Prioritize testing efforts

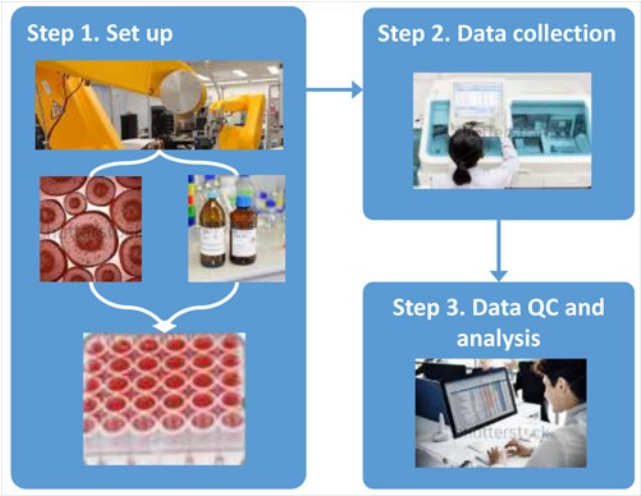
Some examples of questions driving SeqAPASS predictions

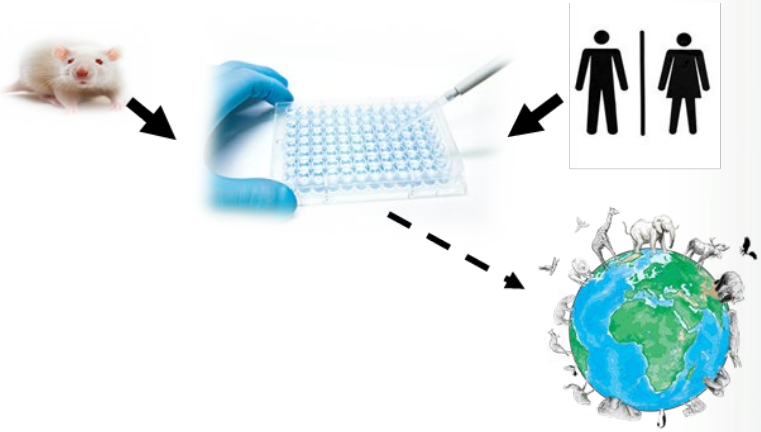
1. Is a chemical (pharmaceutical etc.) that activates a receptor (ER, AR, PPAR, AhR, etc.) in humans predicted to bind to that receptor in other species?
2. If a chemical is binding to a protein target in my model species (zebrafish, medaka, etc.), how does this target compare to humans or other species I care about?
3. Is any species or group of species (endangered species, commercially important species, etc.) predicted to be more or less susceptible to a given chemical with a known protein target?

Applications of SeqAPASS: Extrapolating High-Throughput Screening Data

US EPA Endocrine Disruptor Screening Program: Tasked with assessing thousands of chemicals for potential endocrine activity

US EPA ToxCast Program: Uses mammalian cell-based assays to rapidly screen chemicals and identify those most likely to be endocrine disruptors



