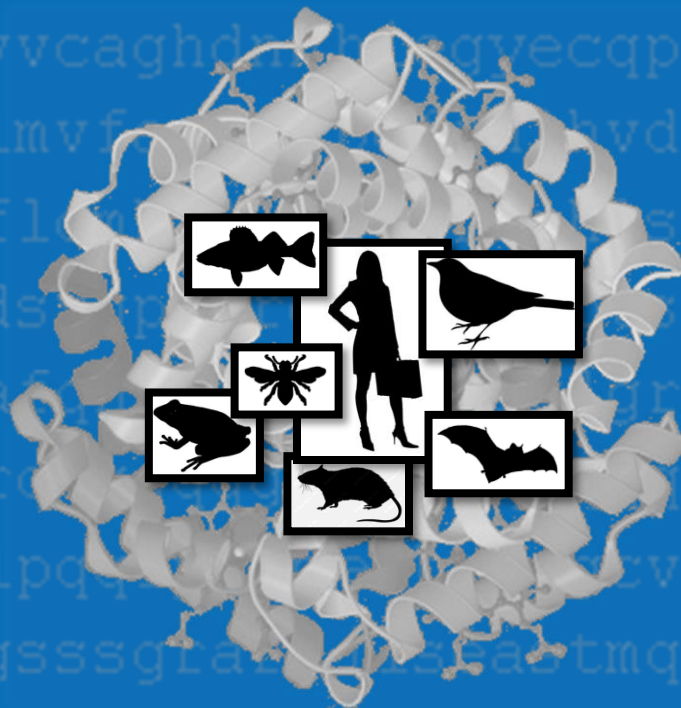


# The Sequence Alignment to Predict Across Species Susceptibility Tool:



**Extrapolating Knowledge Computationally**

Presenter: Carlie A. LaLone, Ph.D.

May 11th 2022

# Overview

- **Need** for cross species extrapolation
- **Bioinformatics** to advance extrapolation
- The Sequence Alignment to Predict Across Species Susceptibility (**SeqAPASS**) tool
- Demonstrated **applications**
- Bringing champions in the field together to advance the science for action through an **international consortium**



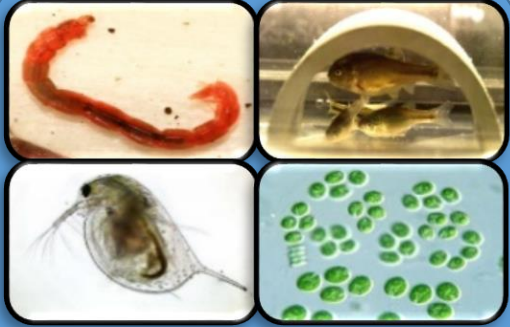


# Toxicity Testing to Understand Chemical Safety

- US EPA Examples:
- *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 21<sup>st</sup> 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*

