treecompareR: an open-source software package to visualize hierarchical chemical classifications

Paul Kruse[1,2], Caroline Ring[1]

1. Center for Computational Toxicology and Exposure, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA

Paul Kruse | kruse.paul@epa.gov | ORCID: 0000-0001-5516-9717

**Abstract number: 3675**
Poster number: P160

Introduction

- Characterizing the chemical universe is important for identifying and prioritizing data needs for exposure, toxicity, and risk assessment [1]
- ClassyFire provides a hierarchical chemical classification taxonomy scheme [2]

treecompareR: a new R package to visualize and explore chemical data using ClassyFire chemical classifications

Exploring the ChemOnt tree

Figure 1. The full ChemOnt taxonomy. The 'Organohalogen compounds' branch is highlighted for reference in Figure 2.
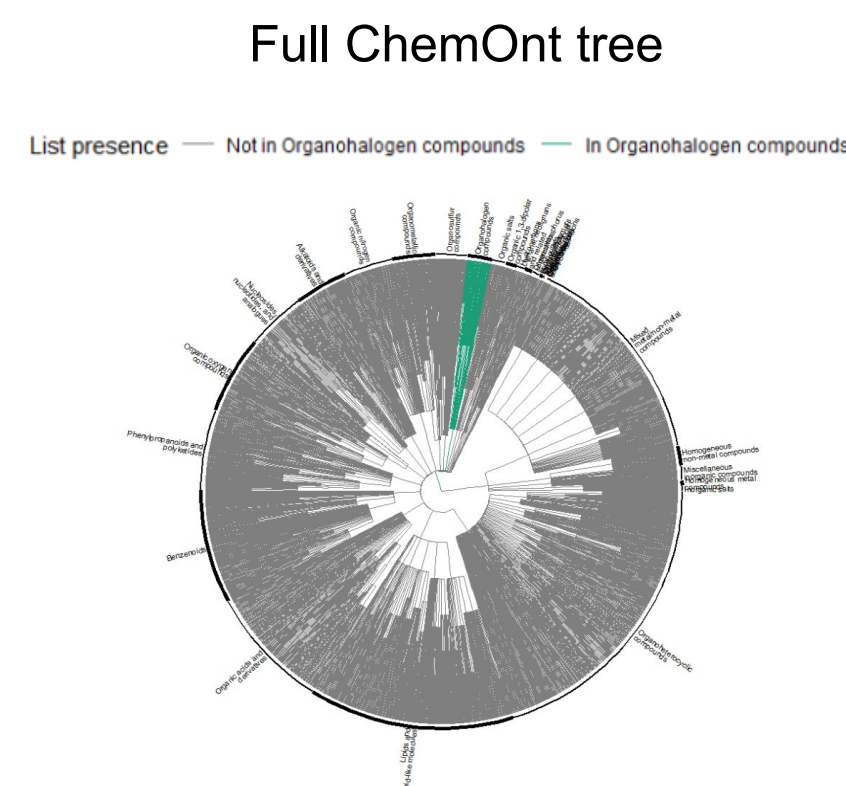
```
display_subtree()
```

- Draws tree
- Annotates clades (at user-specified level)
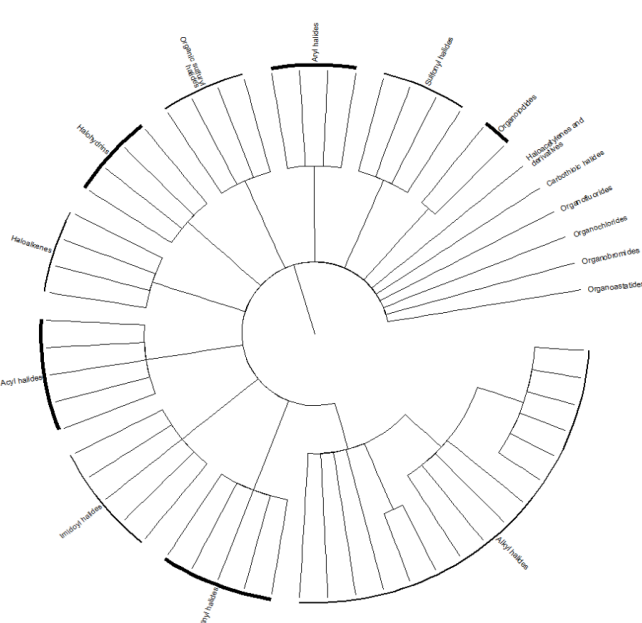 - e.g. superclass, class, subclass….