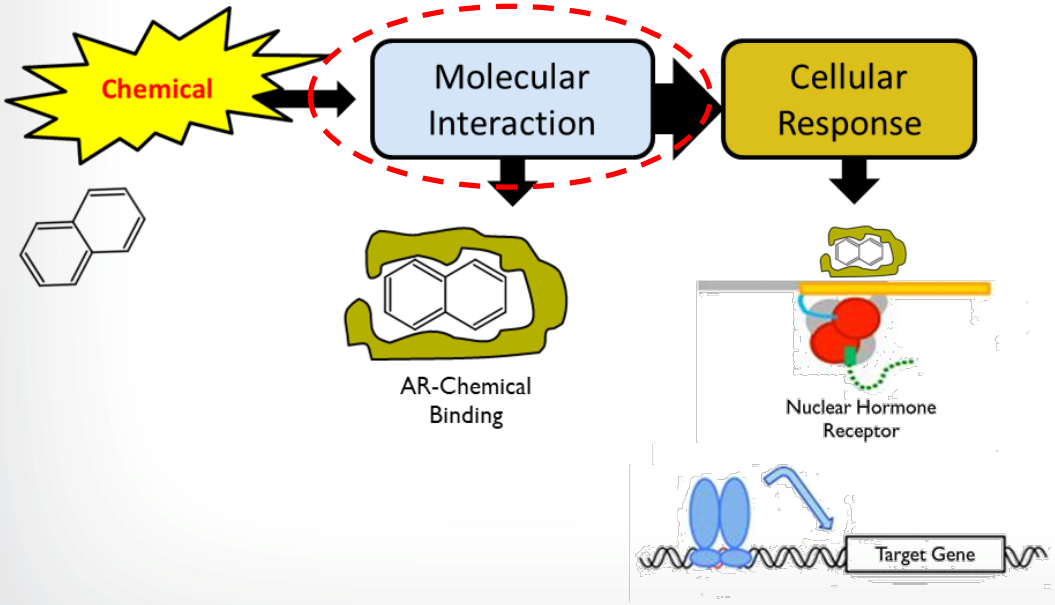


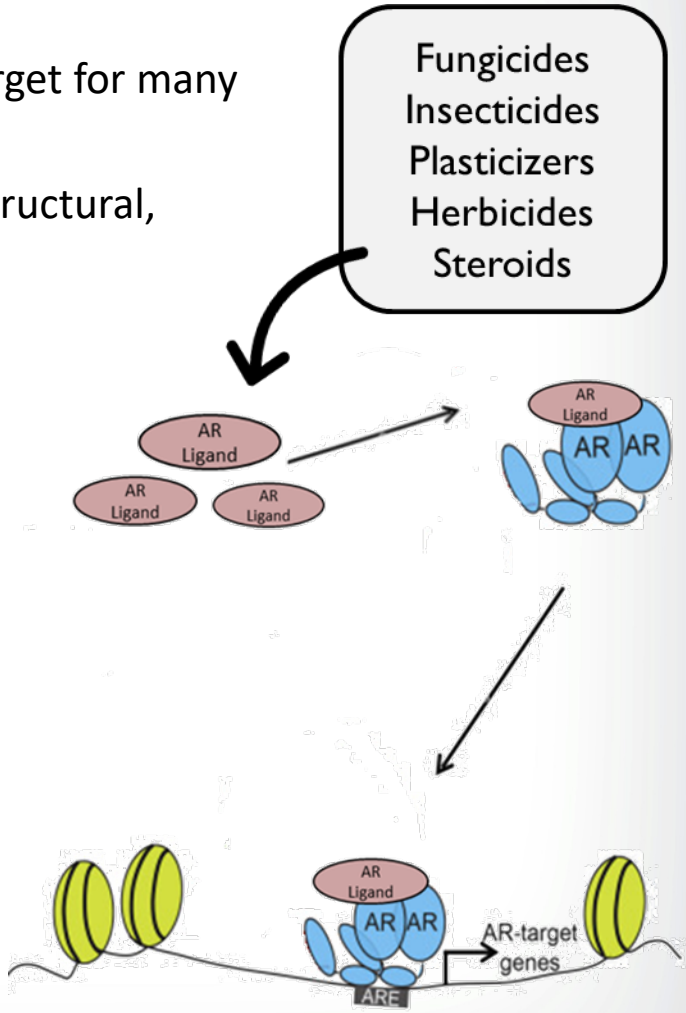
Key Questions for Consideration:

- How well does this mammalian-based prioritization approach reasonably reflect potential impacts on other vertebrates?
- Can we expect chemicals that interact with mammalian receptors to also interact with receptors of other species?

Applications of SeqAPASS: Extrapolating High-Throughput Screening Data

- The androgen receptor (AR) is an important endocrine target for many environmental and industrial chemicals
- AR is a chemical target with a large base of pre-existing structural, molecular target, and toxicity data
- Knowledge of androgenic chemicals is still growing