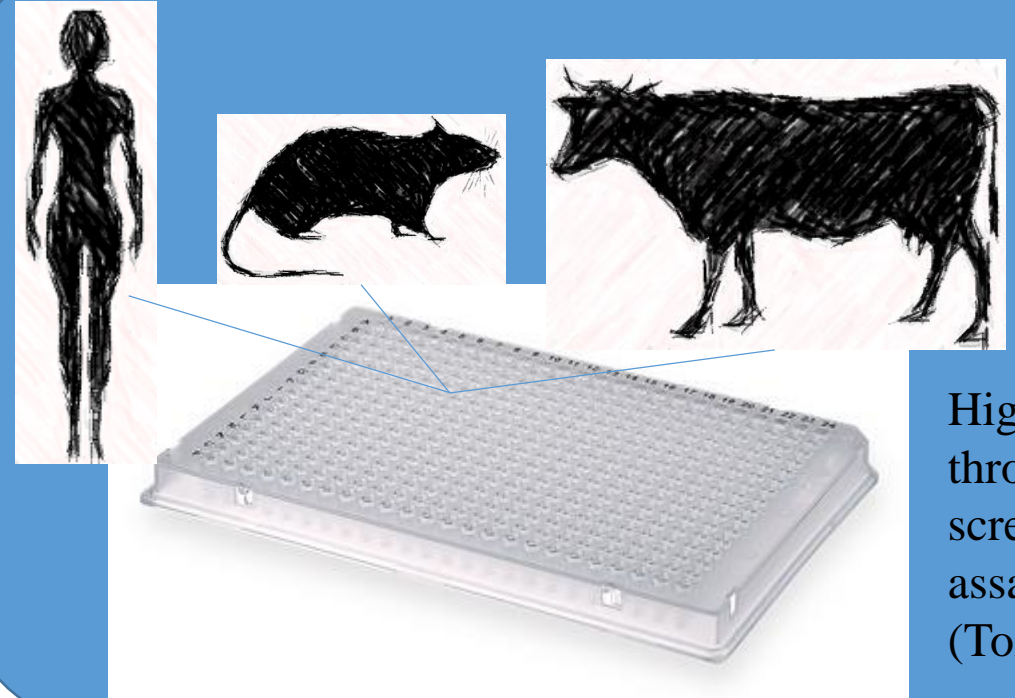


# Need for Advances in Species Extrapolation



High throughput  
transcriptomics

Historic 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**



**easy maintenance and good breeding capabilities**



**short lifespans and rapid life cycles**



**ability to control diet and surroundings**





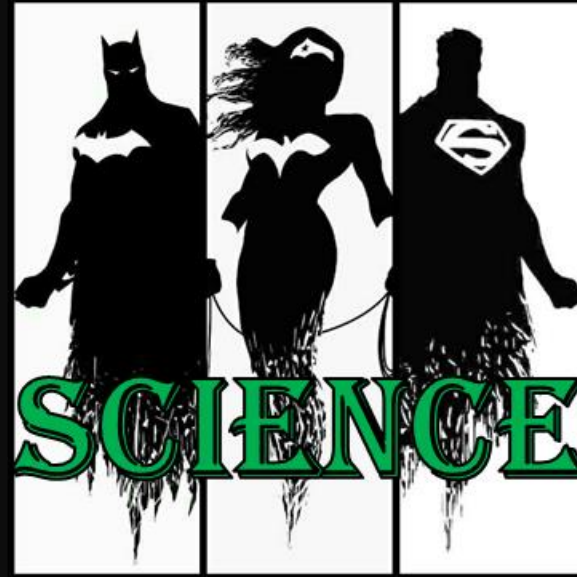
# Species Extrapolation

## 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**



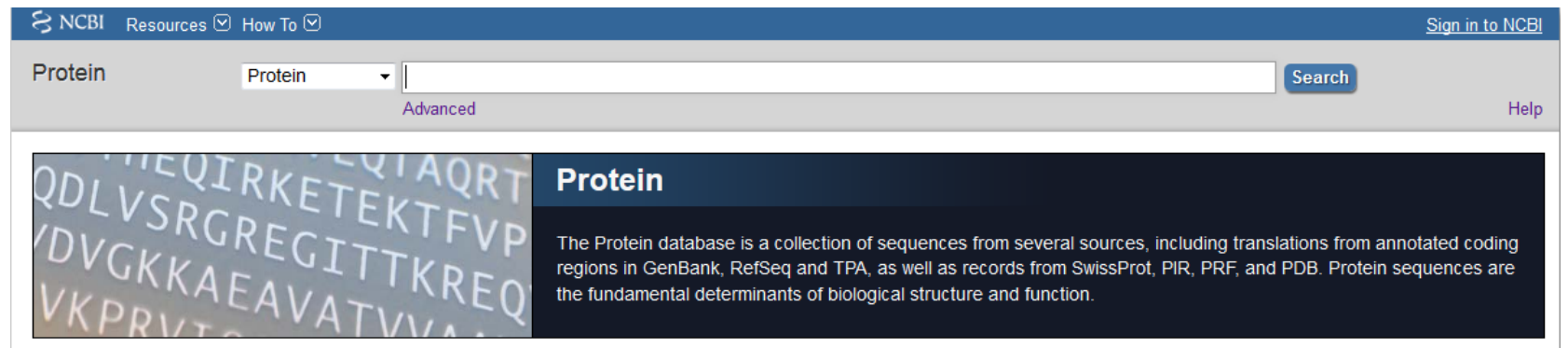

# Where could we begin in understanding species similarities and differences?

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: 224,211,842 Proteins representing 117,030 Organisms**



NCBI Resources How To Sign in to NCBI

Protein Protein Search Advanced Help

**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.

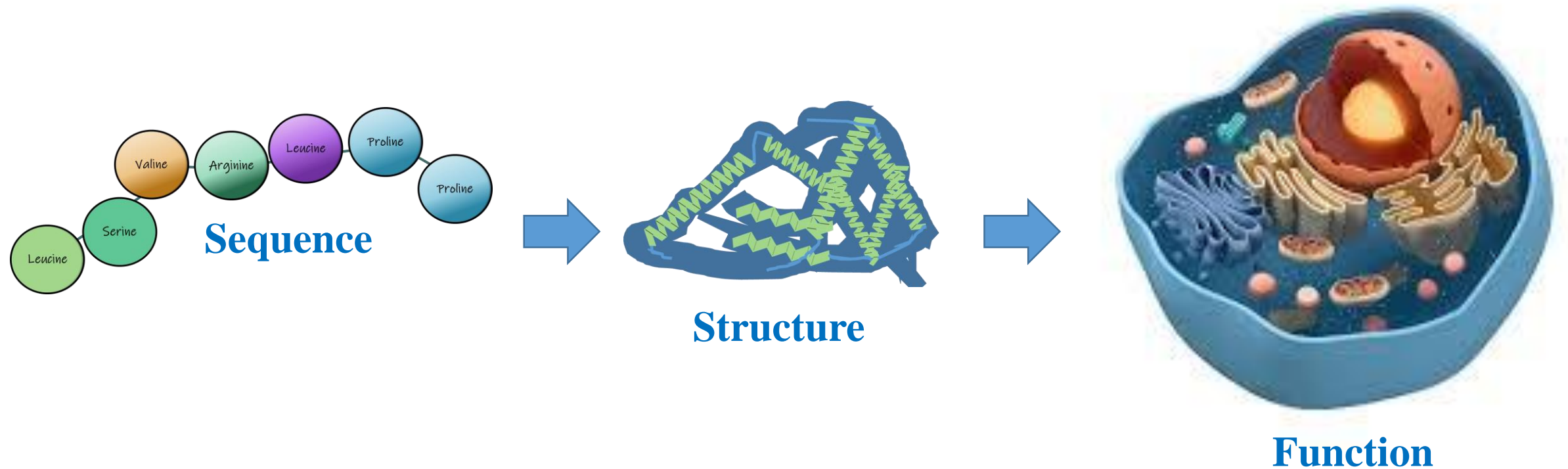
# Bioinformatics

- 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**



# Begin Simple and Advance as the Science Advances

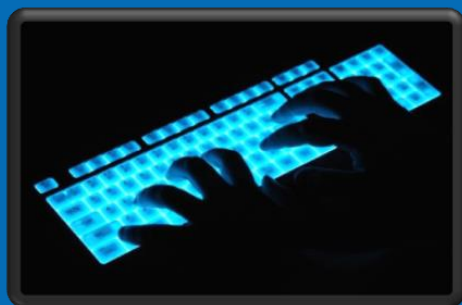


Consider sequence and structural attributes to understand protein conservation across species



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

# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)



