Challenges of utilising complex microbiome data in a regulatory context using plant protection products as an example

Christopher Sweeney
The assessment of microbial community composition within a future risk assessment is outlined in multiple EFSA documents.

Suggested microbial community (phylogenetic or functional) diversity could be a target for protection goals in a future risk assessment.

“Incorporating metagenomics into risk assessment in areas...such as: impact of pesticides on biodiversity (i.e. bacterial soil communities)…”
The risk assessment necessitates a single value representing the “hazard” or, ecotoxicological effect. How do you collapse the complexity of the entire soil microbiome into a simplistic metric suitable for use within a risk assessment?
Governmental or Institutional Concerns

● OECD, 2021:
  - “Number of challenges for using Omics data in a chemical risk assessment…Lack of transparency in how Omics data is generated…Lack of standardisation in…study parameters…Lack of case studies and guidance…”

● Ministry of Environment and Food of Denmark, 2016:
  - “Sharing, harmonizing and consolidating the available knowledge…to develop standardized procedures…”

● Danish Centre for Environment and Energy, 2019
  - “…current OECD... guideline programme does not yet include any 2nd generation ecotoxicogenomics tools…”
  - “Such tests depend on bioinformatics…tools that are not completely mature for routine applications…”

● Canadian Department of Fisheries and Oceans, 2021
  - “The lack of reporting standards for eDNA and inconsistency in reporting…creates challenges for end users…”
Governmental or Institutional Concerns

● OECD, 2021:
  - “Number of challenges for using Omics data in a chemical risk assessment…Lack of transparency in how Omics data is generated…Lack of standardisation in…study parameters…Lack of case studies and guidance…”

● Ministry of Environment and Food of Denmark, 2016:
  - “Sharing, harmonizing and consolidating the available knowledge…to develop standardized procedures…”

● Danish Centre for Environment and Energy, 2019
  - “…current OECD… guideline programme does not yet include any 2nd generation ecotoxicogenomics tools…”
  - “Such tests depend on bioinformatics…tools that are not completely mature for routine applications…”

● Canadian Department of Fisheries and Oceans, 2021
  - “The lack of reporting standards for eDNA and inconsistency in reporting…creates challenges for end users…”

There is a clear lack of understanding as to how to use high resolution technologies to study the microbiome within standardised regulatory contexts
Anatomy of a microbiome study

Existing guidelines to provide starting point

ISO 11063:2020
Standardised protocol for extraction of soil DNA

Earth Microbiome Project
Well established protocols for DNA amplification and preparation of samples for sequencing

Little understanding of how data can be processed and analysed in a standardised, robust and reliable manner suitable for regulatory decision making
Process to derive regulatory conclusions from microbial community data

Raw sequence reads → Sequence processing → Community composition

Complexities in variability associated with differential data processing methods

Per base sequence quality

Taxa A

Taxa B

Taxa...N

Response

Dose

Regulatory endpoint

Complexities in exploring the ecotoxicological response of 1000s of taxa

Complexities in amalgamating responses to derive ecotoxicological inference
Complexities of data processing

Pipeline selection

Sequence quality
Sequencing error

Quality criteria & filtering

Taxonomic assignment

Taxonomic Unit

Dose

1 2 3 4 5 ...

Bioinformatics matters: The accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline


Pipeline selection alters downstream community
e.g. ASVs vs OTUs

Ecological contextualisation
Taxonomic database selection
Complexities of data processing

Main point:

How robust are ecotoxicological inferences to variations in data processing pipelines and parameters?
How do we define a robust data processing pipeline that can take into account individual needs of a dataset?
Complexities of data analysis

Following the generation of a microbial community compositional dataset, this data needs to be curated before it is used further
Complexities of data analysis

Evaluate accuracy of data processing

Remove the influence of rare taxa and non-equal sampling effort

Field community

Inferred community

Sample A

Sample B

Number of ASVs

Sequences per sample
Complexities of data analysis

Main point:
How should we define rules to assess how well the variability in the initial community is captured in the final dataset?

What do we do if it is not well captured?

What steps do we need to take to ensure statistical power of our datasets e.g. filtering?

How do we define rules for the filtering of rare taxa?

Can we accommodate rarefaction in analyses and to what depth should we do this?

How do filtering and rarefaction alter ecotoxicological inferences?
Complexities of generating a regulatory endpoint

Each taxa will have its own response to the perturbation. How do we capture this community response in a single value?

Interpretation of endpoint

What is change?

How large is variability compared to normal operating range of the microbiome?

Relic DNA

Reference compounds?
Complexities of generating a regulatory endpoint

Main point:

What method should be used to calculate an endpoint from an ecotoxicological study?

How much variation is there in ecotoxicological inferences due to different methods to calculate endpoints?

How sensitive is the study system?

What is meaningful change?

Interpretation of endpoint

What is change?

How large is variability compared to normal operating range of the microbiome?

Relic DNA
Research needs and requirements

- Bioinformatics processing and data analysis represents an enormous increase in complexity compared to existing risk assessment methods.

- We do not understand how variability in methods can influence outcomes.

- If next generation sequencing is going to move from an academic exercise into the regulatory domain, clear guidance is needed.

- This community is well placed to help understand how we can better use high resolution techniques to explore the soil microbiome and its response following chemical exposure.