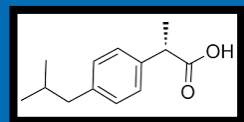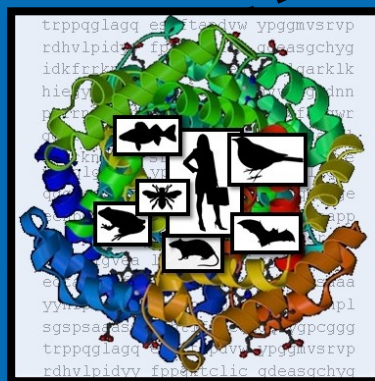# The Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) Tool: Catalyzing a Change in Species Extrapolation

Carlie A. LaLone, Ph.D.
Research Bioinformaticist

Predict Chemical Susceptibility Across All Species

Toxicity Data with Model Organism

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

May 25th, 2021

Chemicals make up the world around us – necessary for our modern society

Toxicity Testing to Understand Chemical Safety

- Regulatory decision-making

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
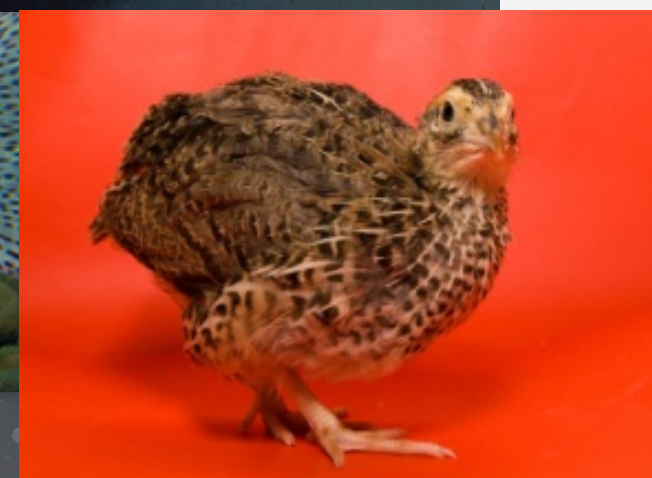
cheap and readily available

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

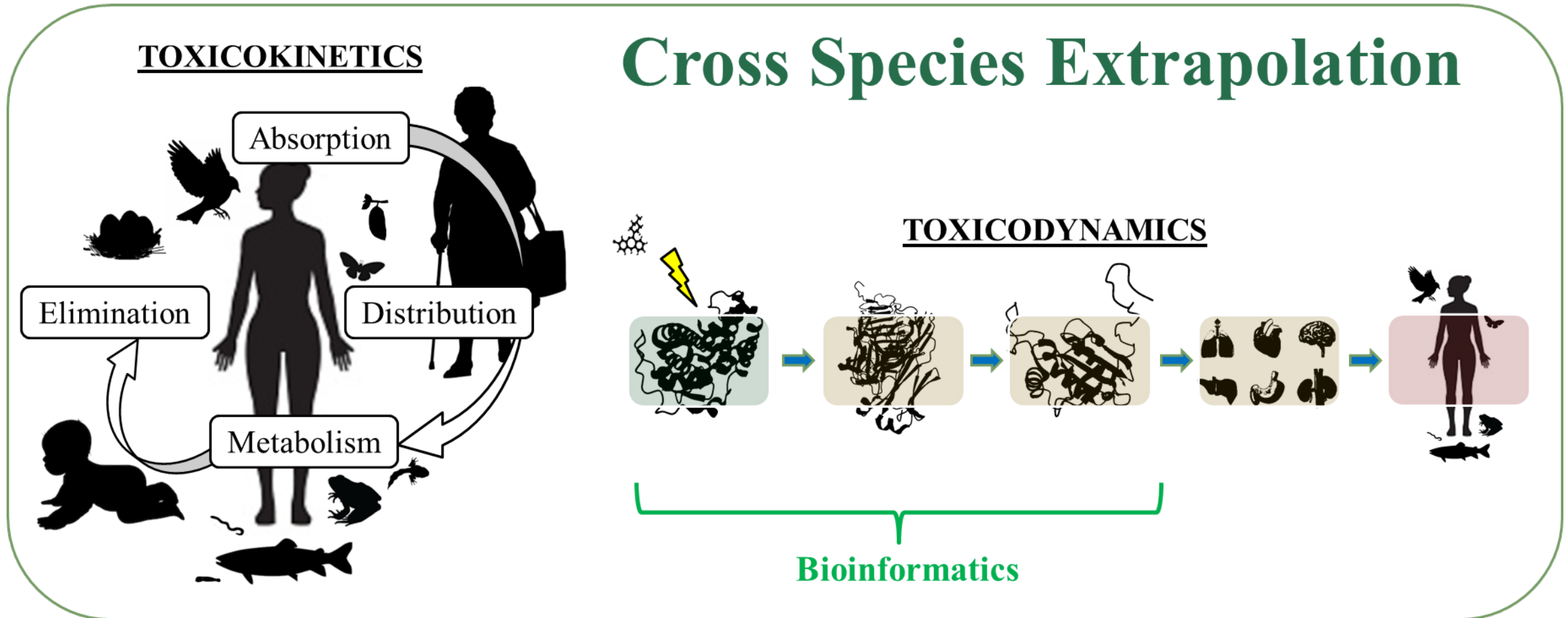Surrogate species (model organism)

