United States Environmental Protection Agency

The Sequence Alignment to Predict Across Species Susceptibility Tool:

Extrapolating Knowledge Computationally

Presenter: Carlie A. LaLone, Ph.D.

May 11th 2022

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

EPA United States Environmental Protection Agency

Overview

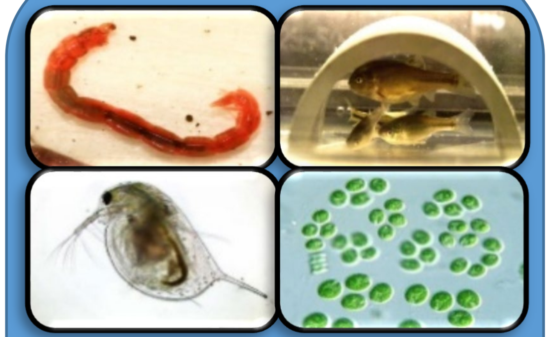
- Need for cross species extrapolation
- Bioinformatics to advance extrapolation
- The Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) tool
- Demonstrated applications
- Bringing champions in the field together to advance the science for action through an international consortium

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 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*

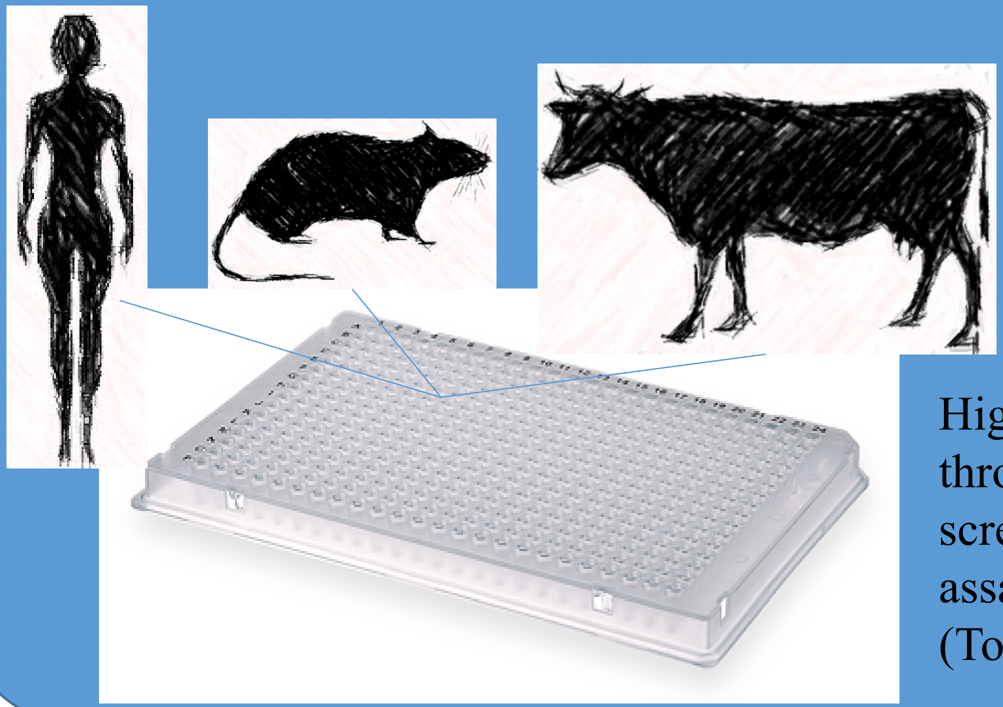
United States Environmental Protection Agency