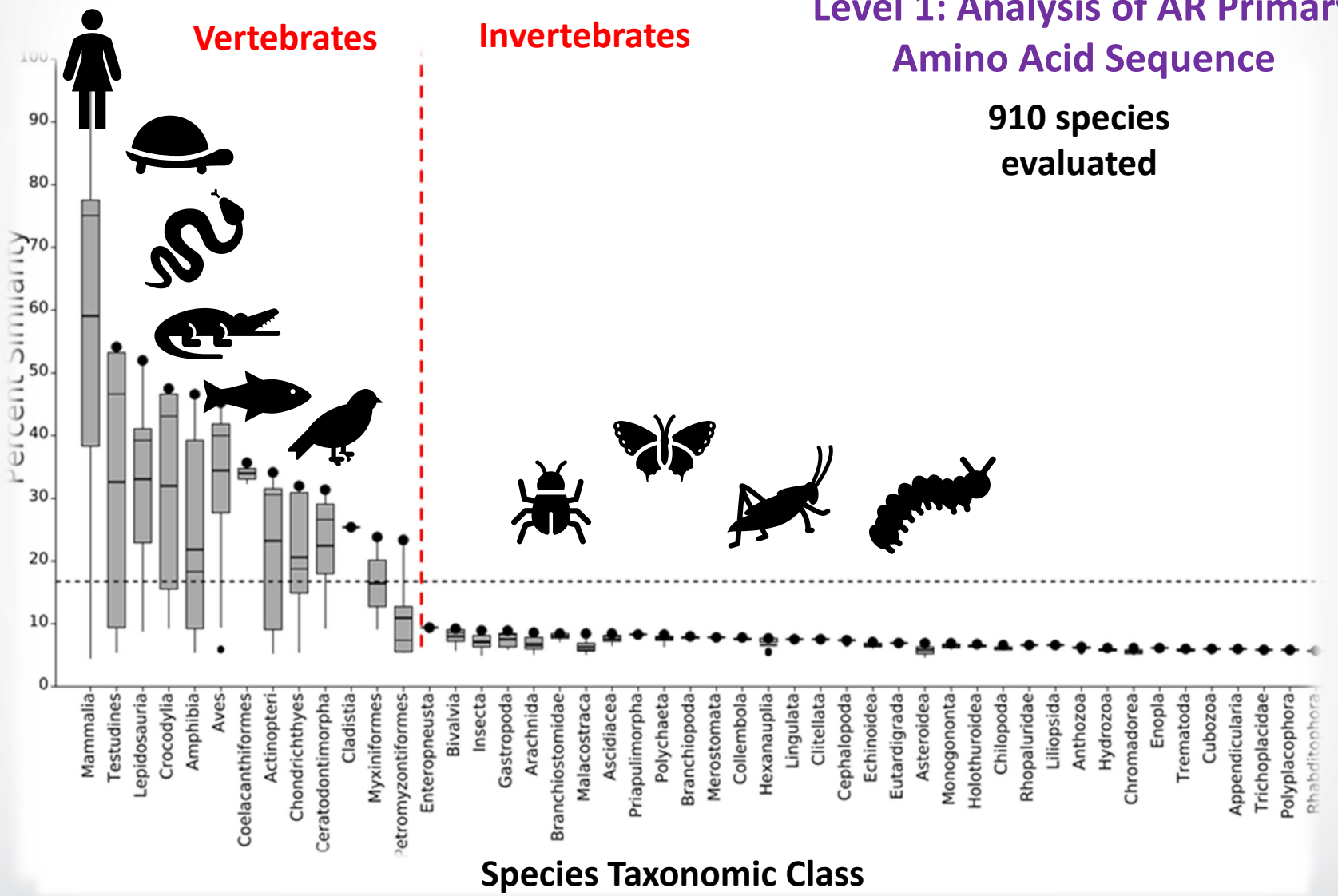


Adapted from Zhang et al. 2015

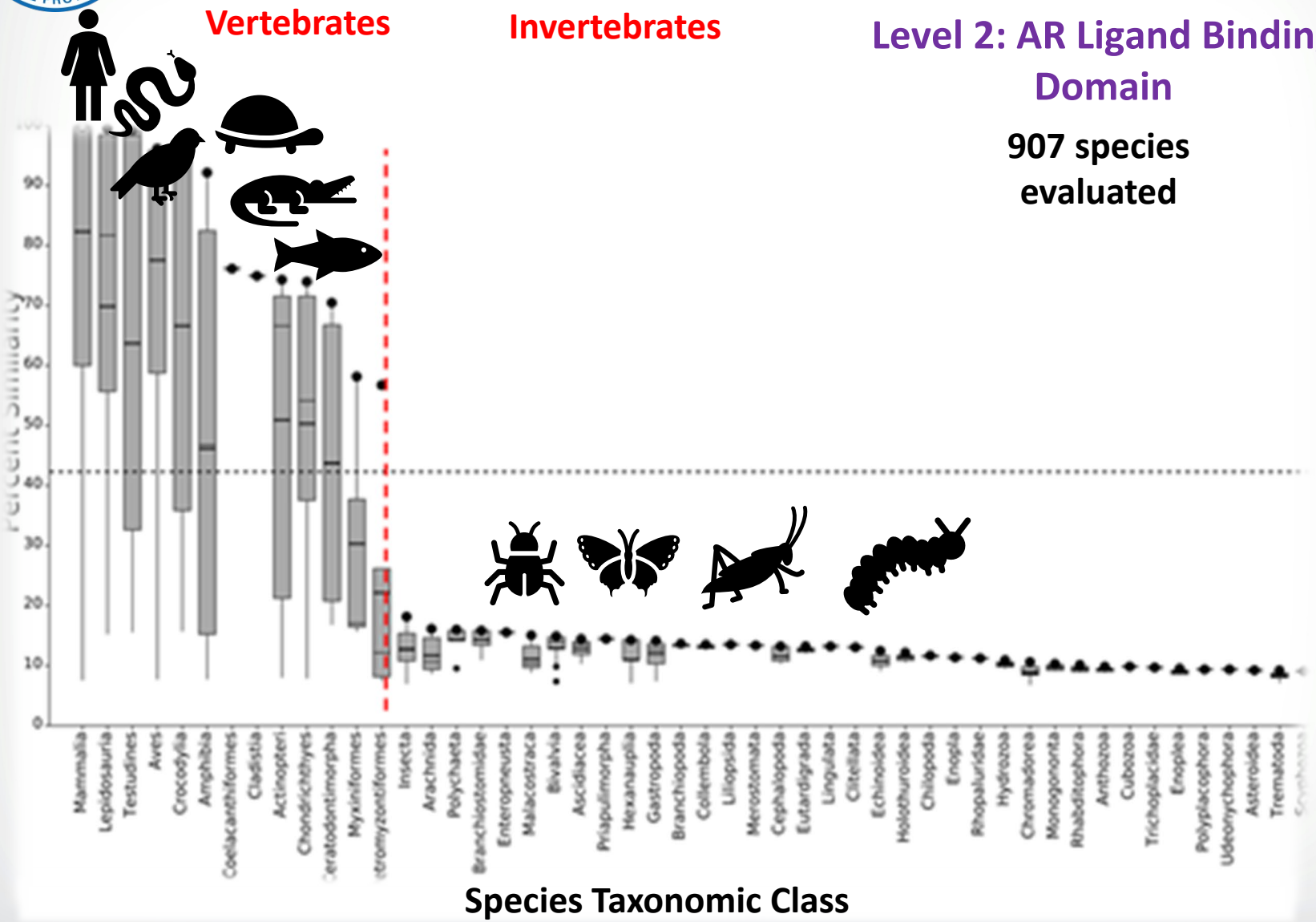
Applications of SeqAPASS: Extrapolating High-Throughput Screening Data