Figure 2. 'Organohalogen compounds' subtree (one superclass).

```
oh <-
prune_and_display_subtree(
prune_to = "Organohalogen
compounds")
```

- "Prune" tree to isolate a subtree
 - e.g., one superclass
- Automatic resizing

Full ChemOnt tree

List presence — Not in Organohalogen compounds — In Organohalogen compounds

Organohalogen subtree

Case study: Chemical contaminants monitored in biosolids and water

"Biosolids": Curated list of chemicals detected in biosolids [3]

"USGS": List of chemicals monitored in surface & ground water by US Geological Survey [4]

Classified using ClassyFire[5]

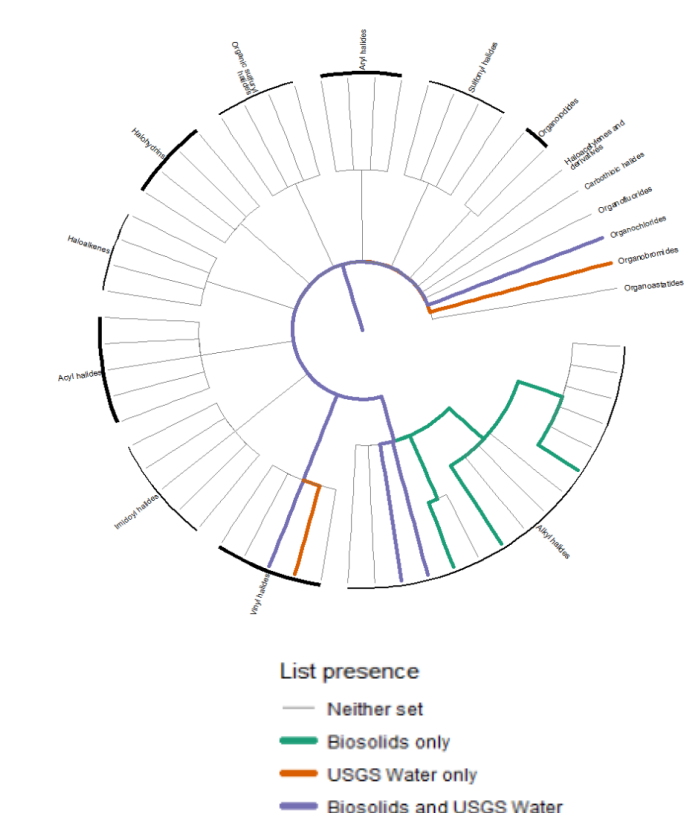**Figure 3.** `Organohalogen tree`, Biosolids and USGS highlighted separately.
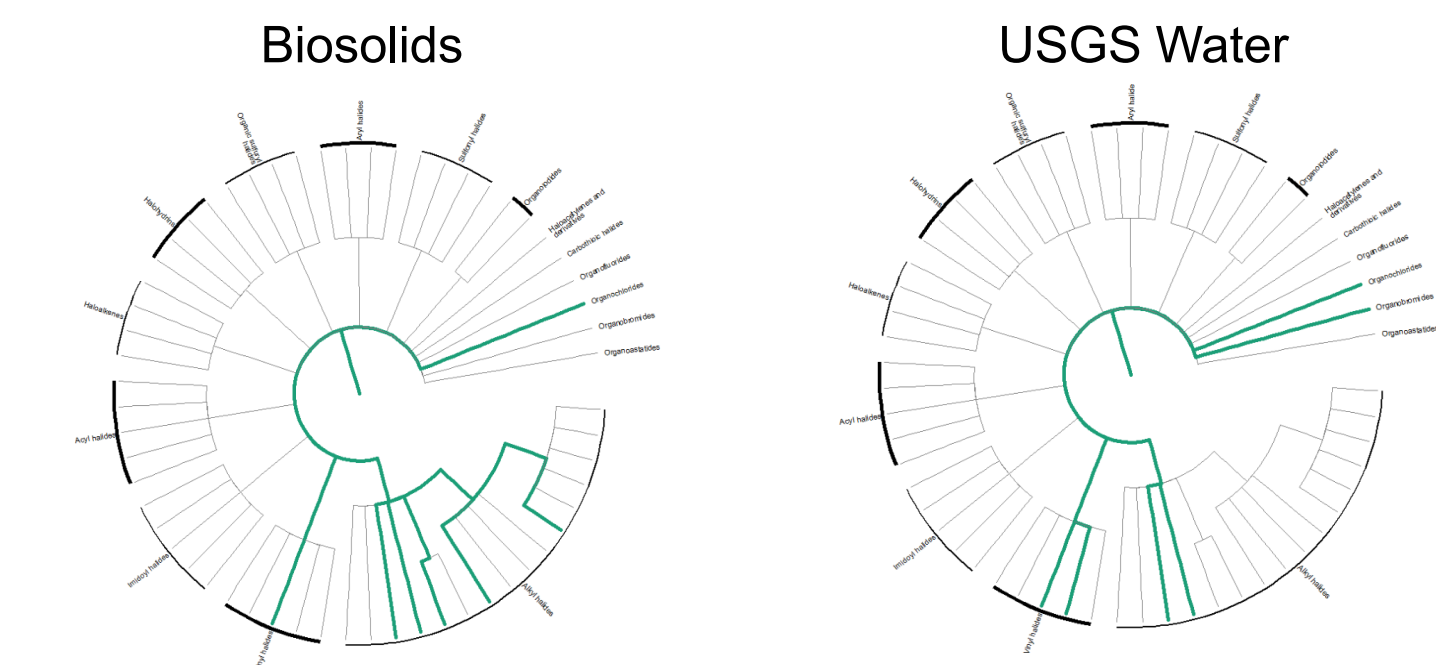
Biosolids


```
display_subtree(
base_tree = oh,
data_1 =
Biosolids)
```