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
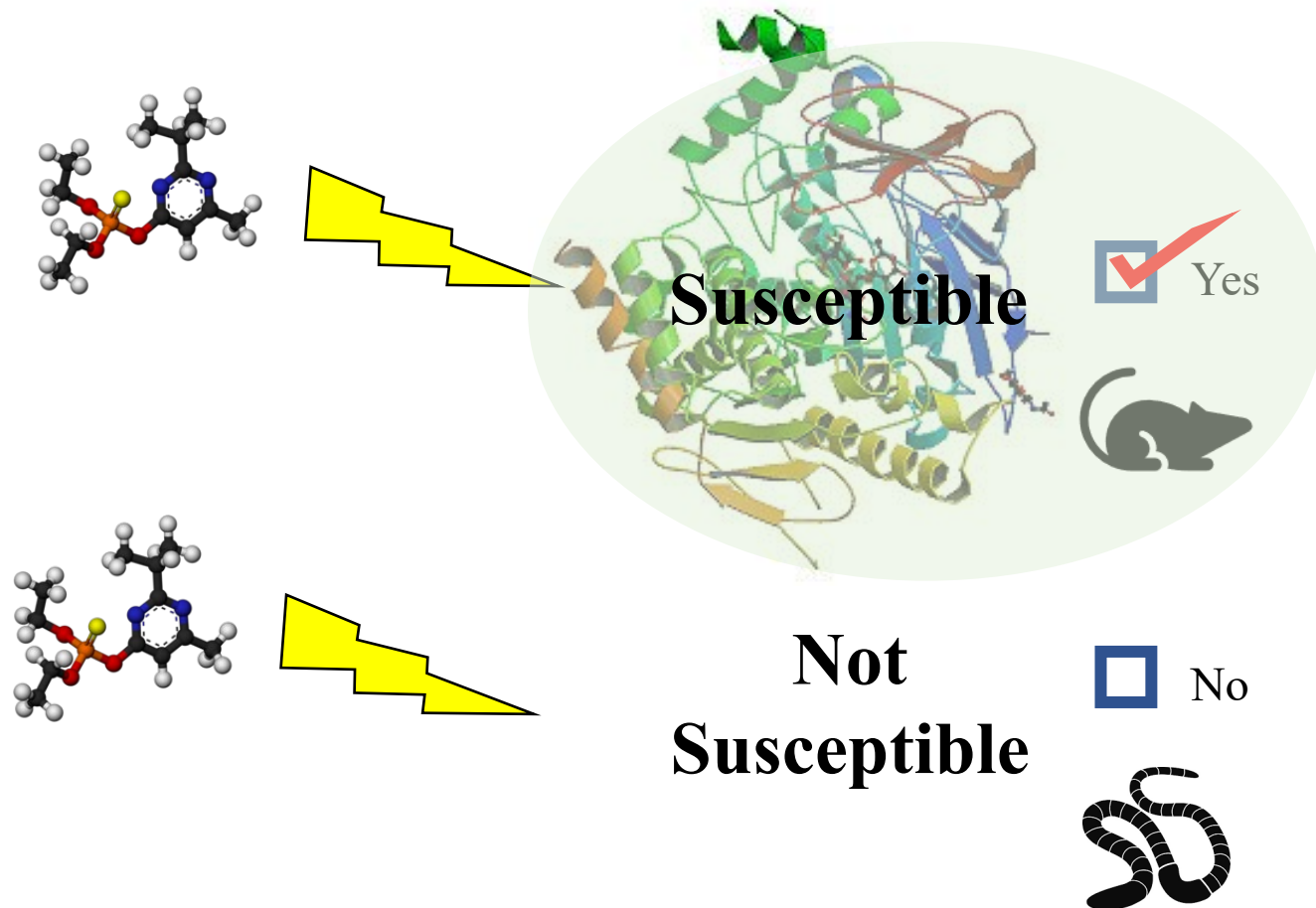 - Important challenge for species listed under the Endangered Species Act

Sensitivity to Chemical Perturbation

Considering chemical sensitivity?

Factors that make a species sensitive
- Exposure
- Dose
- ADME
- Target receptor availability
- Life stage
- Life history
- etc.
- etc.

Susceptible ☑ Yes

Not Susceptible ☐ No

Simple question to address:
Is the known chemical target available in a species for a chemical to act upon?
Yes   or No
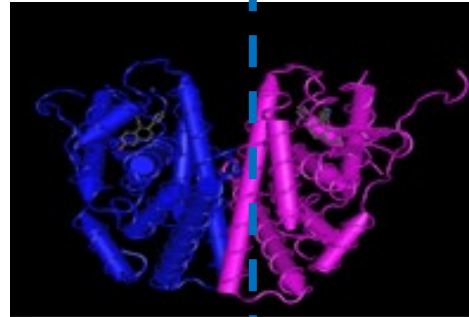Likely susceptible or Not likely susceptible (at least through the known mechanism)

Predictive Approaches

Start simple

Sequence

MTMTLHTKASGMALLHQIQGNELEPLNRPQLKIPLERPLGE
VYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYGQTGLPYG
PGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQ
PHGQQVPYYLENEPSGYTVREAGPPAFYRPNSDNRRQGGR
ERLASTNDKGSMAMESAKETRYCAVCNDYASGYHYGVWSC
EGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQACRLR
KCYEVGMMKGGIRKDRRGGRMLKHKRQRDDGEGRGEVG
SAGDMRAANLWPSPLMIKRSKKNSLALSLTADQMVSALLA
EPPILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKV
PGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFA
PNLLLDRNQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEF
VCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLIHLM
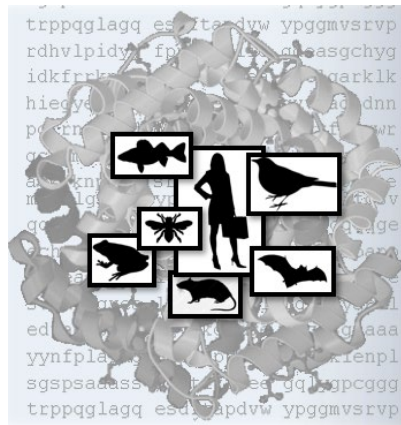
Structure

Function

Bioinformatics

Natural Ligands/Chemicals

Critical amino acids

Functional domain
(e.g. Ligand binding domain)