Level 1: Analysis of AR Primary Amino Acid Sequence

910 species evaluated



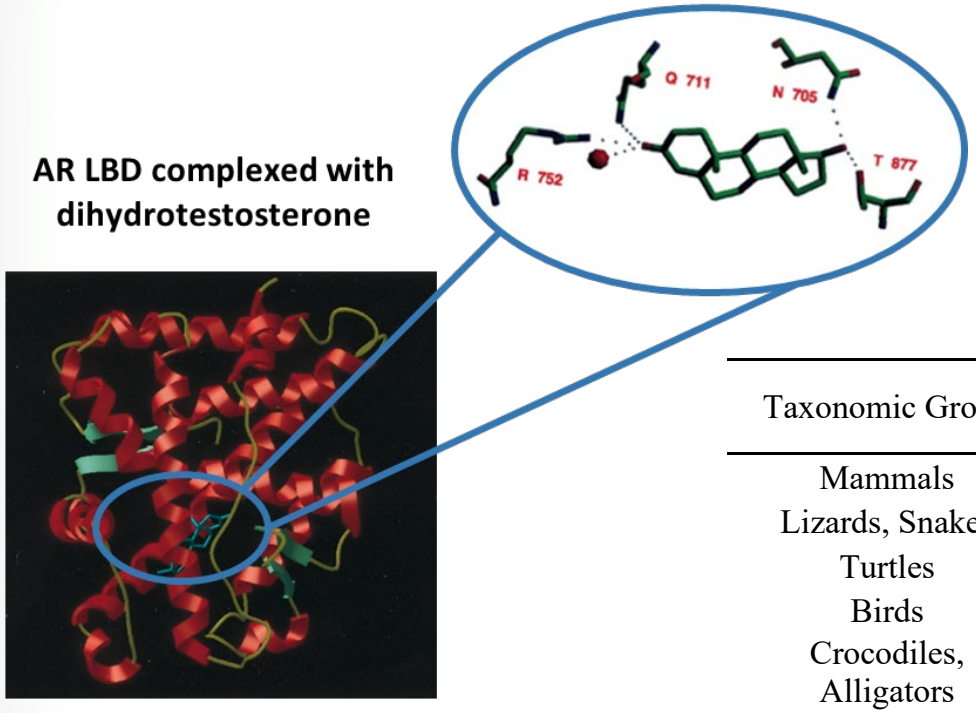
Applications of SeqAPASS: Extrapolating High-Throughput Screening Data