USGS Water

List presence — Not in USGS Water — In USGS Water

```
display_subtree(
base_tree = oh,
data_1 = USGS)
```

List presence — Not in Biosolids — In Biosolids

Figure 4. `Organohalogen tree` and Biosolids and USGS highlighted together.

Biosolids & USGS Water


```
display_subtree(
base_tree = oh,
data_1 =
Biosolids,
data_2 = USGS)
```

List presence
— Neither set
— Biosolids only
— USGS Water only
— Biosolids and USGS Water

Biosolids & USGS Water
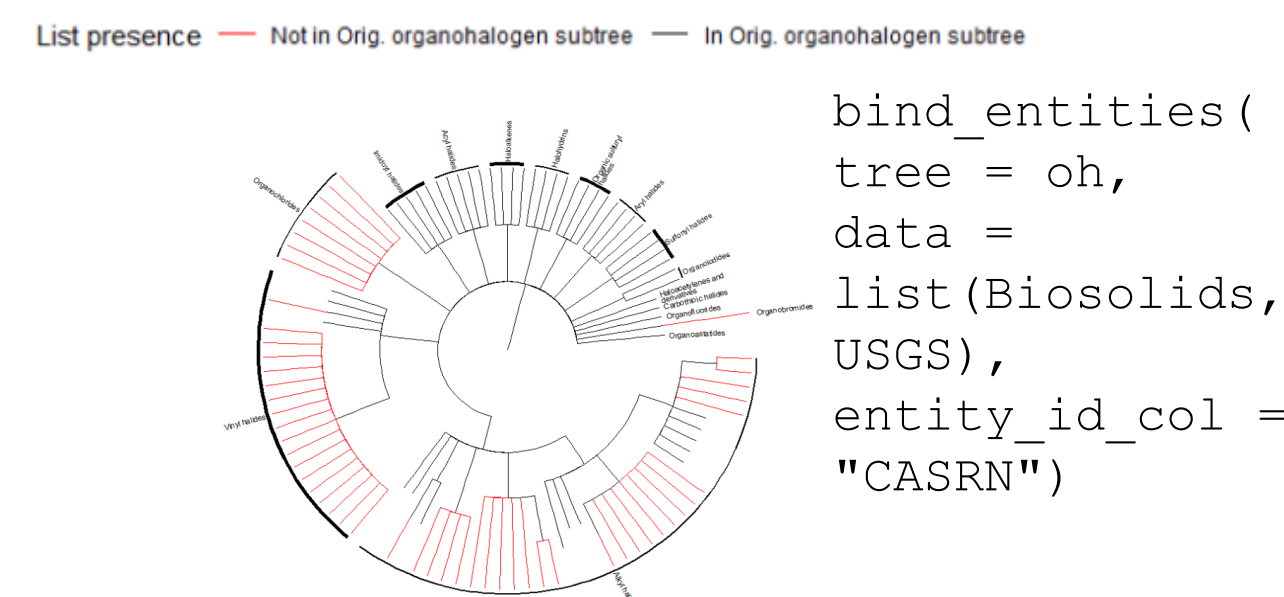
List presence
— Neither set
— Biosolids only
— USGS Water only
— Biosolids and USGS Water