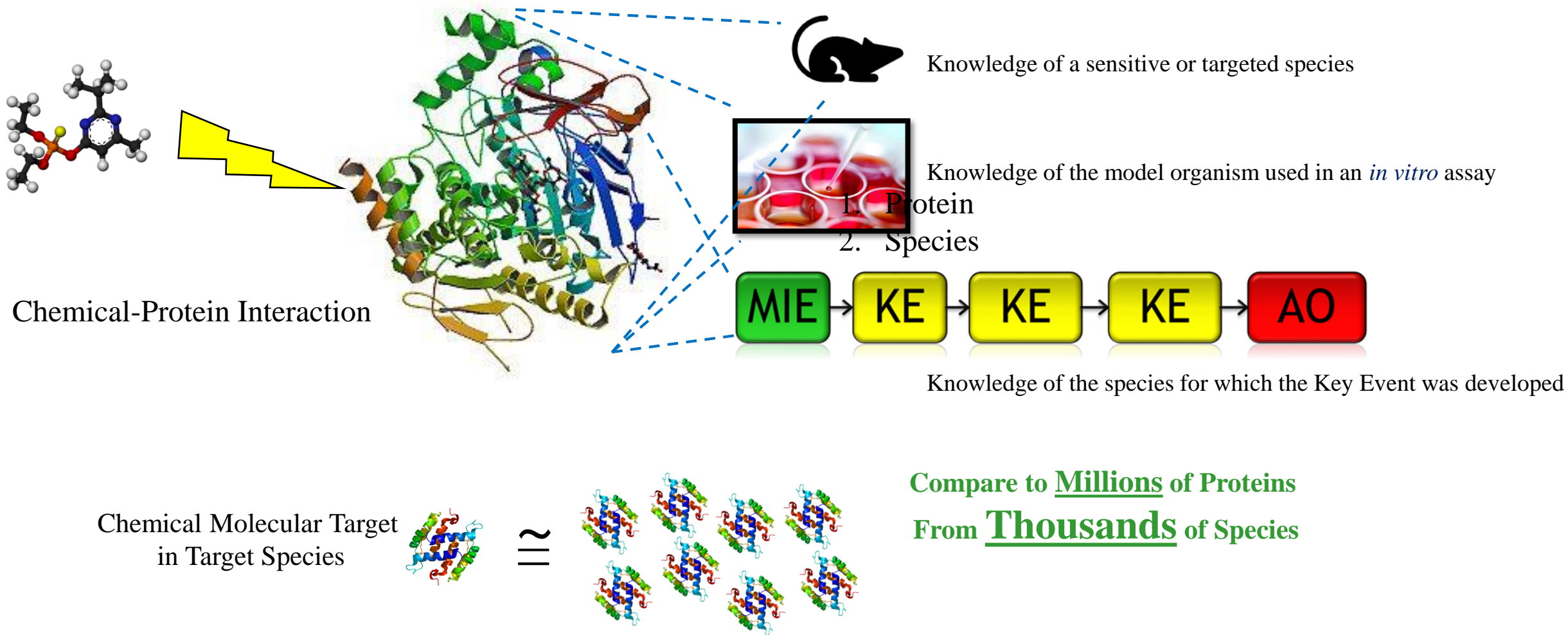
## Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): A Web-Based Tool for Addressing the Challenges of Cross-Species Extrapolation of Chemical Toxicity