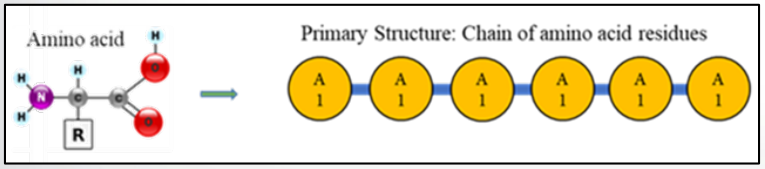


Applications of SeqAPASS: Extrapolating High-Throughput Screening Data

- Critical (close-contact) amino acids required for hydrogen binding to AR ligands identified by mining of existing literature (homology modeling, crystal structures, etc.)

AR LBD complexed with dihydrotestosterone





Level 3: Analysis of Conservation of Individual Amino Acid Residues

250 species evaluated

Taxonomic Group	Class Name	# of Spp.	Shared Susceptibility
Mammals	Mammalia	118	Yes/No
Lizards, Snakes	Lepidosauria	11	Yes
Turtles	Testudines	3	Yes
Birds	Aves	58	Yes
Crocodiles, Alligators	Crocodylia	4	Yes
Amphibians	Amphibia	13	Yes
Coelacanths	Coelacanthiformes	2	Yes
Eel-shaped	Cladistia	1	Yes
Bony Fish	Actinopteri	87/1	Yes/No
Sharks, Rays	Chondrichthyes	4	Yes
Lungfish	Ceratodontimorpha	2	Yes

Applications of SeqAPASS: Extrapolating High-Throughput Screening Data

Class Name	Level 1 Shared Susceptible?	Level 2 Shared Susceptible ?	Level 3		Level 3 Shared Susceptible ?
			Species with Shared Susceptibility	Species not Similarly Susceptible	
Mammalia	Yes	Yes	106	0	Yes
Lepidosauria	Yes	Yes	3	0	Yes
Testudines	Yes	Yes	3	0	Yes
Aves	Yes	Yes	33	0	Yes
Crocodylia	Yes	Yes	3	0	Yes
Amphibia	Yes	Yes	5	0	Yes
Coelacanthiformes	Yes	Yes	2	0	Yes
Cladistia	Yes	Yes	1	0	Yes
Actinopteri	Yes	Yes	88	1	Yes/No
Chondrichthyes	Yes	Yes	3	0	Yes
Ceratodontimorpha	Yes	Yes	2	0	Yes

- Across all three levels, SeqAPASS results suggest conservation of AR across vertebrate species
- Chemicals that bind and activate AR in mammalian-based assays, are likely to interfere with AR in other vertebrate species

How can we "validate" and expand upon these predictions?

“Validation” of SeqAPASS Predictions

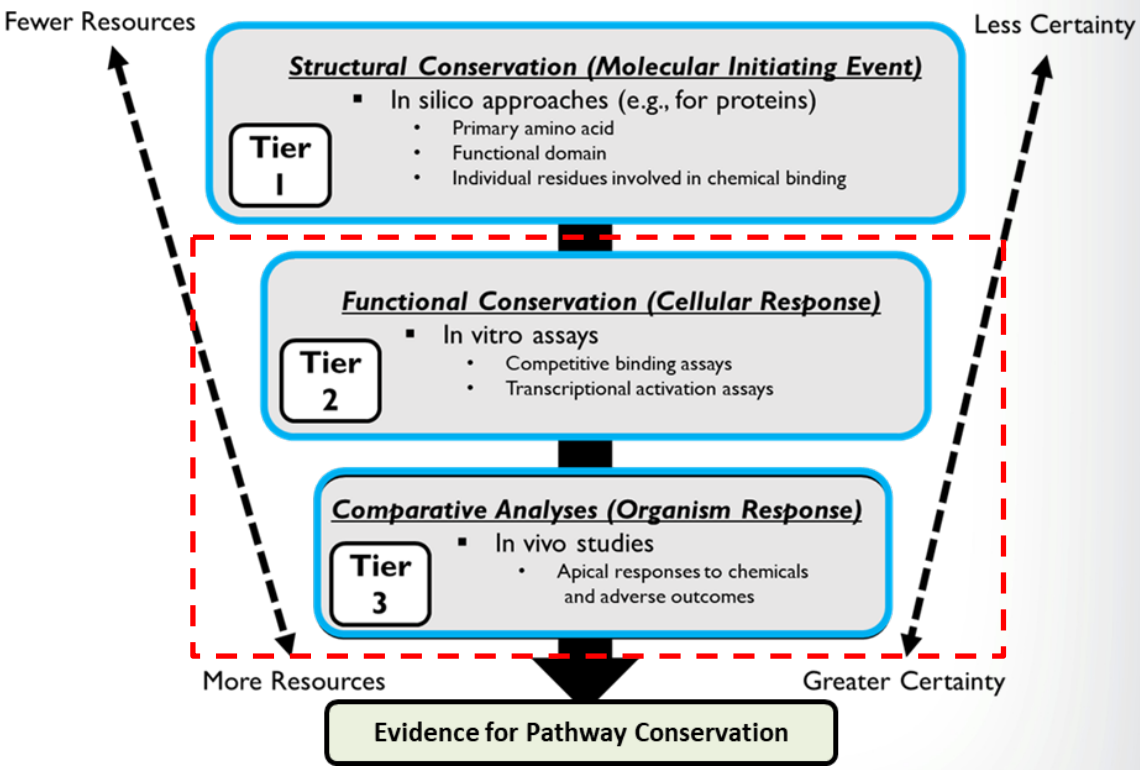
- For SeqAPASS to be used in a regulatory context, it is **essential** to understand how computational predictions relate to empirical data across species