MIE → KE → KE → KE → AO

Similarity across species at the molecular level

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

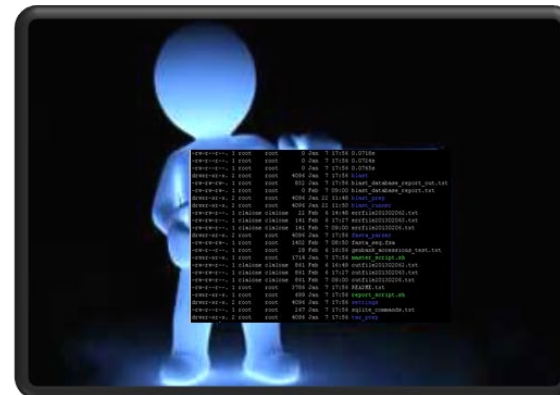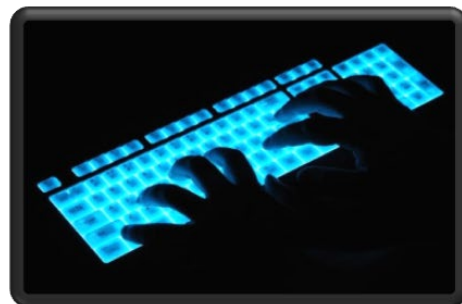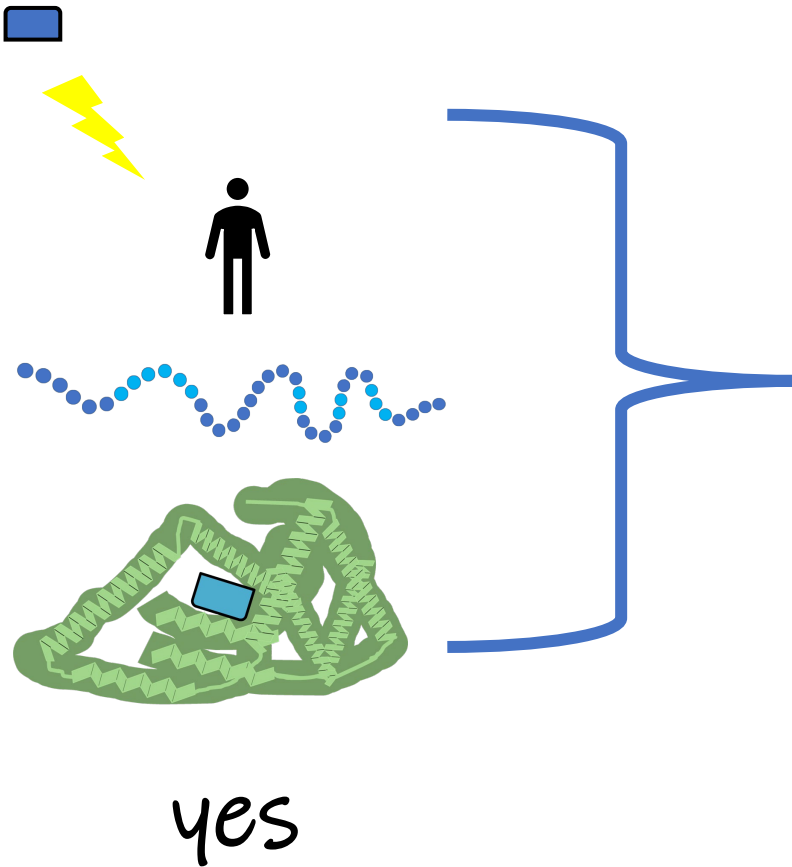# Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)

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

SeqAPASS Predicts Likelihood of Similar Susceptibility based on Sequence Conservation:

yes

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

yes
yes
yes
yes
yes
yes
yes
no
yes
no

SeqAPASS DOES NOT predict the degree of sensitivity/susceptibility:

Factors that make a species sensitive
- Exposure
- Dose
- ADME
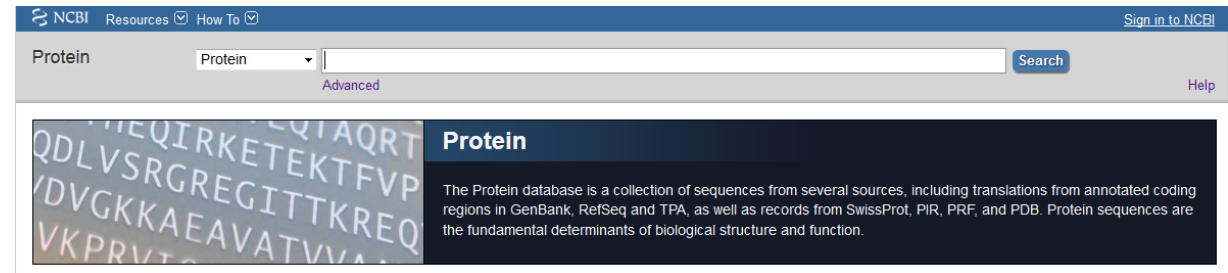- Target receptor availability
- Life stage
- Life history
- etc.
- etc.

Strengths of SeqAPASS

NCBI: 197,232,209 Proteins representing 108,257 Organisms

New tools and technologies
- Improved sequencing technologies
- Large databases of sequence data

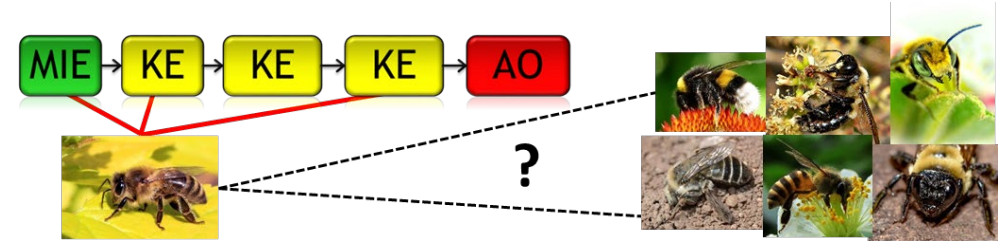- Publicly available to all
- Lines of evidence for conservation for 100s-1000s of species rapidly
- Takes advantage of well-established tools and databases
- Streamlined, consistent, transparent, and published methods
  - Case examples to demonstrate applications
- Guides users to appropriate input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