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





# What information is required for a SeqAPASS query?

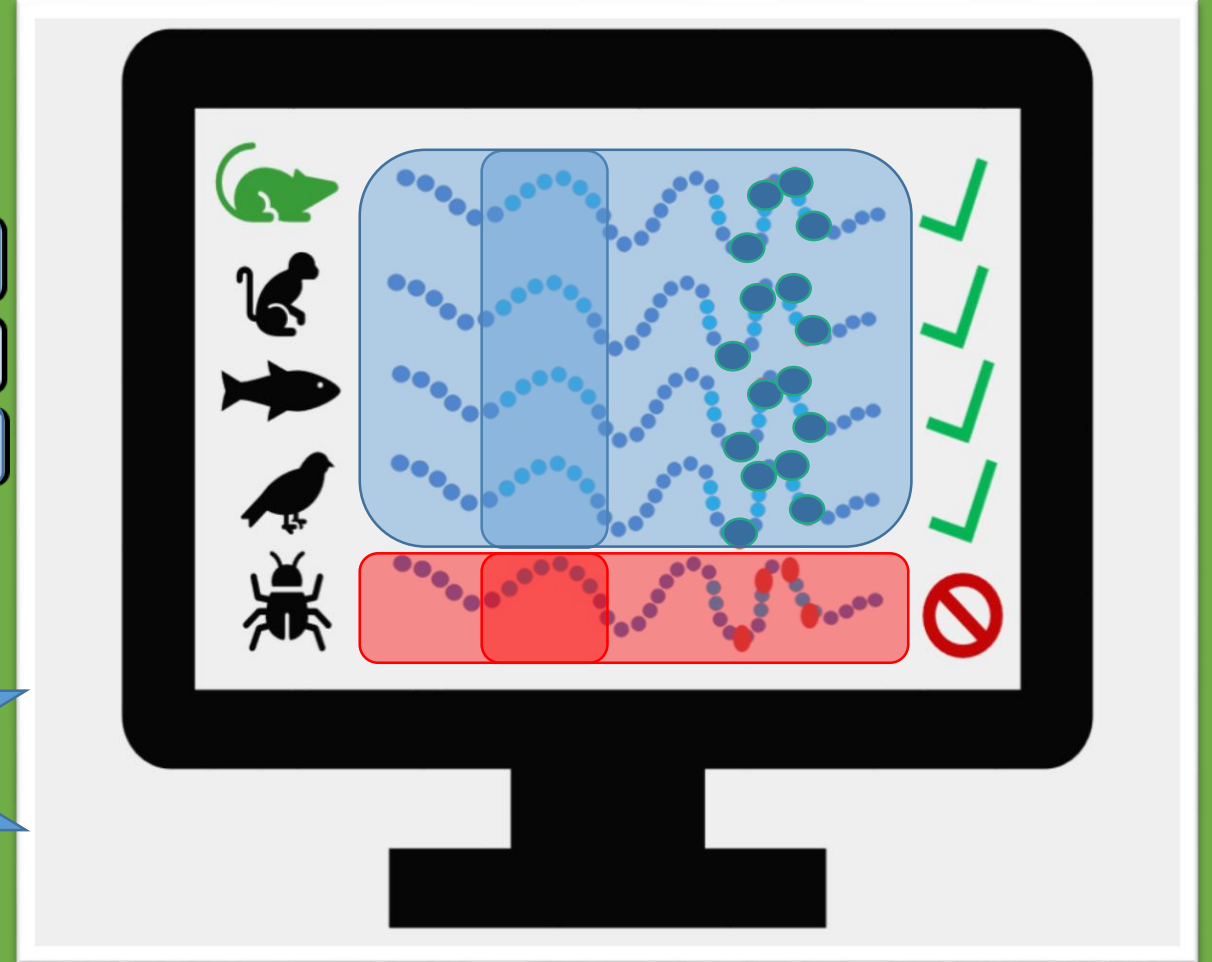


**Greater similarity = Greater likelihood that chemical can act on the protein**  
**Line of Evidence: Predict Potential Chemical Susceptibility Across Species**

### Flexible Analysis Based On Available Data

- Level 1** Primary Amino Acid Sequence Alignments
- Level 2** Conserved Functional Domain Alignments
- Level 3** Critical (Close Contact) Amino Acid Conservation

[seqapass.epa.gov/seqapass/](http://seqapass.epa.gov/seqapass/)