In the lab

- Cross species In vitro studies
- Cross species In vivo studies
- Chemical proteomics
- Etc.

Out of the lab

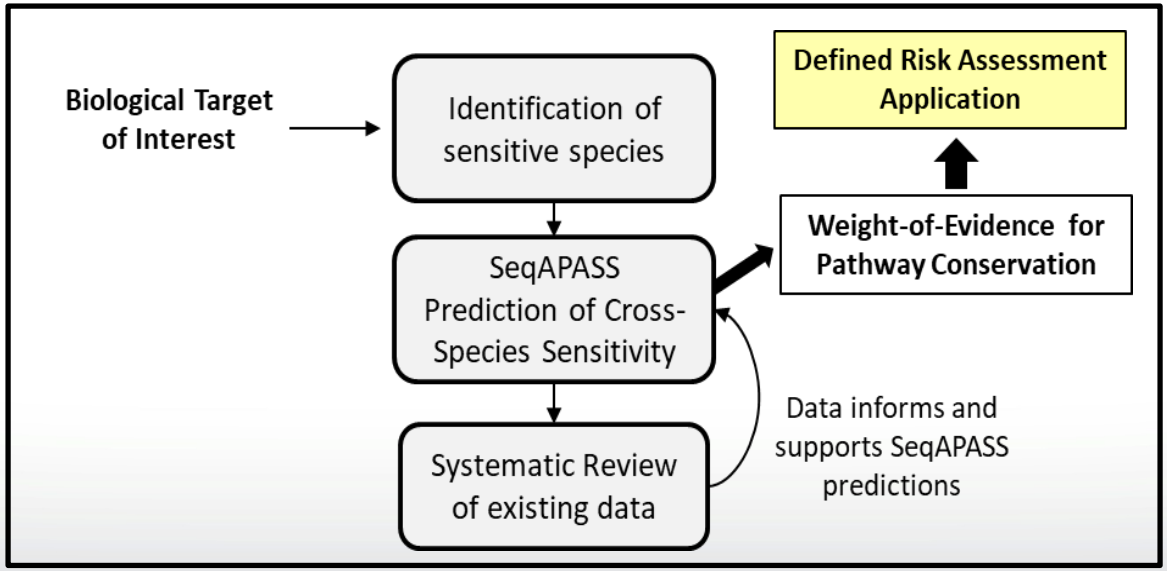
- Homology modelling
- Molecular docking
- Review of existing evidence
- Etc.