Applications of Bioinformatics: Case Studies

- **Extrapolate adverse outcome pathway knowledge across species**
  - Define the taxonomic relevance: Apis vs Non-Apis bees

- **Extrapolate high throughput screening data**
  - Chemicals that target human estrogen receptor alpha, androgen receptor, steroidogenic enzymes, thyroid axis proteins
  - All ToxCast Assay targets
- **Predict relative intrinsic susceptibility**
  - Pesticides
  - Endangered Species Act
  - Derivation of Aquatic Life Criteria
- **Predict chemical bioaccumulation across species**
  - Chemicals of concern: PFAS
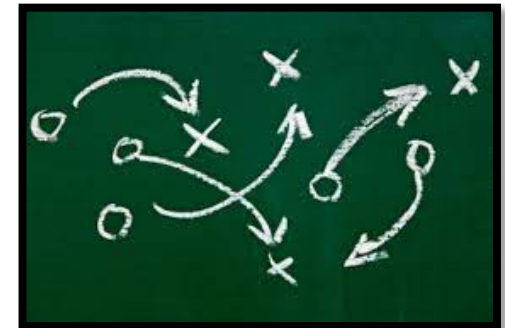- **Generate research hypotheses** Strobilurin fungicides
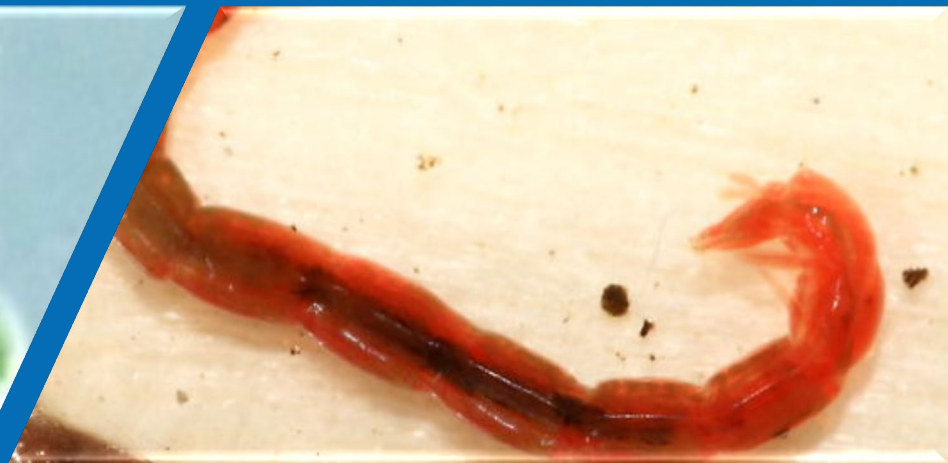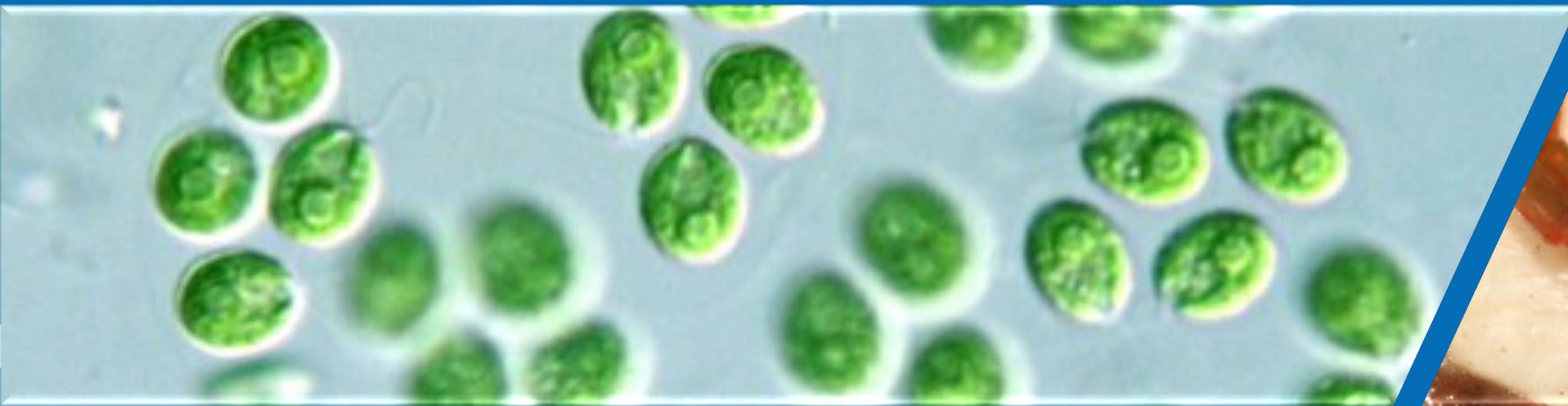- **Prioritization strategies** Pharmaceuticals

Where do we go from here?