Need for Advances in Species Extrapolation



High throughput transcriptomics

Historic whole organism toxicity testing

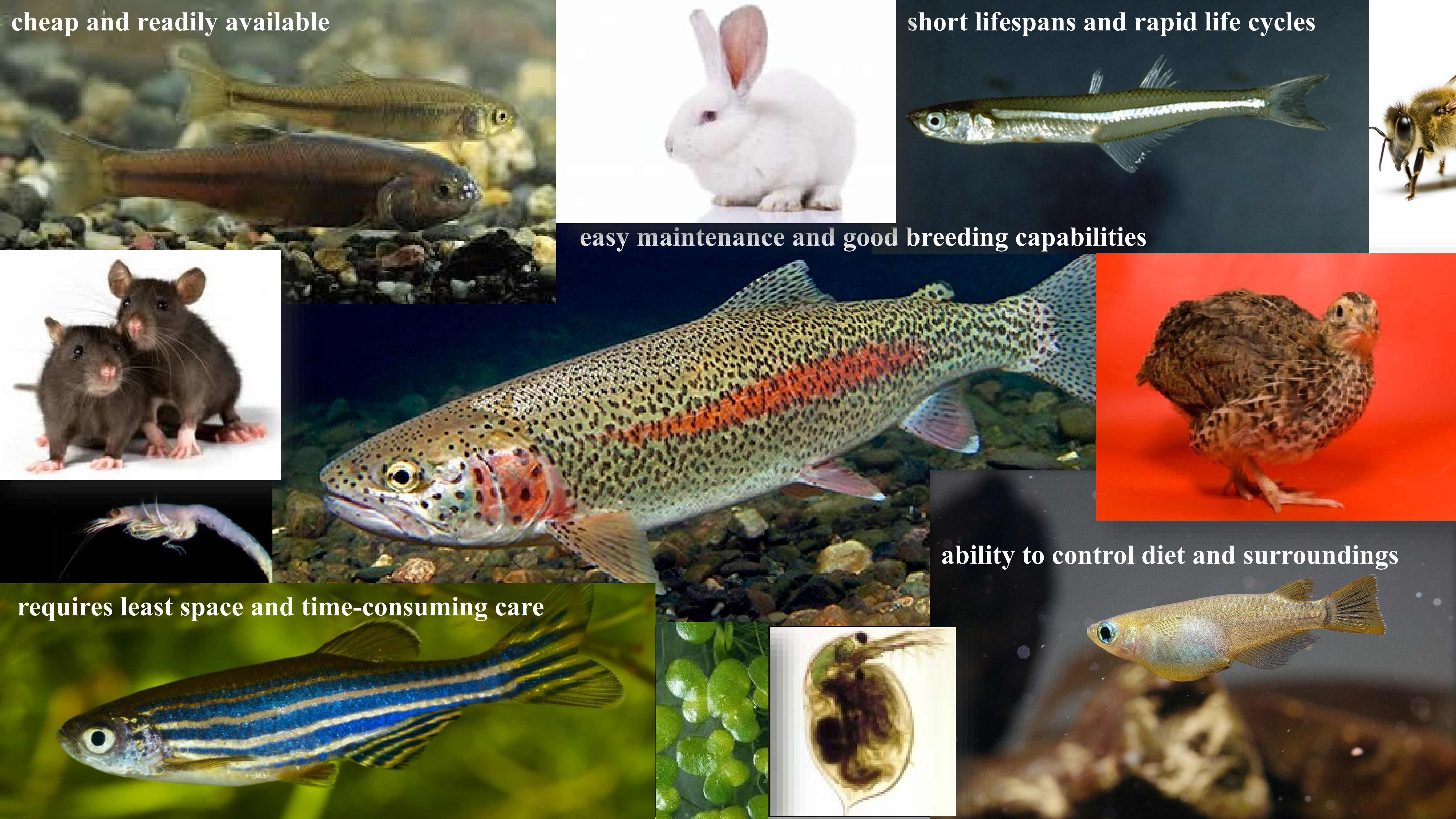


High-throughput screening assays (ToxCast)

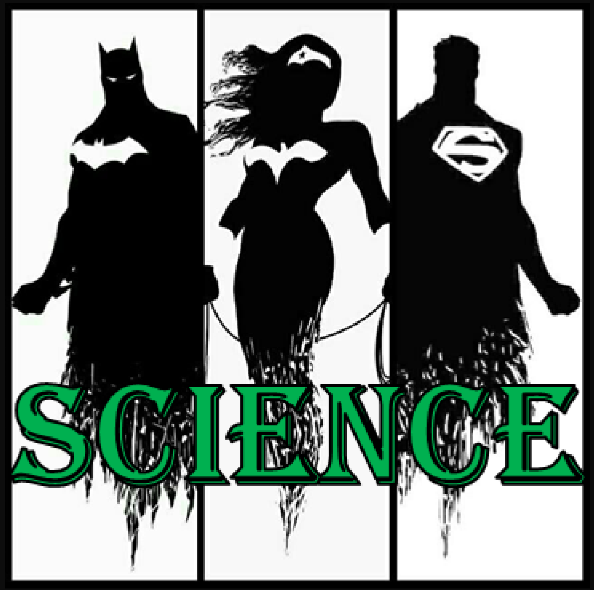
Define the taxonomic domain of applicability in AOP development

MIE → KE → KE → KE → AO

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



<u>What is it?</u>

- 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

<u>Why is it important:</u>

- 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

EPA United States Environmental Protection

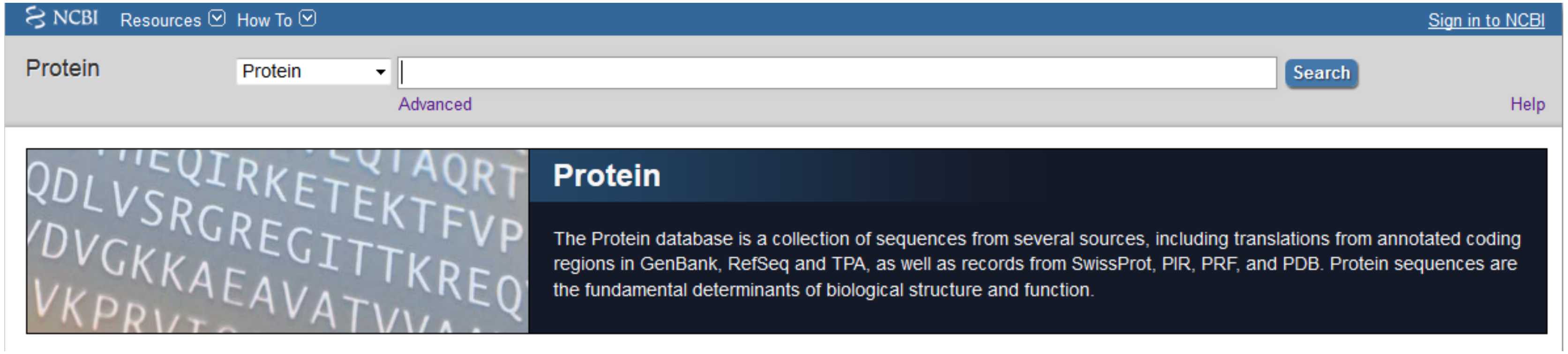
Where could we begin in understanding species similarities and differences?