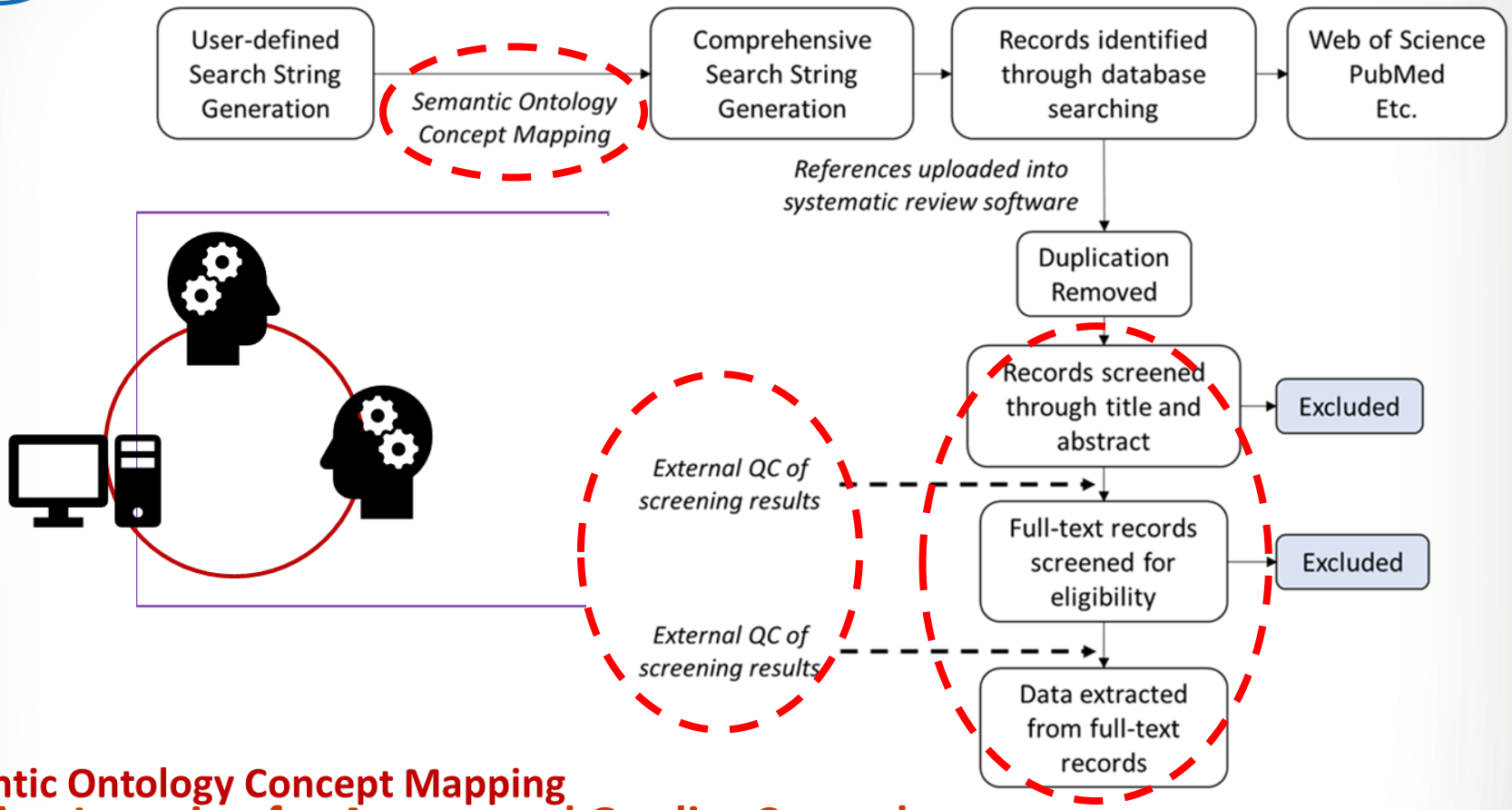
Evaluating and Supporting Cross-Species Predictions

Systematic Literature Review: A type of literature review that uses systematic methods to collect secondary data, critically appraise research studies, and synthesize findings

- Using existing evidence (literature), we can evaluate the scientific basis of our cross-species predictions
- Gathering in vivo and in vitro data on vertebrate species exposed to known androgenic compounds, additional lines of evidence can be compiled for the conservation of the androgen receptor across species



Incorporation of Technical Advances and Tools for Improved Systematic Review:



Semantic Ontology Concept Mapping

- Using accurate search terms is essential for conducting an unbiased, comprehensive survey of the literature landscape
- Scientific language is often non-standard, redundant, and mischaracterized
- Semantic ontology concept mapping approaches can develop comprehensive literature search strings by expanding the computer's knowledge