- Connect bioinformatics with general informatics including systematic methods
  - ECOTOX Knowledgebase
  - Systematic literature review for WOE
- Advance structural evaluations (computing power and storage)
  - Move from docking to virtual screening
- Specific laboratory studies to support bioinformatics
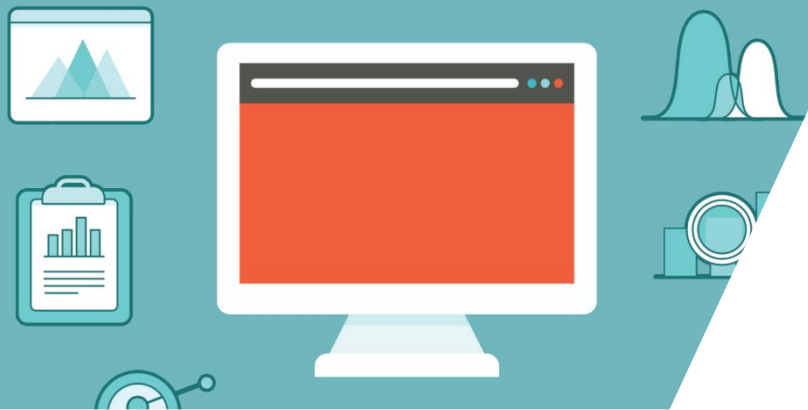 - Site-directed mutagenesis

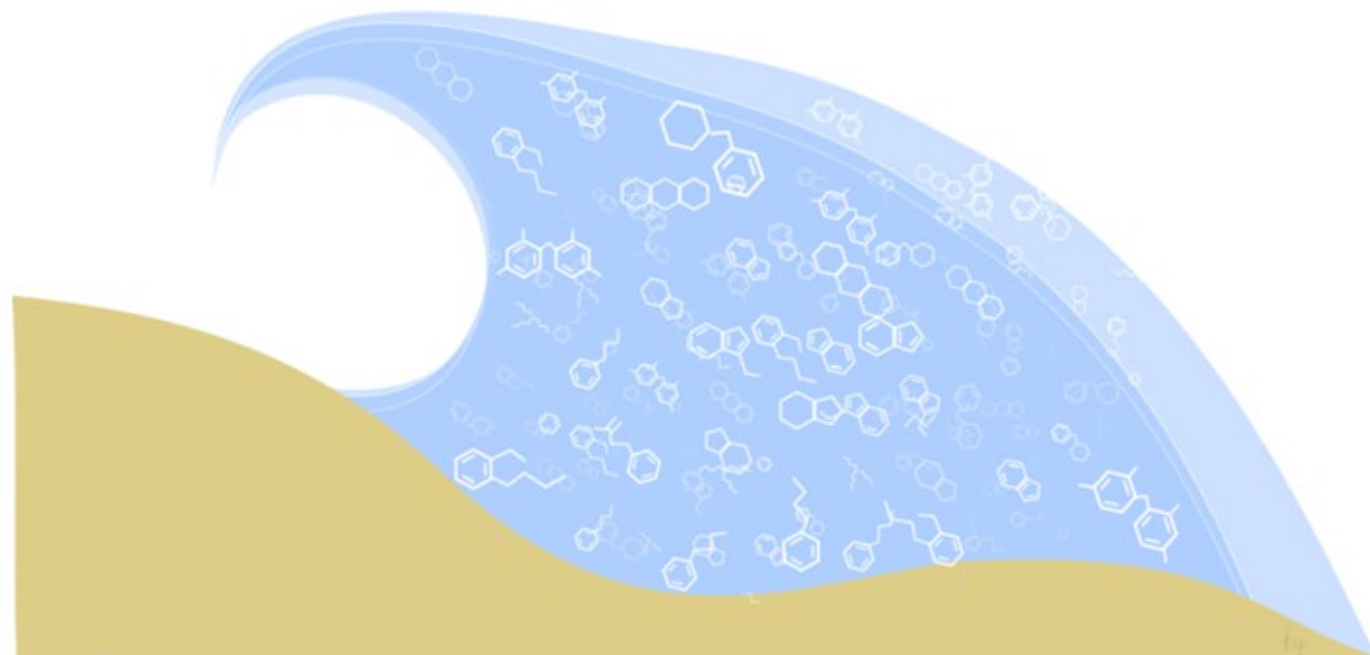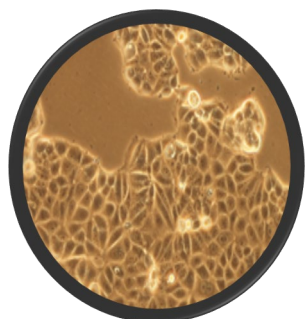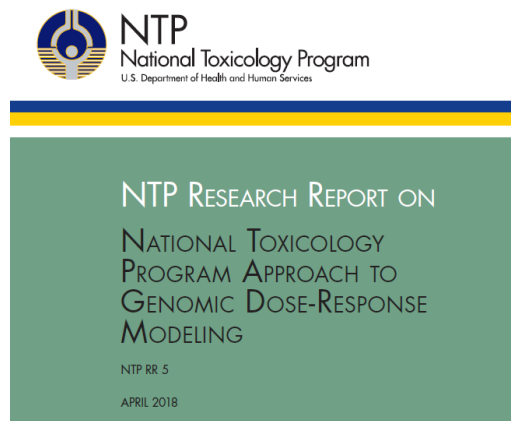**Problem Statement**:

Tens of thousands of chemicals are currently in use and hundreds more are introduced to the market every year. Only a small fraction has been thoroughly evaluated for potential risks to human health and the environment.

- Need for high-throughput assays to evaluate chemical safety.

- Provide adequate coverage of relevant pathways through which chemicals can cause toxicity.

Toxicogenomic Approach

Whole human transcriptome

NTP
National Toxicology Program
U.S. Department of Health and Human Services

NTP RESEARCH REPORT ON

NATIONAL TOXICOLOGY
PROGRAM APPROACH TO
GENOMIC DOSE-RESPONSE
MODELING

NTP RR 5

APRIL 2018

SOT | Society of Toxicology
OXFORD
academic.oup.com/toxsci

High-Throughput Transcriptomics Platform for Screening Environmental Chemicals

Joshua A. Harrill [*,1] Logan J. Everett,* Derik E. Haggard [*,†]
Thomas Sheffield,*,† Joseph L. Bundy,* Clinton M. Willis,*,‡
Russell S. Thomas [*] Imran Shah [*] and Richard S. Judson [*]

- Number of mammalian studies have shown short-term transcriptomics-based PODs are predictive of apical potency.
- Generally, within ½ log.
- Health protective points of departure.