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



Bioinformatics

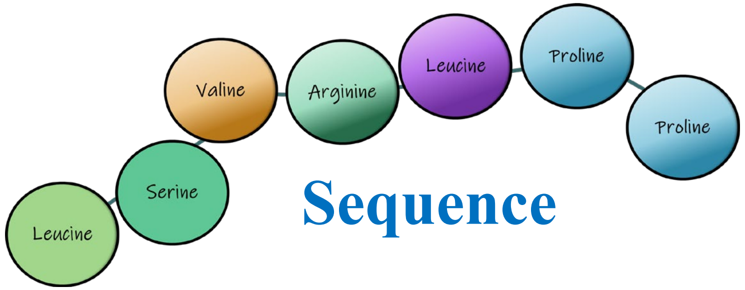
- Combines mathematics, information science, and biology to **answer biological questions**

- 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**

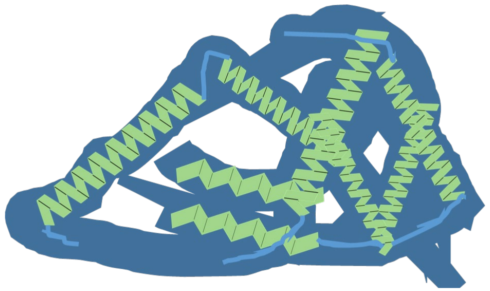
Let the computers do the work

EPA United States Environmental Protection Agency

Begin Simple and Advance as the Science Advances



Sequence

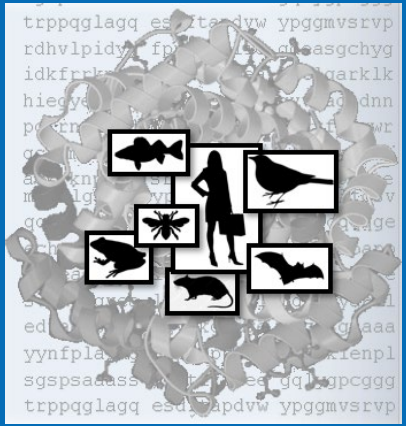


Structure

Function

Consider sequence and structural attributes to understand protein conservation across species

EPA
United States Environmental Protection Agency



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

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)

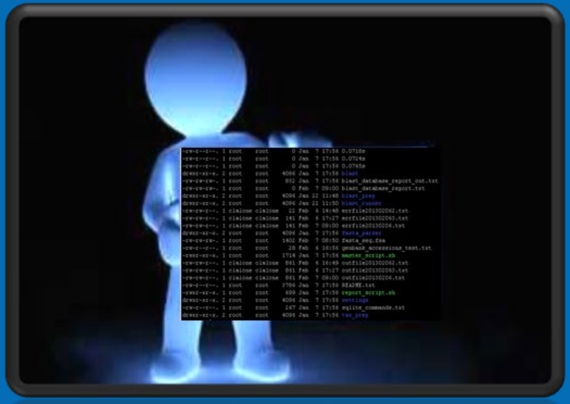
OXFORD
SOT Society of Toxicology
www.toxsci.oxfordjournals.org