Vinyl chlorides

Alkyl fluorides
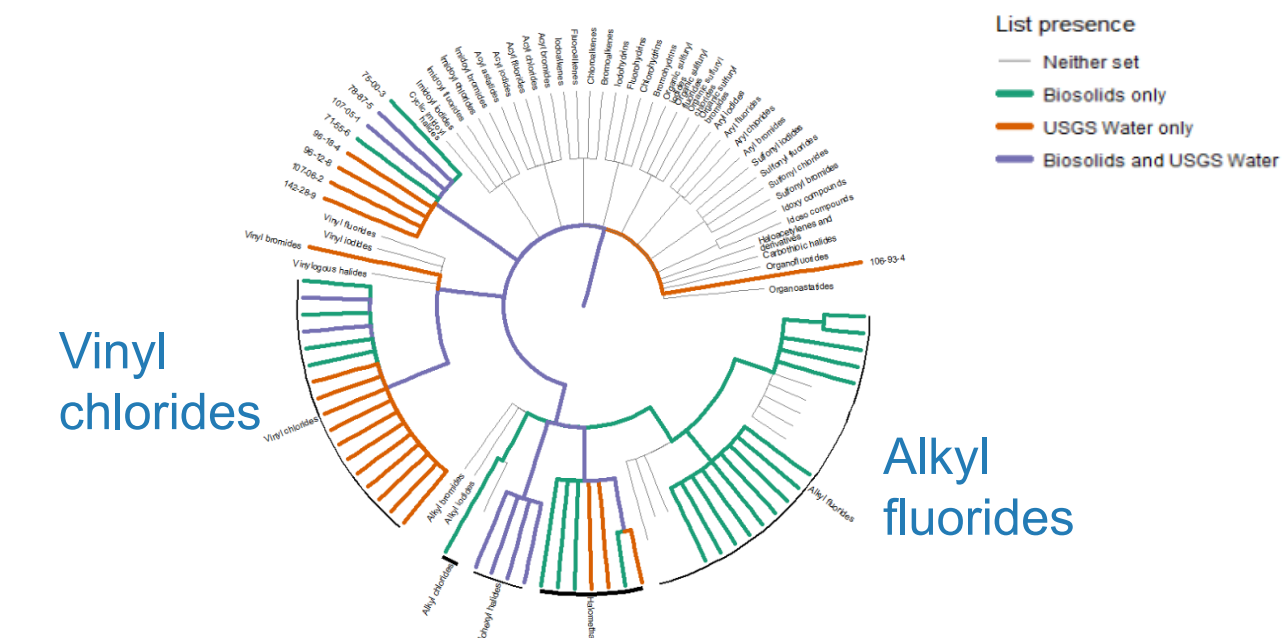
Figure 5. One way of visualizing individual chemicals & similarity. Top: Organohalogen tree. Bottom: Tree highlighted according to list membership (Biosolids, USGS, both, neither).