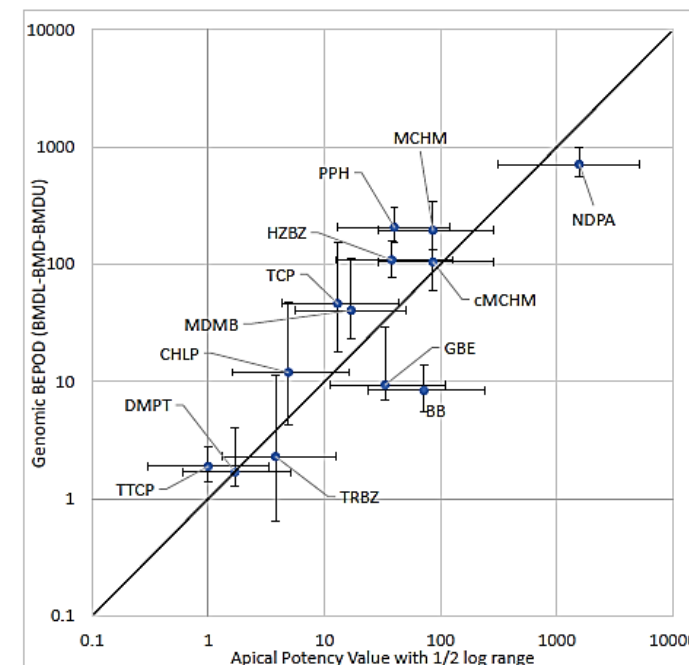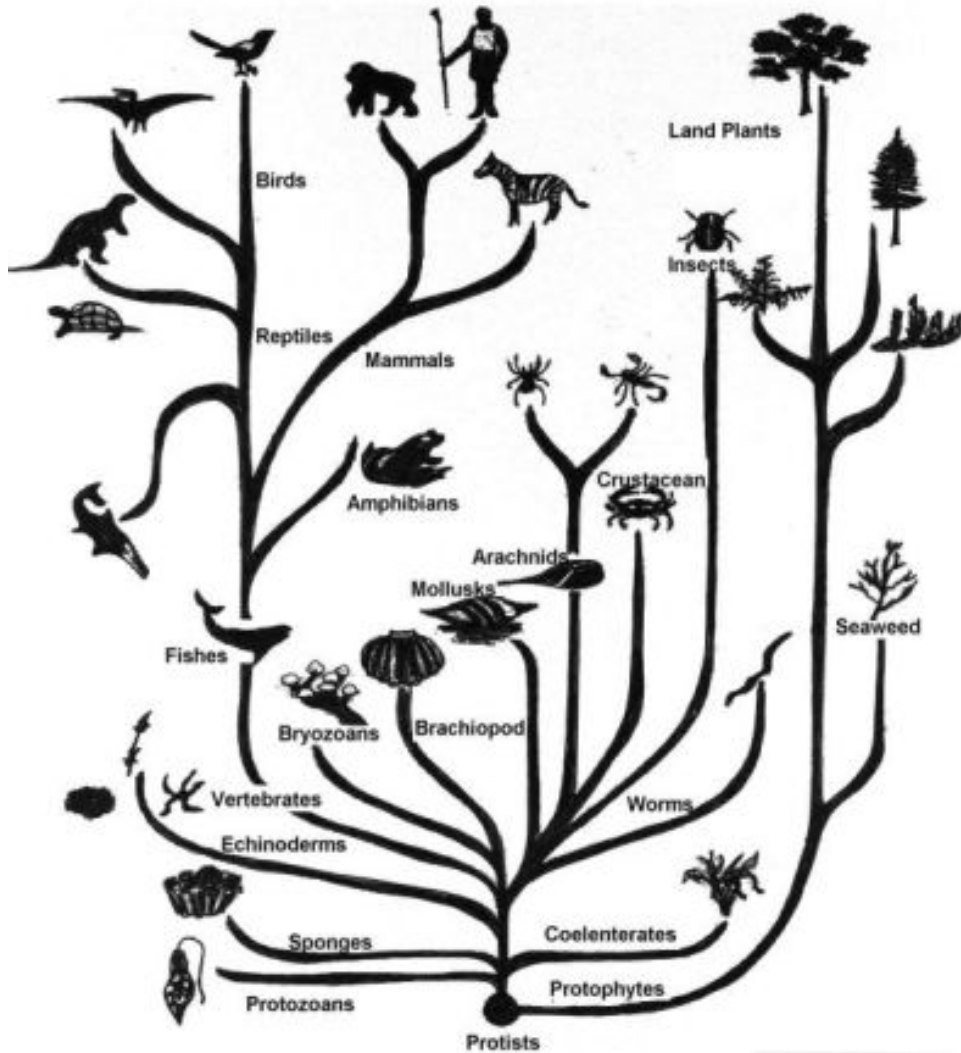
Figure 14. Comparison of the Most Sensitive Apical ½ Log Potency Range to the Most Sensitive GO Biological Processes BEPOD

Data from Figure 1–Figure 13 in this document were compiled to allow a larger scale comparison of apical and gene set-based biological potency estimates. The most sensitive apical potency values (NOAEL or BMD) from guideline toxicity assessments are plotted on the x-axis and the BEPOD range (BMD$_L$-BMD-BMD$_U$) from the GO Biological Processes analysis from 4- or 5-day GDRS studies are plotted on the y-axis. A diagonal 1-to-1 line is drawn as reference to perfect agreement between the potency values. The points to the left of the line demonstrate more sensitive apical endpoints, whereas those to the right exhibited more sensitive BEPODs. Overall, the apical and BEPOD values strongly agree, as indicated by $R^2 = 0.89$.