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

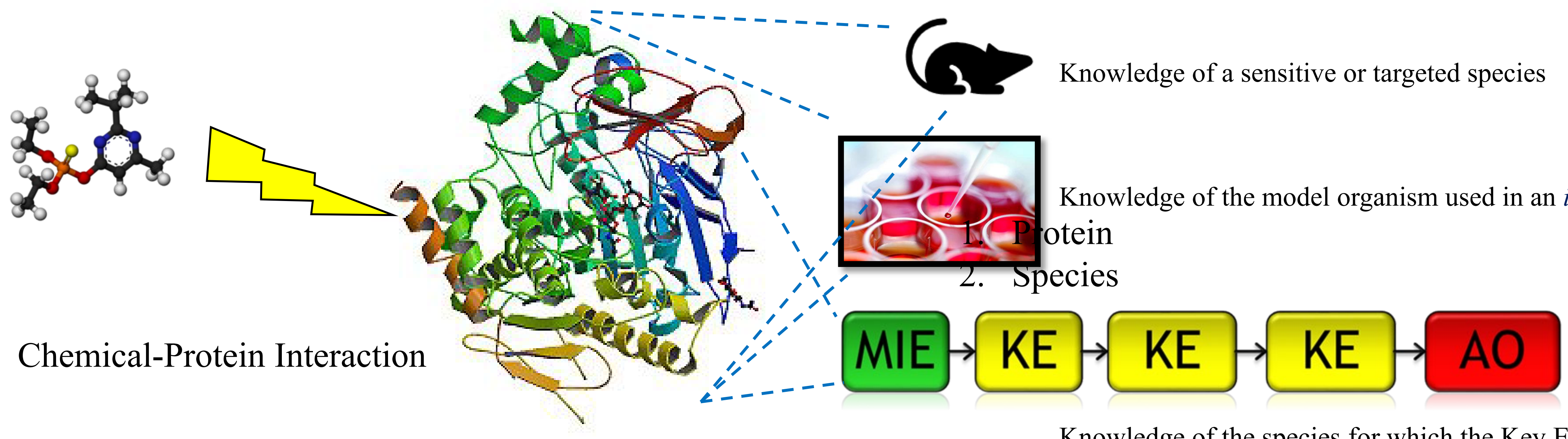
doi: 10.1093/toxsci/kfw119
Advance Access Publication Date: June 30, 2016
Research article

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[*]



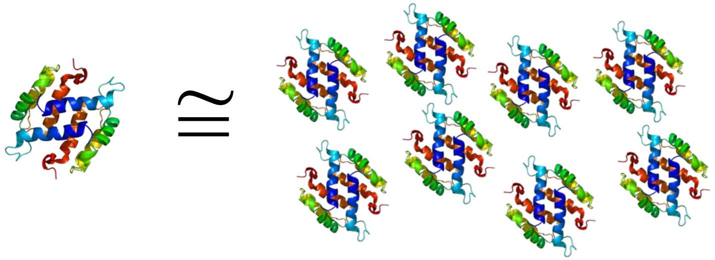
What information is required for a SeqAPASS query?



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

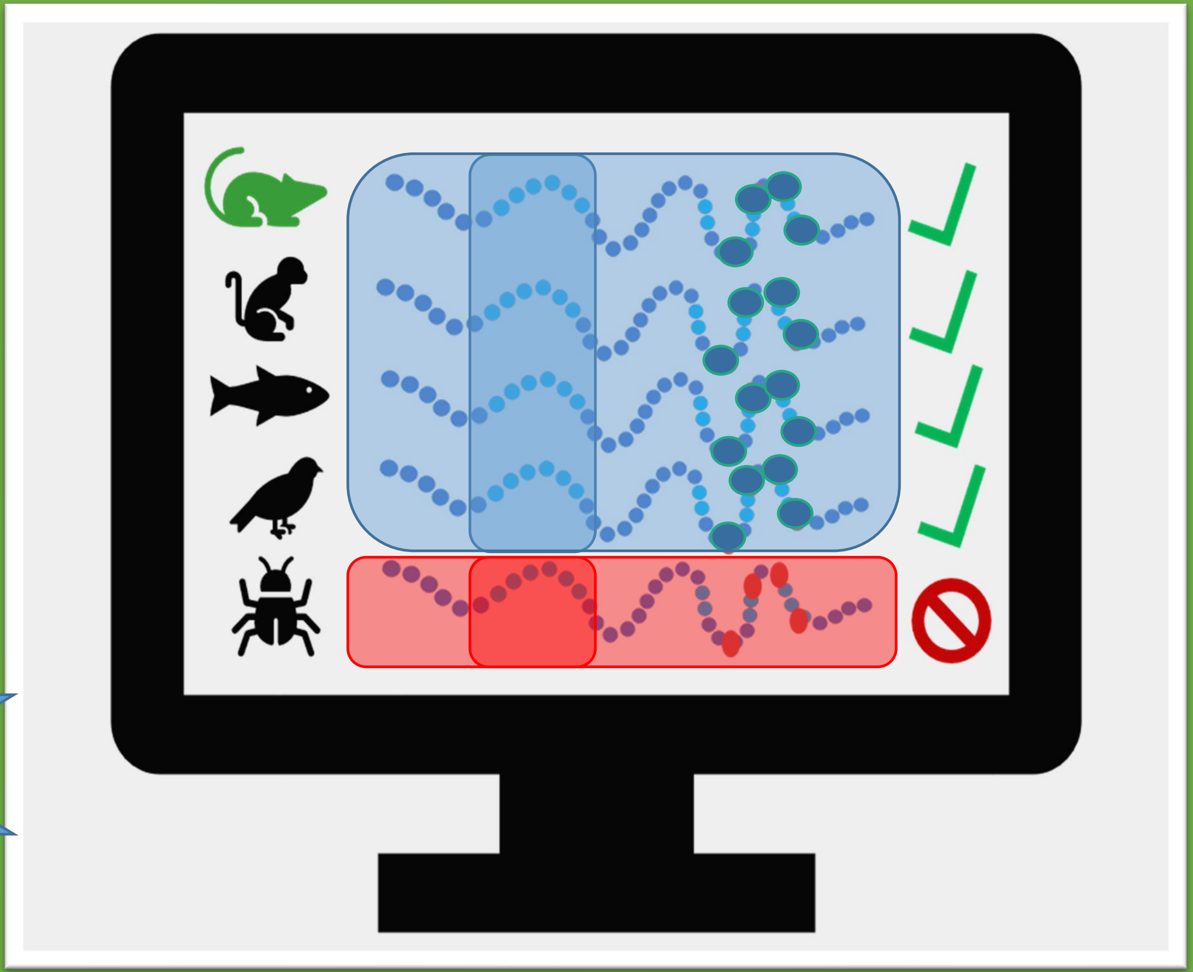
Flexible Analysis Based On Available Data