```
display_overlap(
base_tree = oh,
data_1 =
Biosolids,
data_2 = USGS)
```
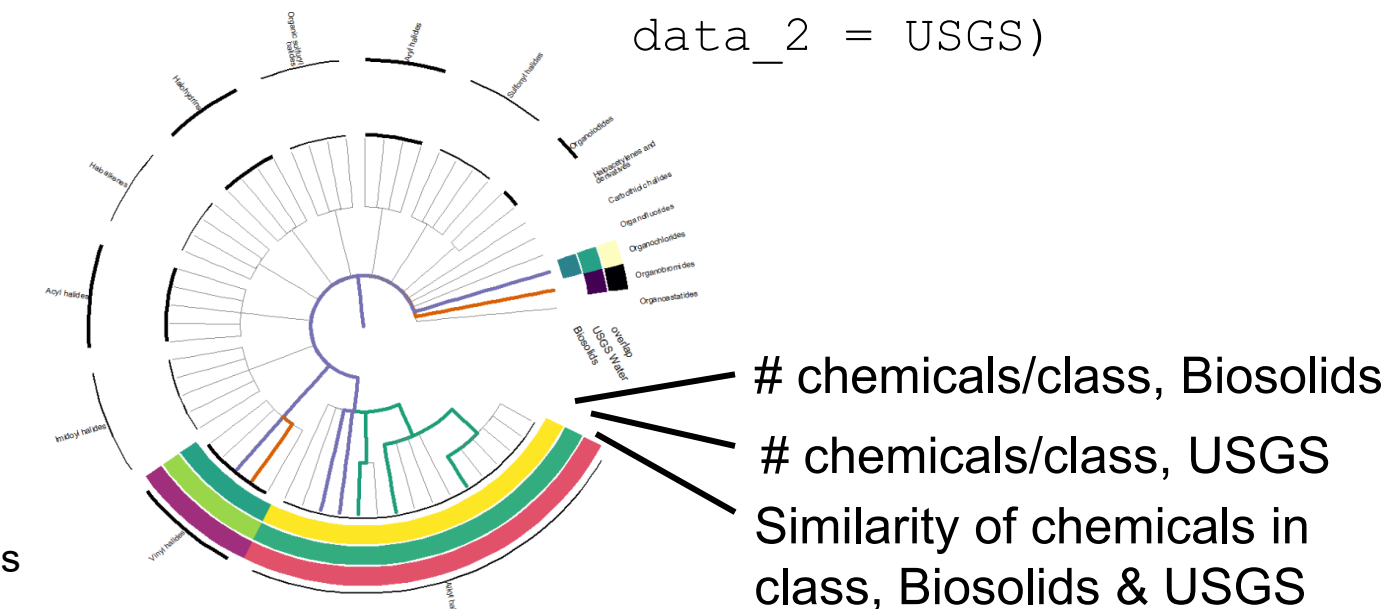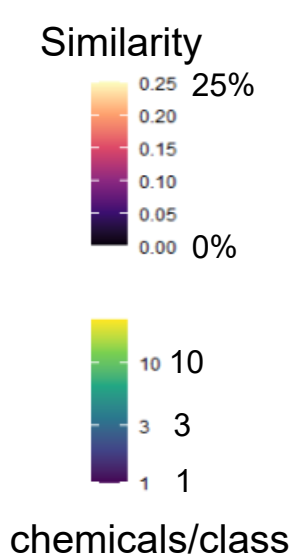
Similarity
0.25 25%
0.20
0.15
0.10
0.05
0.00 0%