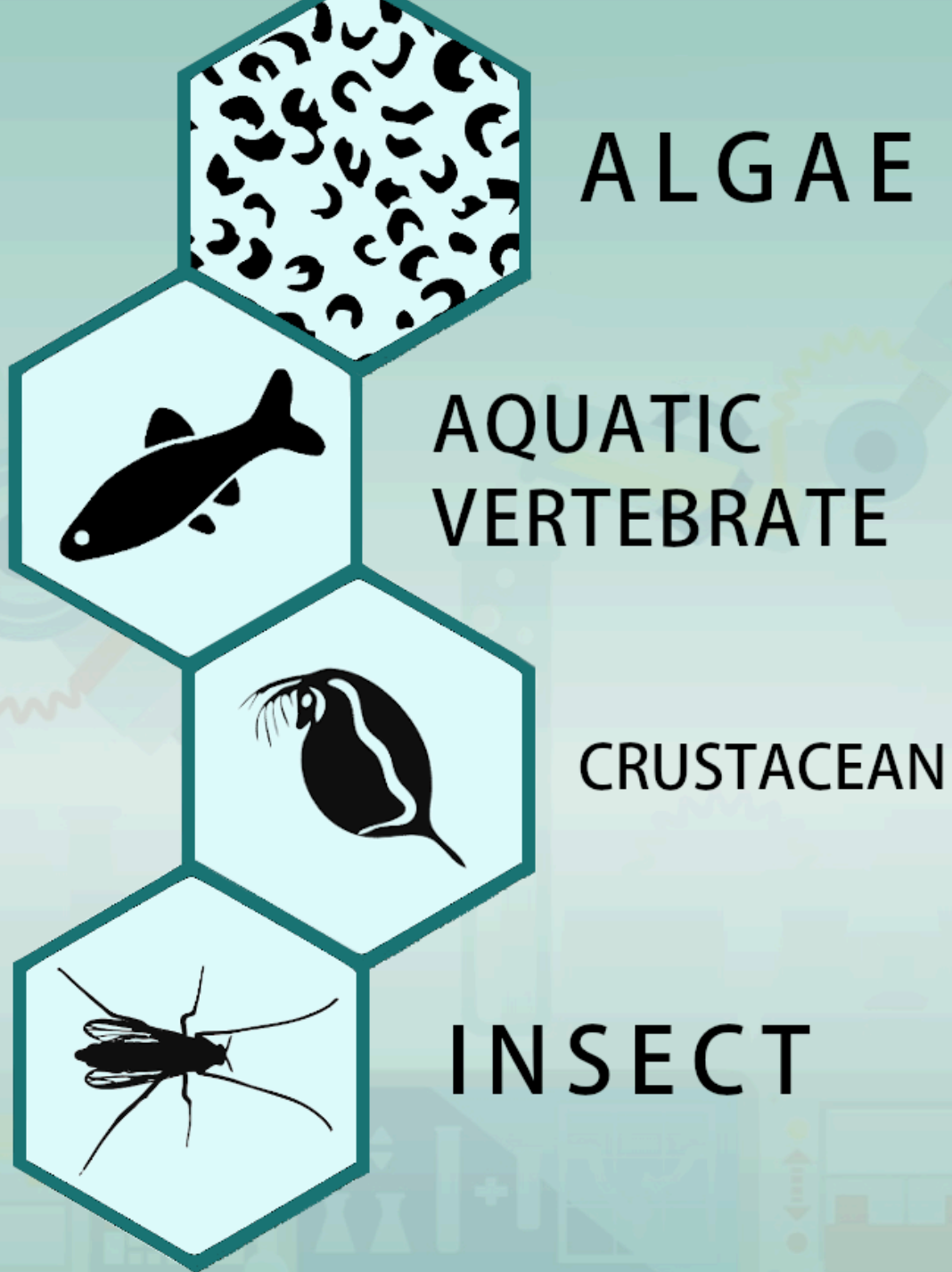# Ecotoxicology Perspective

- Humans are just a tiny fraction of the biological diversity we are charged to protect.

- Many genes/pathways are conserved

- Unique physiology in other kingdoms, phyla, classes…

- How do we assure those pathways are covered?
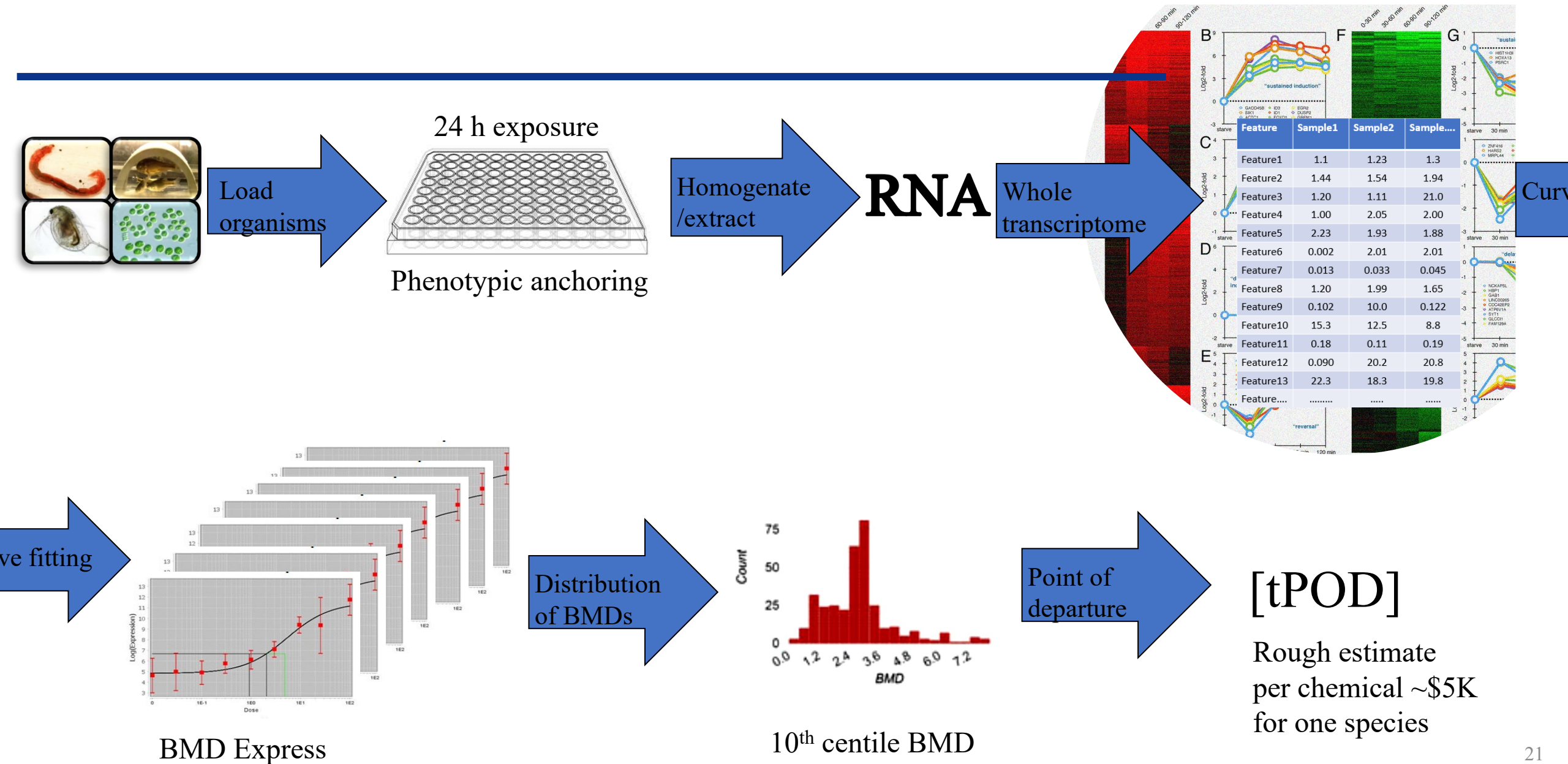
ALGAE