Level 1 Primary Amino Acid Sequence Alignments

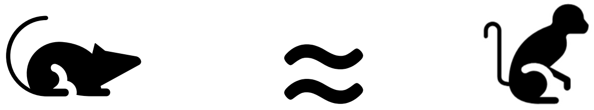
Level 2 Conserved Functional Domain Alignments

Level 3 Critical (Close Contact) Amino Acid Conservation

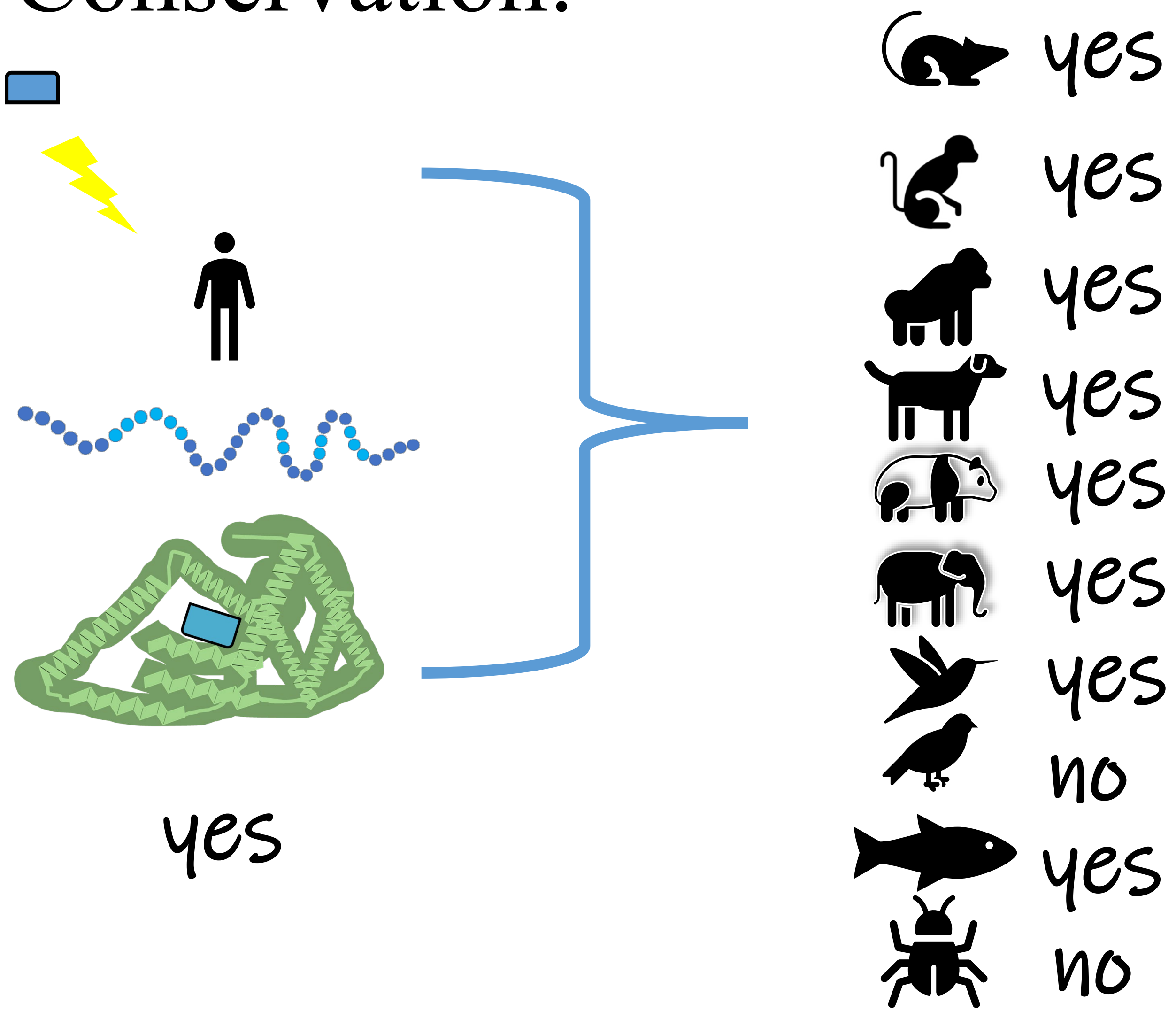