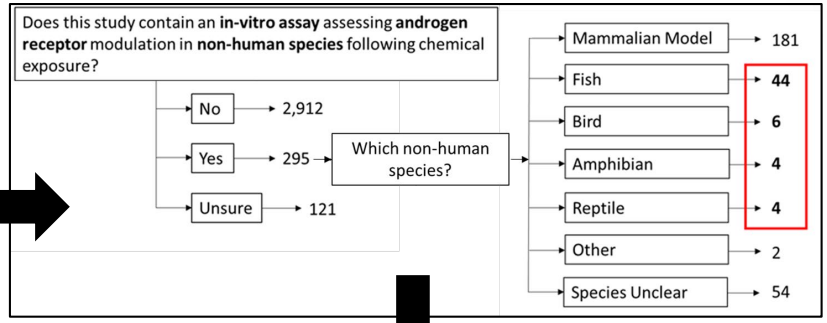
Machine Learning for Automated Quality Control

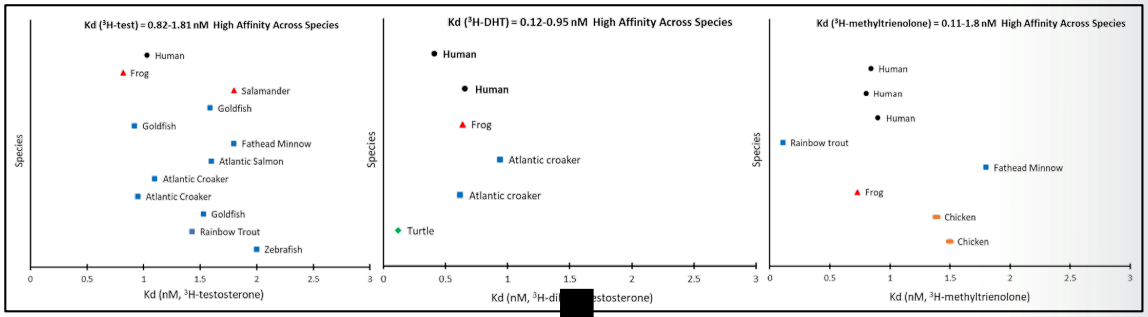
- Development of machine learning models to screen through identified articles and identify possible errors (compared with external review by scientists)
- Facilitates blind Quality Assurance/Quality Control
- Does the computer agree with the human that this article is relevant/irrelevant

Incorporation of Technical Advances and Tools for Improved Systematic Review:

What does this look like?

Class Name	Level 1 Shared Susceptible?	Level 2 Shared Susceptible ?	Level 3 Species with Shared Susceptibility	Level 3 Species not Similarly Susceptible	Level 3 Shared Susceptible ?
Mammalia	Yes	Yes	106	0	Yes
Lepidosauria	Yes	Yes	3	0	Yes
Testudines	Yes	Yes	3	0	Yes
Aves	Yes	Yes	33	0	Yes
Crocodylia	Yes	Yes	3	0	Yes
Amphibia	Yes	Yes	5	0	Yes
Coelacanthiformes	Yes	Yes	2	0	Yes
Cladistia	Yes	Yes	1	0	Yes
Actinopteri	Yes	Yes	88	1	Yes/No
Chondrichthyes	Yes	Yes	3	0	Yes
Ceratodontimorpha	Yes	Yes	2	0	Yes





Tier 1 Structural Conservation (Molecular Initiating Event)

Tier 2 Functional Conservation (Cellular Response)

Tier 3 Comparative Analyses (Organism Response)

Assemble Evidence for Pathway Conservation for Defined Risk Assessment Application

Chemical	Observations by Vertebrate Class: Fish	Reptile	Amphibian	Bird
Testosterone (endogenous androgen)	• Female development of male sex traits	• Female development of male sex traits • Male-biased populations (Skewed sex-ratios)	• Male-biased populations (Skewed sex-ratios)	• Female development of male sex traits • Increase in male behaviors
Methyltestosterone (synthetic androgen)	• Female development of male sex traits • Reduced gonadosomatic index (GSI)	• Female development of male sex traits • Male-biased populations (Skewed sex-ratios)	• Male-biased populations (Skewed sex-ratios)	• Reduced egg laying in females
17ß-trenbolone (environmental androgen)	• Female development of male sex traits • Reduced circulating E2 and vitellogenin levels	• Female development of male sex traits • Male-biased populations (Skewed sex-ratios)	• Female development of male sex traits • Male-biased populations (Skewed sex-ratios)	• Female development of male sex traits • Male-biased populations (Skewed sex-ratios)

Acknowledgements
Carlie LaLone, EPA, ORD
Kristan Markey, EPA, OCSPP
Scott Lynn, EPA, OCSPP

Thanks!

Any questions?

SeqAPASS v4.0

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

Anyone can use SeqAPASS to help inform their own research questions! If you are interested in using SeqAPASS we are happy to help!

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