AQUATIC VERTEBRATE

CRUSTACEAN

INSECT

High throughput assays for three major trophic levels of aquatic ecosystems

- Primary producers (e.g., algae)
- Primary consumers (e.g., zooplankton, aquatic inverts)
- Secondary consumers (e.g., fish)

Commonly used for GHS classification and labeling of chemicals for environmental hazard

Aquatic organisms highly vulnerable to exposure

Incorporating transcriptomics as assessment endpoint

Load organisms

24 h exposure

Phenotypic anchoring

Homogenate /extract

RNA

Whole transcriptome

Curv

ve fitting

BMD Express

Distribution of BMDs

10th centile BMD

Point of departure

[tPOD]

Rough estimate per chemical ~$5K for one species

Evaluating the approach

[tPOD] $\leq$ [Most sensitive chronic endpoint] 👍

[tPOD] $>$ [Most sensitive chronic endpoint]
[tPOD] $<<<<$ [Most sensitive chronic endpoint] 👎

Approach is being explored with cell-lines for eco species as well

APCRA Case study: Transcriptomics-based PODs for Ecotoxicology

1. Generate transcriptomic PODs for ≈ 20 chemicals

   - Initial focus on fathead minnow
   - Parallel assays with additional taxa – for future analyses

2. Compare tPODs with available acute and chronic toxic toxicity data

3. Compare tPODs with in vitro-derived PODs

Acknowledgements

U.S. EPA, ORD
    Dan Villeneuve
    Kevin Flynn
    Adam Biales
    Donovan Blatz (ORISE)
    Sara Vliet (ORISE)
    Sally Mayasich (ORISE)
    Marissa Jensen (Univ. Minnesota Duluth)
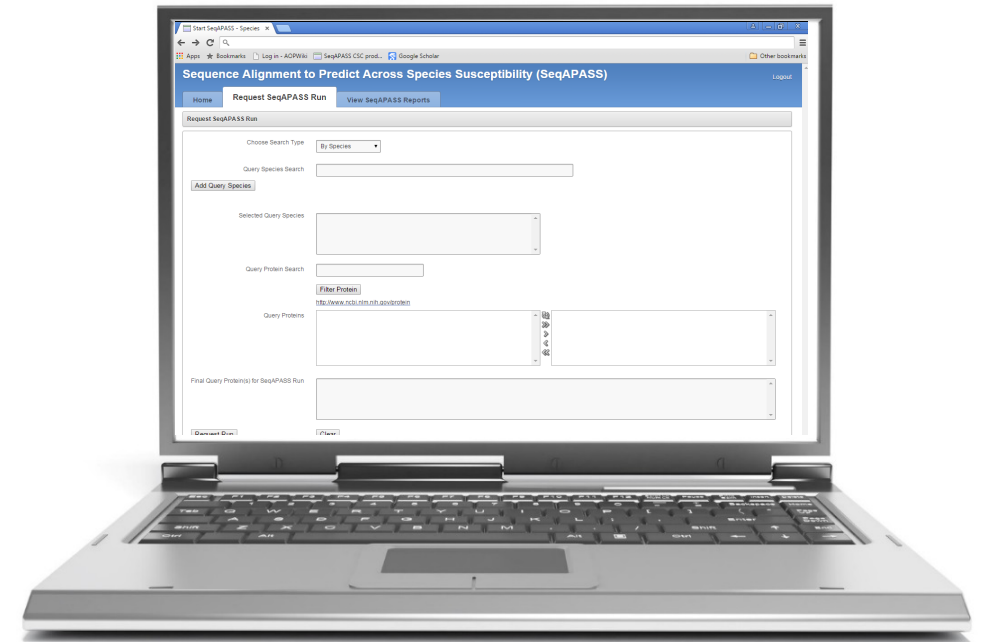
GDIT
    Thomas Transue
    Cody Simmons
    Audrey Wilkinson

University of Pittsburgh
    Carla Ng
    Weixiao Cheng

SeqAPASS v5.0

LaLone.Carlie@epa.gov

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