seqapass.epa.gov/seqapass/



Gather Lines of Evidence Toward Protein Conservation

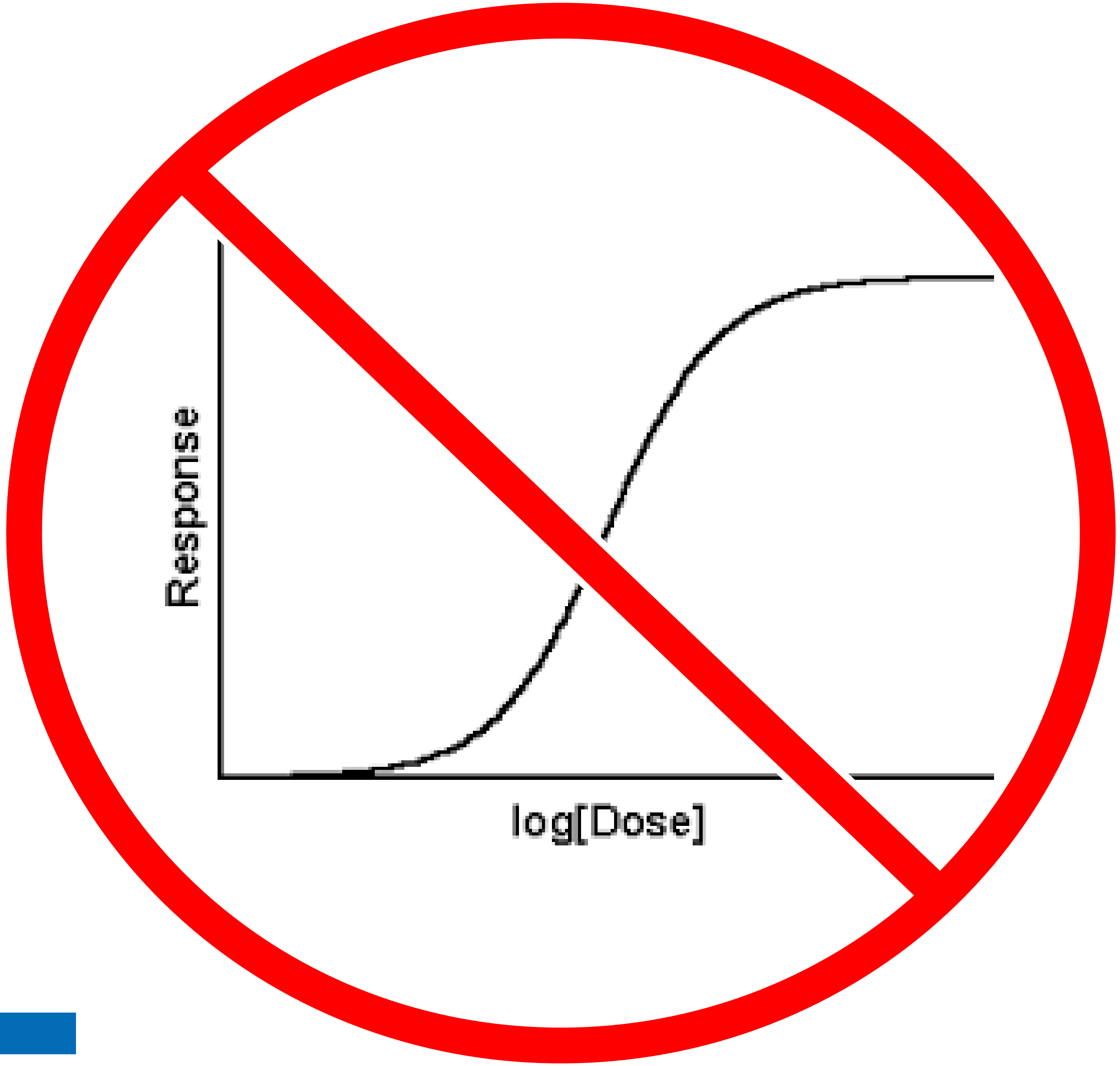


SeqAPASS Predicts Likelihood of Similar Susceptibility based on Sequence Conservation:



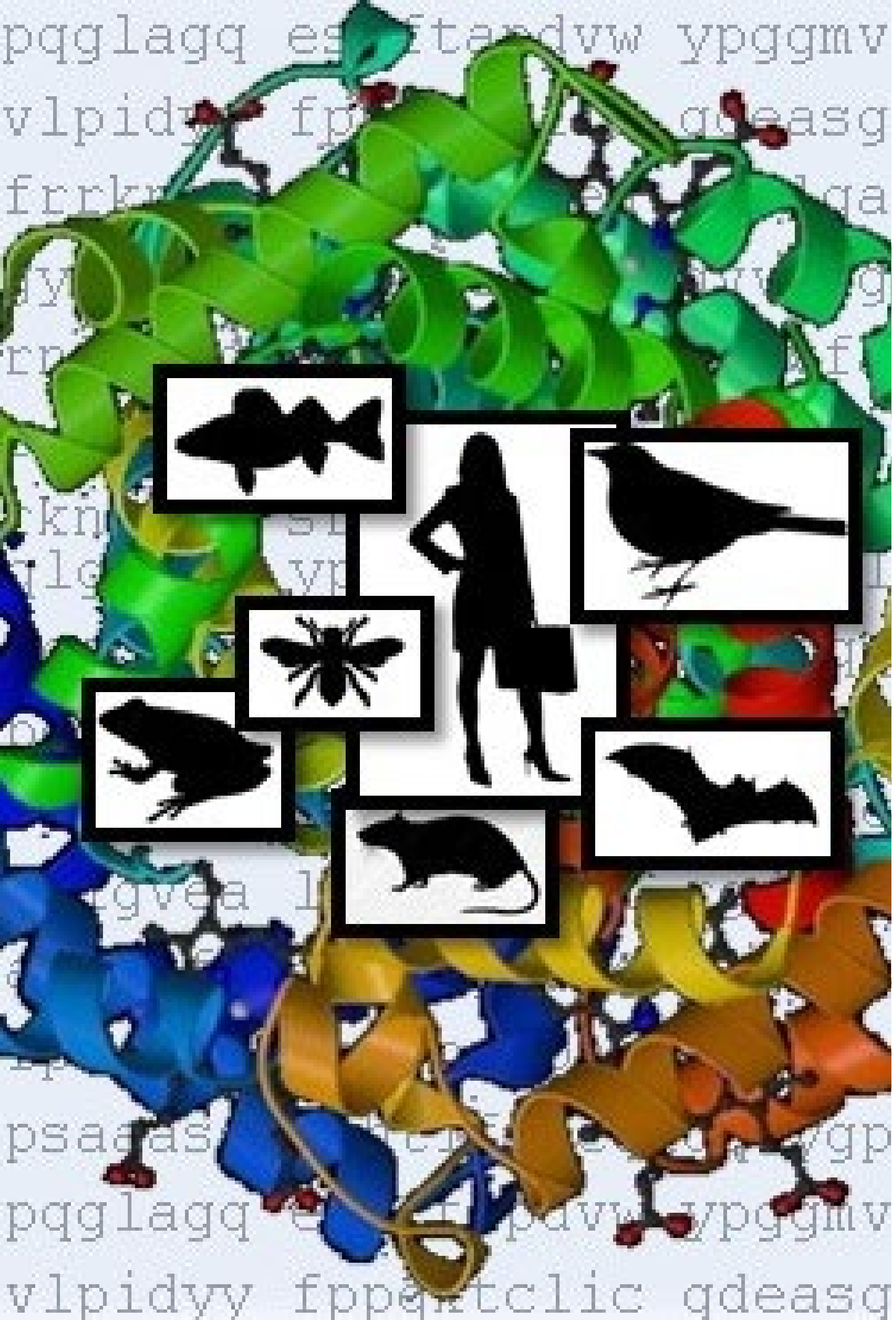
Line(s) of evidence indicate
- The protein is conserved
- The protein is NOT conserved

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

Application of SeqAPASS

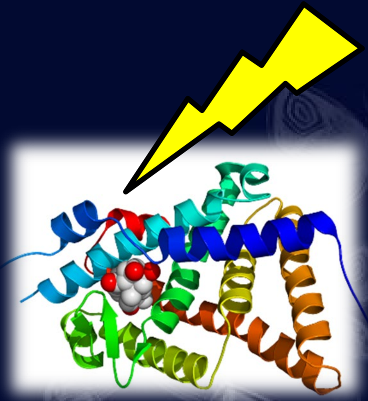
SeqAPASS for the butterflies

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



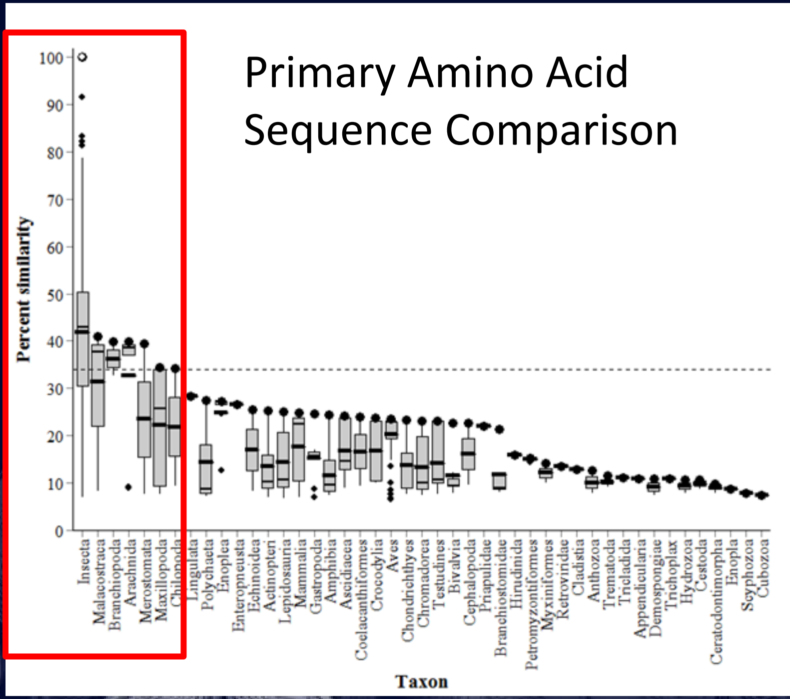
What about beneficial Lepidopterans?