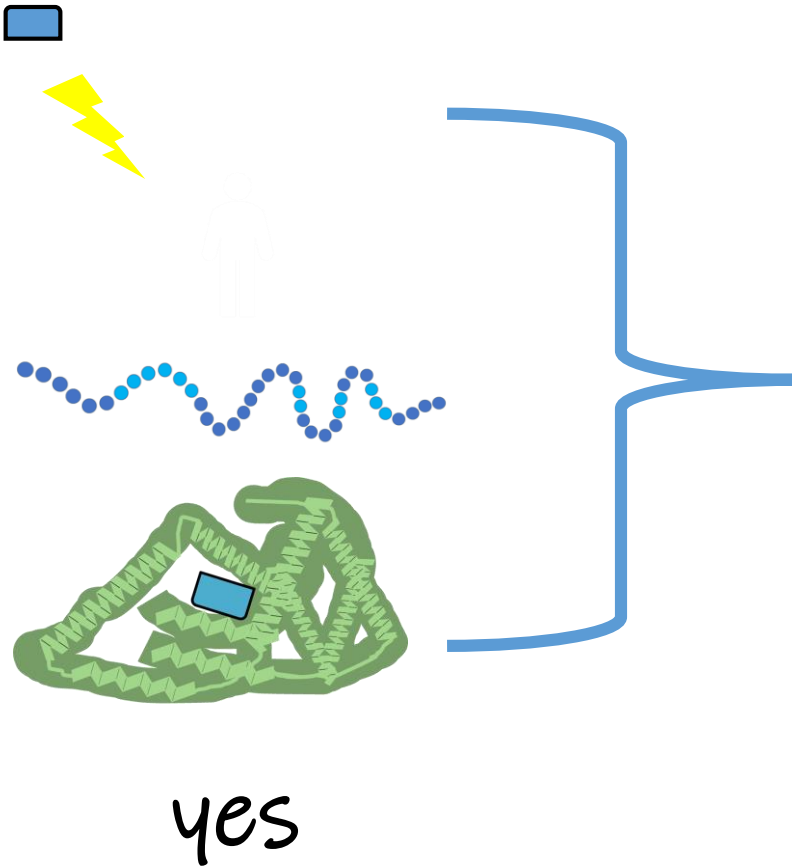






## Gather Lines of Evidence Toward Protein Conservation





# SeqAPASS Predicts Likelihood of Similar Susceptibility based on Sequence Conservation:

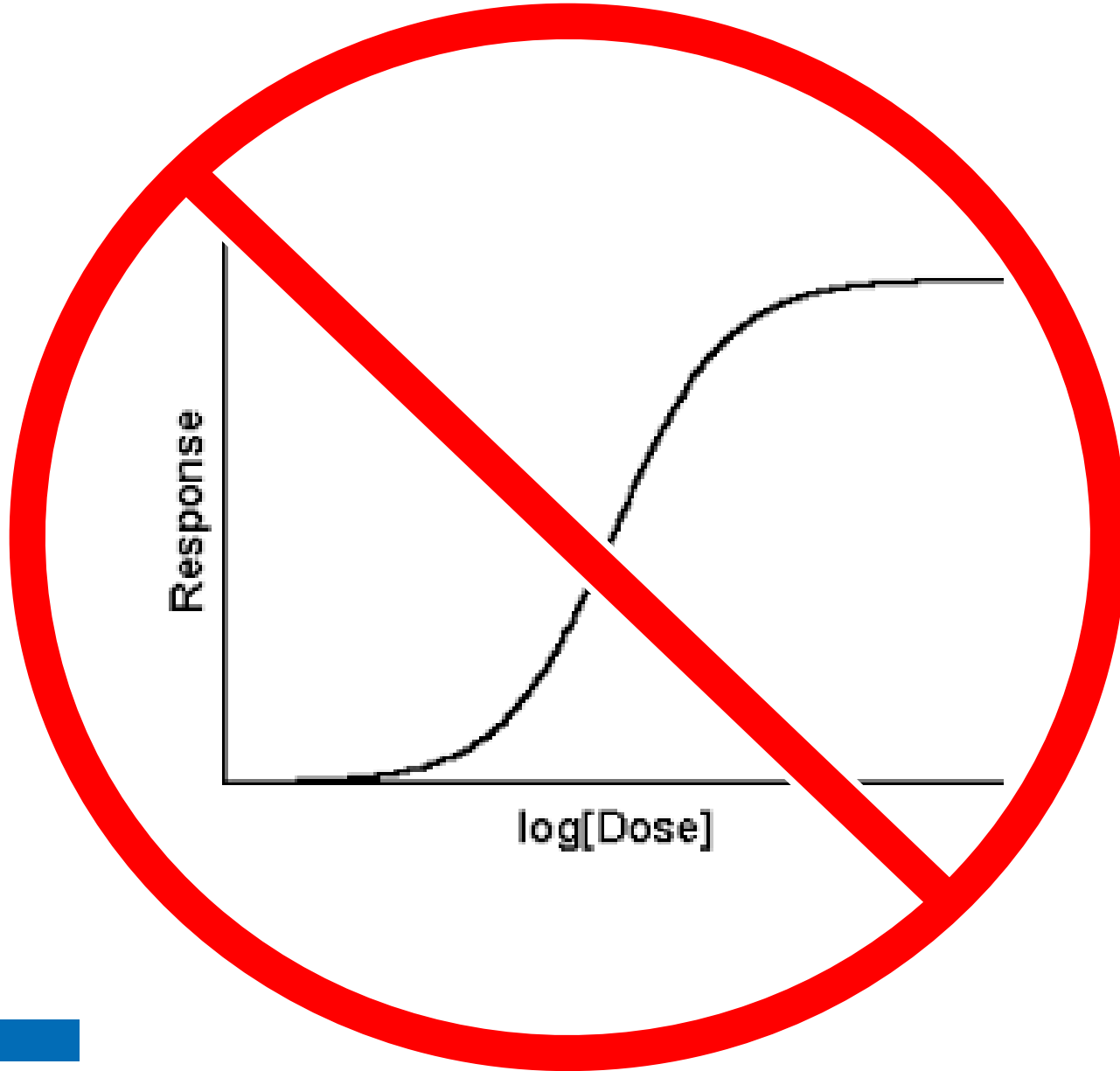


	yes
	yes
	yes
	yes
	yes
	yes
	yes
	no
	yes
	no

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.