10 10
3   3
1   1
chemicals/class

chemicals/class, Biosolids
chemicals/class, USGS
Similarity of chemicals in class, Biosolids & USGS

Figure 6. Another way of visualizing individual chemicals & similarity. Similarity = (# shared chemicals) / (# total chemicals)

Case study, visualizing chemical data: biosolids chemical concentrations

How do biosolids concentrations vary by chemical class?

Biosolids concentration data (mg/kg dry wt) as median from National Sewage Sludge Surveys (NSSS) 1988, 2001, and 2009 [6]

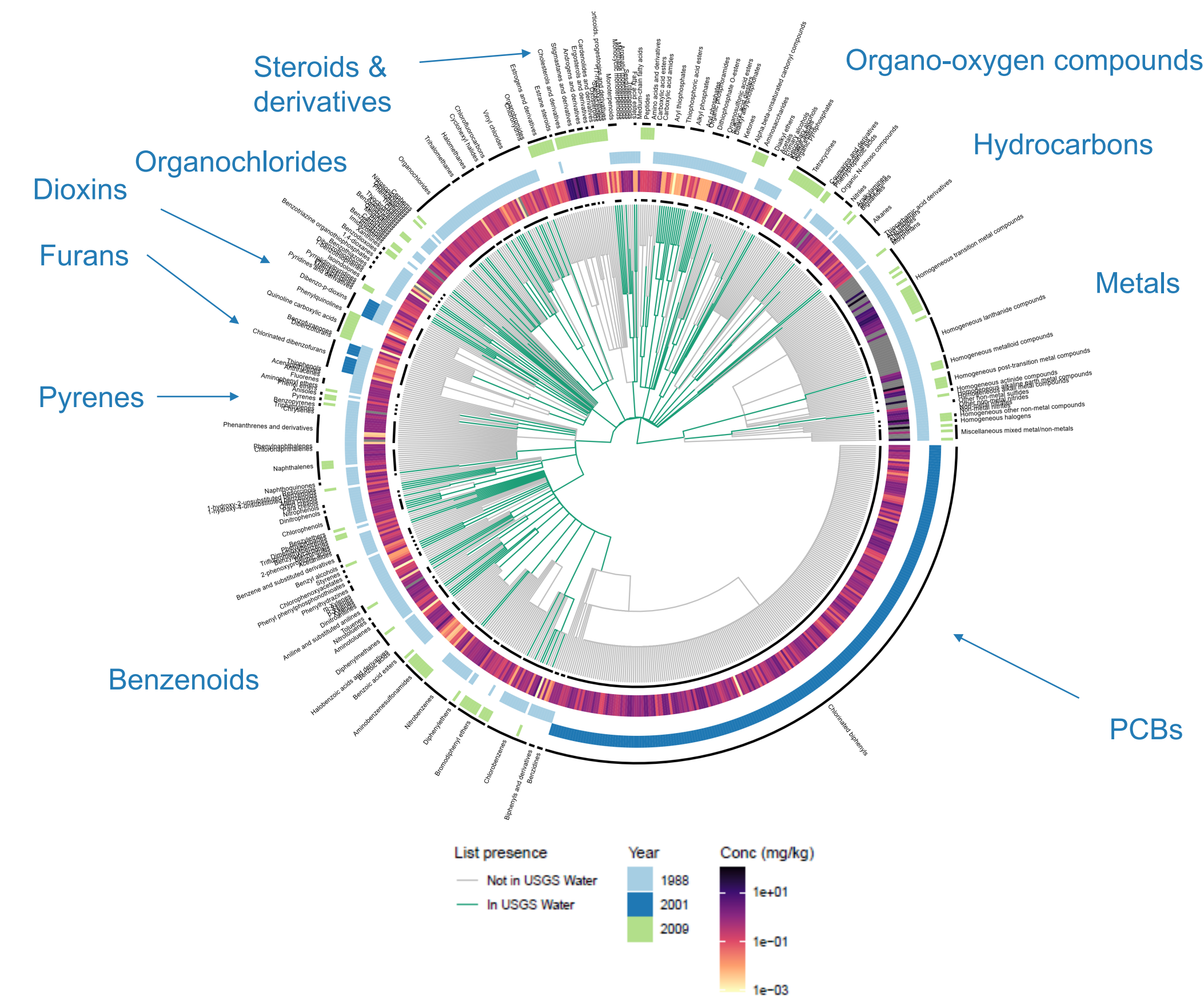Not all analytes were monitored in all years

Biosolids & USGS Water

Steroids & derivatives
Organo-oxygen compounds
Hydrocarbons
Organochlorides
Dioxins
Furans
Metals
Pyrenes
Benzenoids
PCBs

List presence
— Not in USGS Water
— In USGS Water

Year
1988
2001
2009

Conc (mg/kg)
1e+01
1e-01
1e-03

Figure 7. A tree diagram with individual chemicals displayed as tips. The inner most concentric circle displays mean chemical concentration while the three outer circles display membership within the three NSSS data sets. The branch shading expressed TSCA membership of the chemicals in the NSSS data.

```
display_subtree(base_tree = chemont_tree,
data_1 = NSSS) +
ggtree::geom_fruit(data = NSSS_concs, geom = "tile") +
ggtree::geom_fruit(data = NSSS_years, geom = "tile")
```

Conclusion

- Chemical data + ClassyFire trees = information-dense visualizations that can illustrate trends and connections that are otherwise obscured
- `treecompareR` package: easy tree visualizations of chemical data
  - built on `ggtree` and `ggplot2` [7-15]
- `treecompareR` can be extended to visualize other taxonomies (not just ClassyFire!)
- `treecompareR` also contains similarity measures for quantitative comparison of trees [16-20]
- For an additional case study use, see poster #P305

References