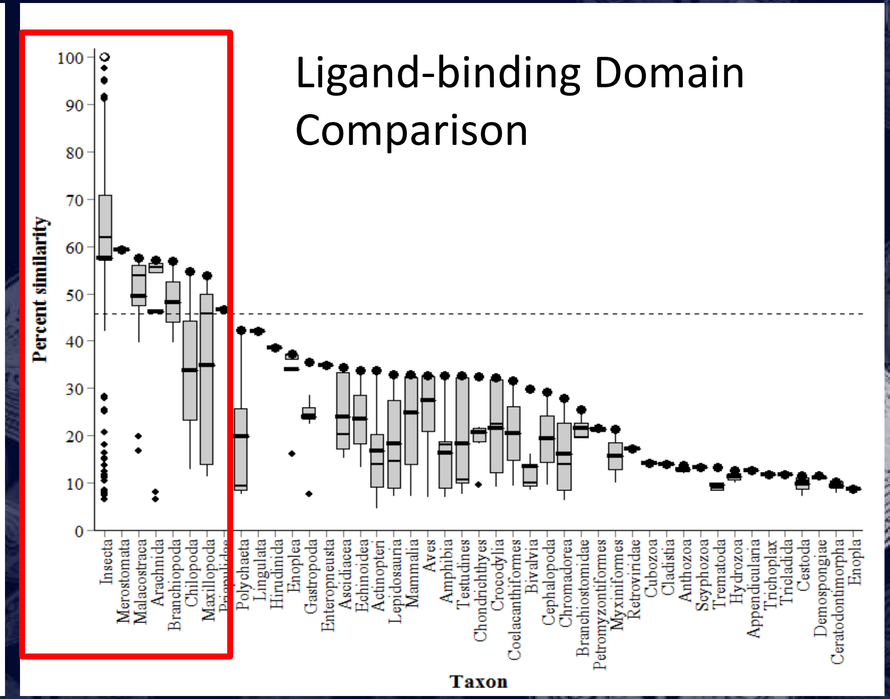
?

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

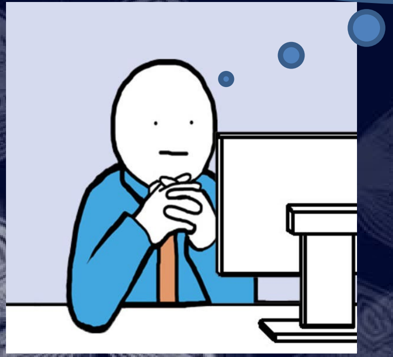
Insect Classification	Scientific Name	Common Name	Order	Protein Name	[b]Q331	[c]P353	[c]M360	[a]V402	[b]V434	[b]Q521	[b]M525	[c]I527	[c]K532	[b]L536
Pest	*Heliothis virescens*	tobacco budworm	Lepidoptera	Ecdysone receptor	Q	P	M	V	V	Q	M	I	K	L
Pest	*Helicoverpa armigera*	cotton bollworm	Lepidoptera	ecdysone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

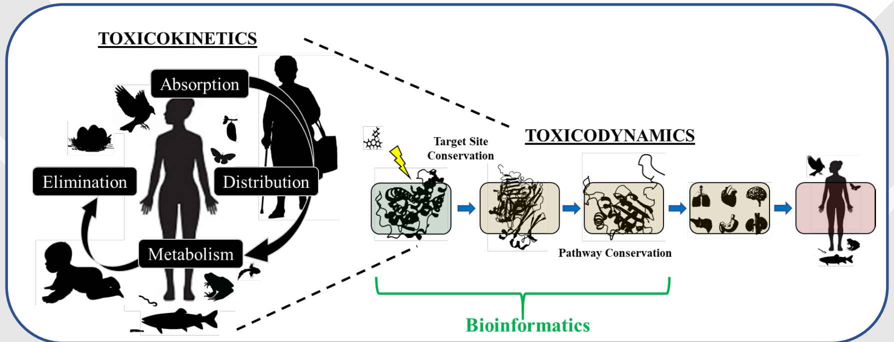
Industry

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 or Geoff.Hodges@unilever.com



Academia

NGO

EPA
United States
Environmental Protection
Agency

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

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