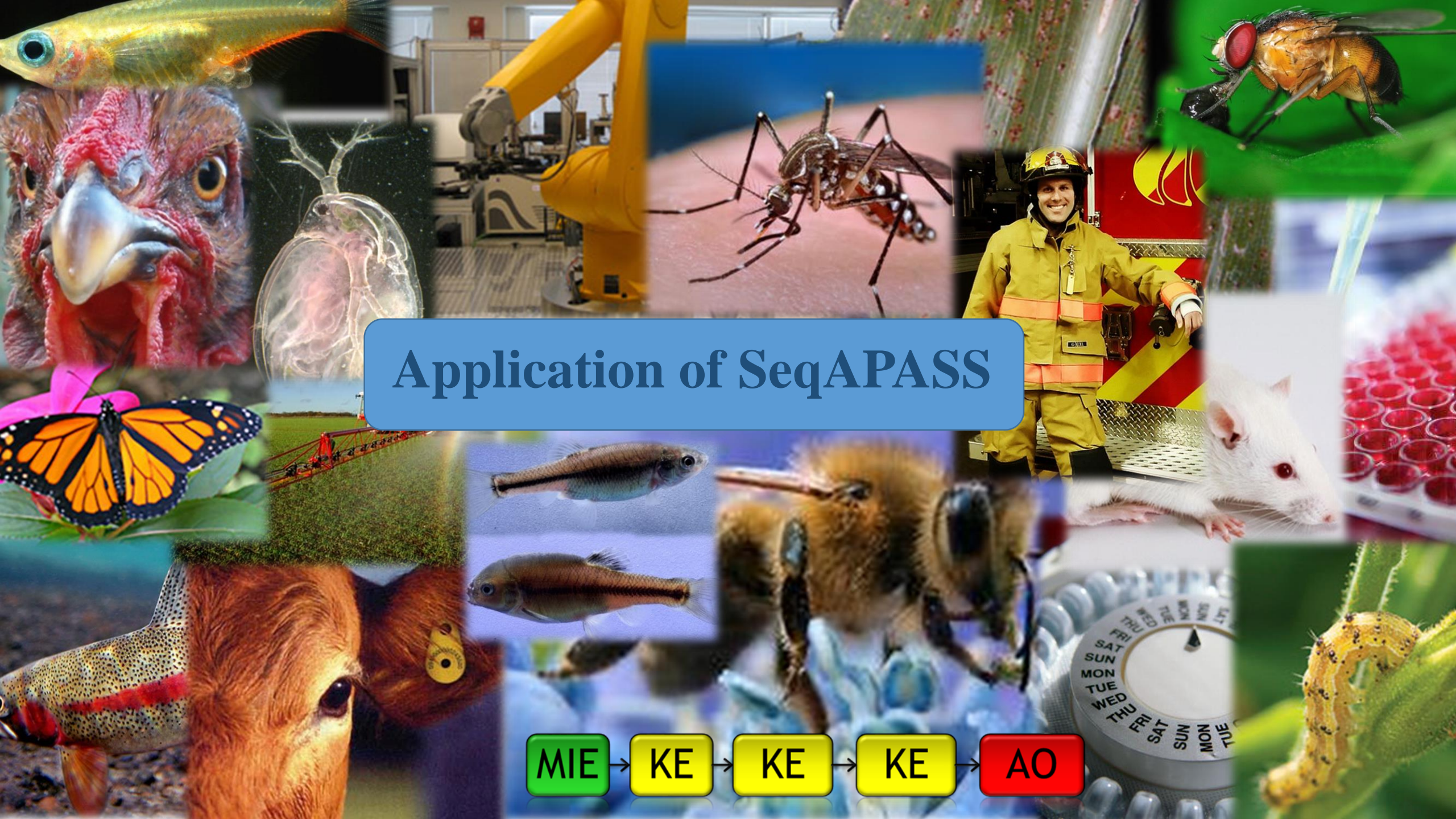




- Publicly available to all
- Lines of evidence for conservation for 100s-1000s of species rapidly
- Takes advantage of well-established tools and databases
- Streamlined, consistent, transparent, and published methods
  - Case examples to demonstrate applications
- Guides users to appropriate input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

- 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





# Application of SeqAPASS





# SeqAPASS for the butterflies

Evaluation to Predict Chemical Susceptibility  
to Threatened/Endangered Species

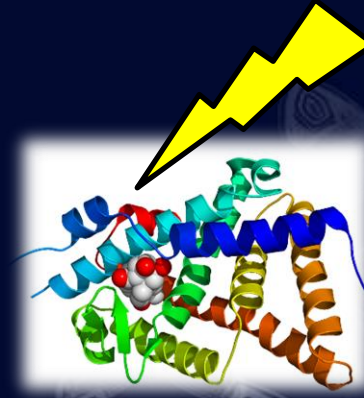


Order: Lepidoptera



# Molt-accelerating Insecticides

- Methoxyfenozide and Tebufenozide
  - Molecular target ecdysteroid receptor (EcR)
  - Target species: pest insects from **Lepidopteran** order
    - Armyworm, budworm, moth, corn borer
  - **Low toxicity** to non-target organism
    - Vertebrates, honey bees, earthworms



What about beneficial Lepidopterans?



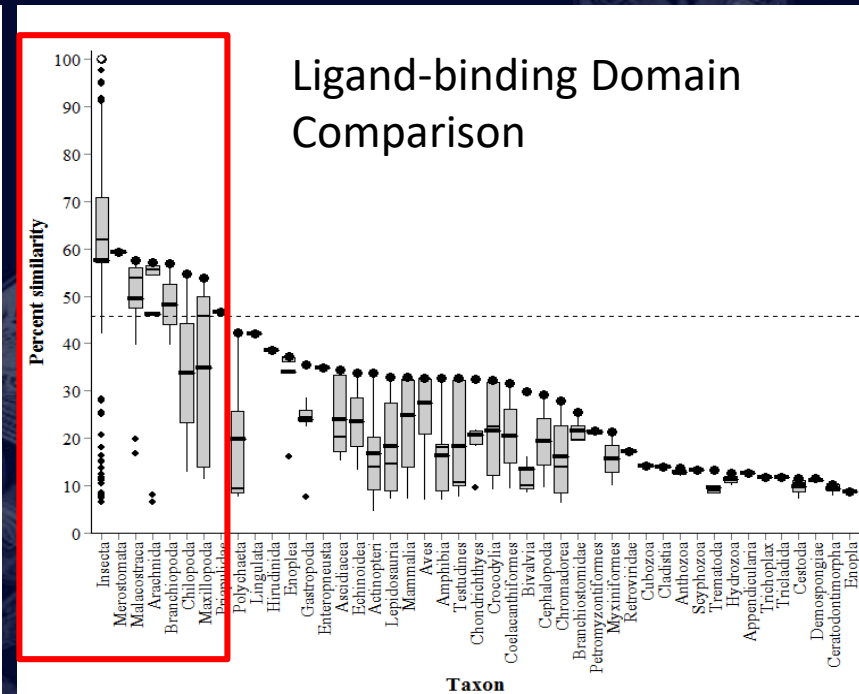
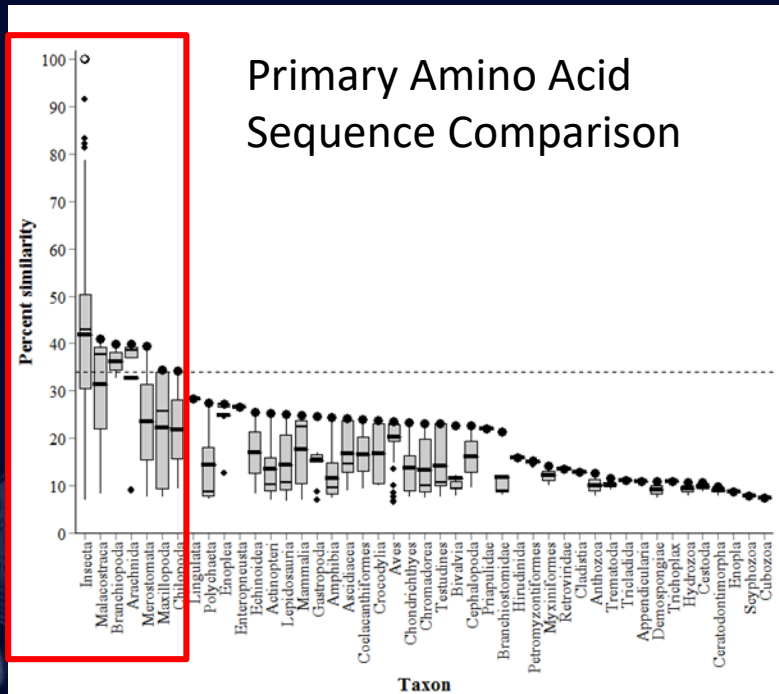
Threatened or Endangered  
Butterfly Species



?

# SeqAPASS Results

- Level 1
  - Ecdysone receptor is well conserved across invertebrates
- Level 2
  - Conservation of the ligand binding domain across Insect, Merostomata, Malacostraca, Arachnida, Branchiopoda, Maxillopoda, Priapulidae





# SeqAPASS Level 3

Insect Classification	Scientific Name	Common Name	Order	Protein Name	<sup>a</sup> Q331	<sup>c</sup> P353	<sup>c</sup> M360	<sup>a</sup> V402	<sup>b</sup> V434	<sup>b</sup> Q521	<sup>b</sup> M525	<sup>c</sup> I527	<sup>c</sup> K532	<sup>b</sup> L536
Pest	<i>Heliothis virescens</i>	tobacco budworm	Lepidoptera	Ecdysone receptor	Q	P	M	V	V	Q	M	I	K	L
Pest	<i>Helicoverpa armigera</i>	cotton bollworm	Lepidoptera	ecdysone receptor	Q	P	M	V	V	Q	M	I	K	L
Pest	<i>Agrotis ipsilon</i>	black cutworm moth	Lepidoptera	EcR B-like protein	Q	P	M	V	V	Q	M	I	K	L
Pest	<i>Spodoptera littoralis</i>	African cotton leafworm	Lepidoptera	ecdysone receptor	Q	P	M	V	V	Q	M	I	K	L
Pest	<i>Spodoptera exigua</i>	beet armyworm	Lepidoptera	ecdysone receptor	Q	P	M	V	V	Q	M	I	K	L
Pest	<i>Spodoptera litura</i>	moths	Lepidoptera	ecdysone receptor	Q	P	M	V	V	Q	M	I	K	L
Pest	<i>Amyelois transitella</i>	moths	Lepidoptera	PREDICTED: ecdysone receptor isoform X1	Q	P	M	V	V	Q	M	I	K	L
Pest	<i>Scirpophaga incertulas</i>	moths	Lepidoptera	ecdysone receptor	Q	P	M	V	V	Q	M	I	K	L
Pest	<i>Chilo suppressalis</i>	striped riceborer	Lepidoptera	ecdysone receptor B1 isoform	Q	P	M	V	V	Q	M	I	K	L
Pest	<i>Omphisa fuscidentalis</i>	bamboo borer	Lepidoptera	ecdysone receptor B1 isoform	Q	P	M	V	V	Q	M	I	K	L
Beneficial	<i>Papilio machaon</i>	common yellow swallowtail	Lepidoptera	PREDICTED: ecdysone receptor isoform X1	Q	P	M	V	V	Q	M	I	K	L
Beneficial	<i>Papilio xuthus</i>	Asian swallowtail	Lepidoptera	PREDICTED: ecdysone receptor isoform X1	Q	P	M	V	V	Q	M	I	K	L
Pest	<i>Plodia interpunctella</i>	Indianmeal moth	Lepidoptera	ecdysone receptor	Q	P	M	V	V	Q	M	I	K	L
Beneficial	<i>Danaus plexippus</i>	monarch butterfly	Lepidoptera	ecdysteroid receptor EcR-B	Q	P	M	V	V	Q	M	I	K	L
Pest	<i>Plutella xylostella</i>	diamondback moth	Lepidoptera	ecdysteroid receptor	Q	P	M	V	V	Q	M	I	K	L
Pest	<i>Choristoneura fumiferana</i>	eastern spruce budworm	Lepidoptera	ecdysteroid receptor EcR-B	Q	P	M	V	V	Q	M	I	K	L
Pest	<i>Ectropis obliqua</i>	moths	Lepidoptera	ecdysteroid receptor	Q	P	M	V	V	Q	M	I	K	L
Pest	<i>Grapholitha molesta</i>	moths	Lepidoptera	ecdysone receptor	Q	P	M	V	V	Q	M	I	K	L
Pest	<i>Manduca sexta</i>	tobacco hornworm	Lepidoptera	Ecdysone receptor	Q	P	M	V	V	Q	M	I	K	L
Beneficial	<i>Bombyx mori</i>	domestic silkworm	Lepidoptera	EcRB1	Q	P	M	V	V	Q	M	I	K	L
Pest	<i>Operophtera brumata</i>	winter moth	Lepidoptera	Ecdysone receptor A	Q	P	M	V	V	Q	M	I	K	L
Pest	<i>Sesamia nonagrioides</i>	Mediterranean corn borer	Lepidoptera	ecdysone receptor	Q	P	M	V	V	Q	M	I	K	L
Beneficial	<i>Bicyclus anynana</i>	quinting bush brown	Lepidoptera	ecdysteroid receptor	Q	P	M	V	V	Q	M	I	K	L
Beneficial	<i>Junonia coenia</i>	huckeye	Lepidoptera	ecdysteroid receptor	Q	P	M	V	V	Q	M	I	K	L
Beneficial	<i>Apis mellifera</i>	honey bee	Hymenoptera	ecdysteroid receptor A isoform	Q	K	I	M	T	Q	M	F	K	L
Beneficial	<i>Apis dorsata</i>	giant honeybee	Hymenoptera	PREDICTED: ecdysone receptor-like	Q	K	I	M	T	Q	M	F	K	L
Beneficial	<i>Apis florea</i>	little honeybee	Hymenoptera	PREDICTED: ecdysone receptor isoform X2	Q	K	I	M	T	Q	M	F	K	L
Beneficial	<i>Bombus terrestris</i>	buff-tailed bumblebee	Hymenoptera	PREDICTED: ecdysone receptor isoform X2	Q	K	I	M	T	Q	M	F	K	L
Beneficial	<i>Bombus impatiens</i>	common eastern bumble bee	Hymenoptera	PREDICTED: ecdysone receptor isoform X2	Q	K	I	M	T	Q	M	F	K	L
Beneficial	<i>Melipona quadrifasciata</i>	bees	Hymenoptera	Ecdysone receptor	Q	K	I	M	T	Q	M	F	K	L
Pest	<i>Polistes dominula</i>	European paper wasp	Hymenoptera	EcR	Q	R	I	M	T	Q	M	Y	K	L
Beneficial	<i>Habropoda laboriosa</i>	bees	Hymenoptera	Ecdysone receptor	Q	K	I	M	T	Q	M	F	K	L
Beneficial	<i>Orius laevigatus</i>	Insidious flower bug	Hemiptera	Amar et al. 2012	H	K	I	M	V	Q	M	F	Q	L
Pest	<i>Nezara viridula</i>	southern green stink bug	Hemiptera	ecdysone receptor isoform	H	R	I	M	V	Q	M	F	Q	L
Pest	<i>Nilaparvata lugens</i>	brown planthopper	Hemiptera	ecdysteroid receptor	H	R	I	T	V	Q	M	F	K	L
Pest	<i>Bemisia tabaci</i>	sweet potato white fly	Hemiptera	-	H	R	I	M	T	M	T	F	K	V
Beneficial	<i>Hydropsyche incognita</i>	caddisflies	Trichoptera	EcR	Q	P	I	M	I	R	M	I	K	L
Non-insect	<i>Daphnia magna</i>	crustaceans	Branchiopoda	ecdysone receptor A1	Q	K	M	C	T	L	M	F	K	L
Non-insect	<i>Ixodes scapularis</i>	black-legged tick	Arachnida	AamEcRA1, putative	S	R	I	G	S	M	M	F	Q	L
Non-insect	<i>Oreochromis niloticus</i>	Nile tilapia	Actinopteri	PREDICTED: oxysterols receptor LXR-alpha	-	R	L	T	F	V	Q	F	Q	L
Non-insect	<i>Xenopus (Silurana) tropicalis</i>	western clawed frog	Amphibia	oxysterols receptor LXR-beta	K	R	L	T	F	V	Q	F	Q	L
Non-insect	<i>Homo sapiens</i>	human	Mammalia	oxysterols receptor LXR-alpha isoform 4	R	R	L	T	F	V	Q	F	Q	L
Non-insect	<i>Picoides pubescens</i>	downy woodpecker	Aves	Oxysterols receptor LXR-alpha	Q	E	L	T	F	V	Q	F	Q	L

- Conservation of key amino acid residues across Lepidoptera, including beneficial butterfly species, therefore predicting intrinsic susceptibility

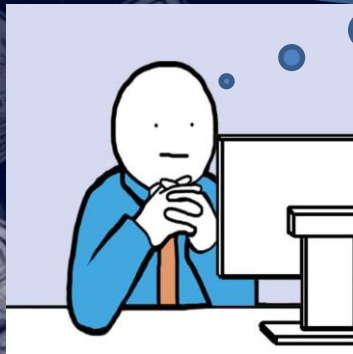


# EcR Conservation across Lepidopteran Order



SeqAPASS Level 1, 2, 3: EcR Conserved – Consistent line of evidence based on all butterfly species with sequence information available that the EcR is present for a chemical to act upon

Ah-ha....another line of evidence!!!



Endangered Species Act (ESA)

- Determination of “may affect”
- Determination of “may affect and is likely to adversely affect”
  - Line of evidence for chemical susceptibility

Government

Industry

# Consortium to Advance Cross Species Extrapolation in Regulation

## Steering Committee:

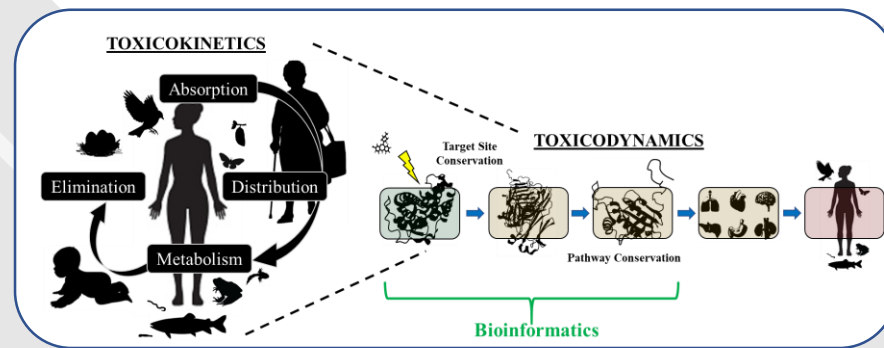
Carlie LaLone (US EPA)  
Geoff Hodges (Unilever)  
Nil Basu (McGill U)  
Steve Edwards (RTI)  
Fiona Sewell (NC3Rs)  
Michelle Embry (HESI)  
Patience Browne (OECD)

1. Define the taxonomic domain of applicability
2. Define the global regulatory landscape/need
3. Develop a bioinformatics toolbox
4. Communicate a shared scientific vision

Interested in Learning more or Joining: Contact [LaLone.Carlie@epa.gov](mailto:LaLone.Carlie@epa.gov) or [Geoff.Hodges@unilever.com](mailto:Geoff.Hodges@unilever.com)

Academia

NGO



# Acknowledgements

## U.S. EPA, ORD

Marissa Jensen (University of Minnesota Duluth)

Sally Mayasich (University of Wisconsin)

Monique Hazemi (ORISE)

Sara Vliet (past ORISE 2021)

Jon Doering (past NRC 2018)

Colin Finnegan (past ORISE 2018)

Donovan Blatz (past ORISE 2021)

## GDIT

Cody Simmons

Audrey Wilkinson

Wilson Menendez

Thomas Transue (past GDIT 2022)

SeqAPASS v6.0 (Released Sept. 2021)



[LaLone.Carlie@epa.gov](mailto:LaLone.Carlie@epa.gov)

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