

Vladimir Mikryukov<sup>1</sup>, Kessy Abarenkov<sup>1,2</sup>, Shawn Laffan<sup>3</sup>, Tim Robertson<sup>4</sup>, Emily Jane McTavish<sup>5</sup>, John Waller<sup>4</sup>, Joseph T. Miller<sup>4</sup>

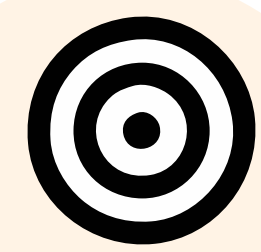
1, Institute of Ecology and Earth Sciences, University of Tartu, Estonia  
 2, Natural History Museum, University of Tartu, Estonia  
 3, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, Australia  
 4, Global Biodiversity Information Facility, Copenhagen, Denmark  
 5, School of Natural Sciences, University of California, Merced, USA

<https://github.com/vmik/PhyloNext>



## Introduction

Phylogenetic diversity indices evaluate biological communities from the perspective of the evolutionary interrelationship between species, and thus, help to maximize taxonomic, genetic, or functional diversity conservation effectiveness.



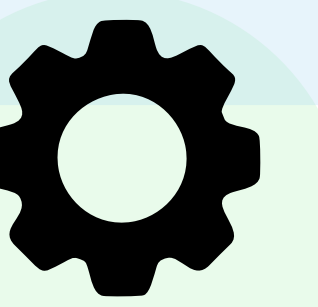
## Objectives

To bring together two critical research data infrastructures, the Global Biodiversity Information Facility (**GBIF**) and Open Tree of Life (**OToL**), to make them more accessible to non-experts, field ecologists, and policy makers.

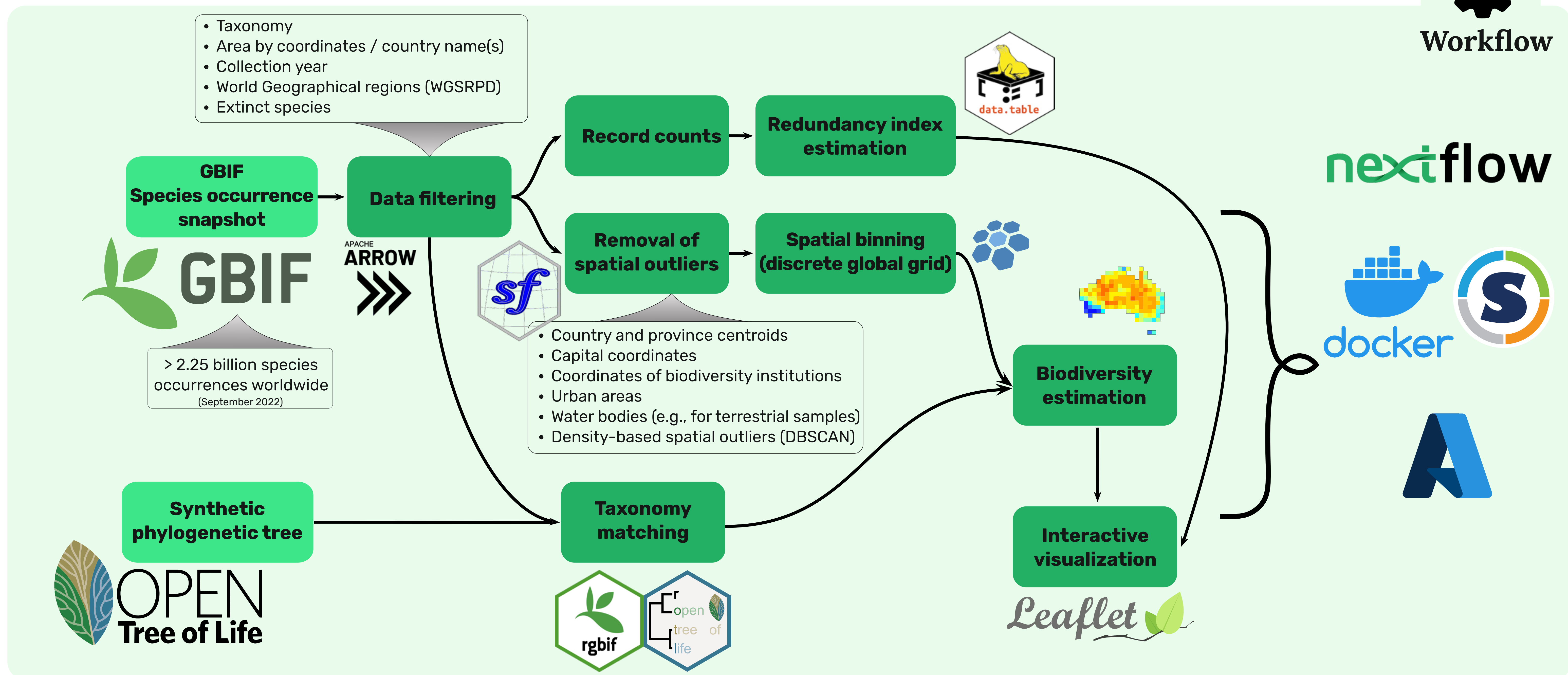


## Methods

- **Apache Arrow, data.table** - data filtering
- **OToL API** - phylogenetic tree preparation
- **rgbif, rotl** - species name matching
- **sf** - spatial data processing
- **Uber's H3** - spatial binning
- **Biodiverse** - diversity estimation
- **Leaflet** - interactive visualization
- **Nextflow** - workflow management
- **MS Azure** - cloud computing



## Workflow



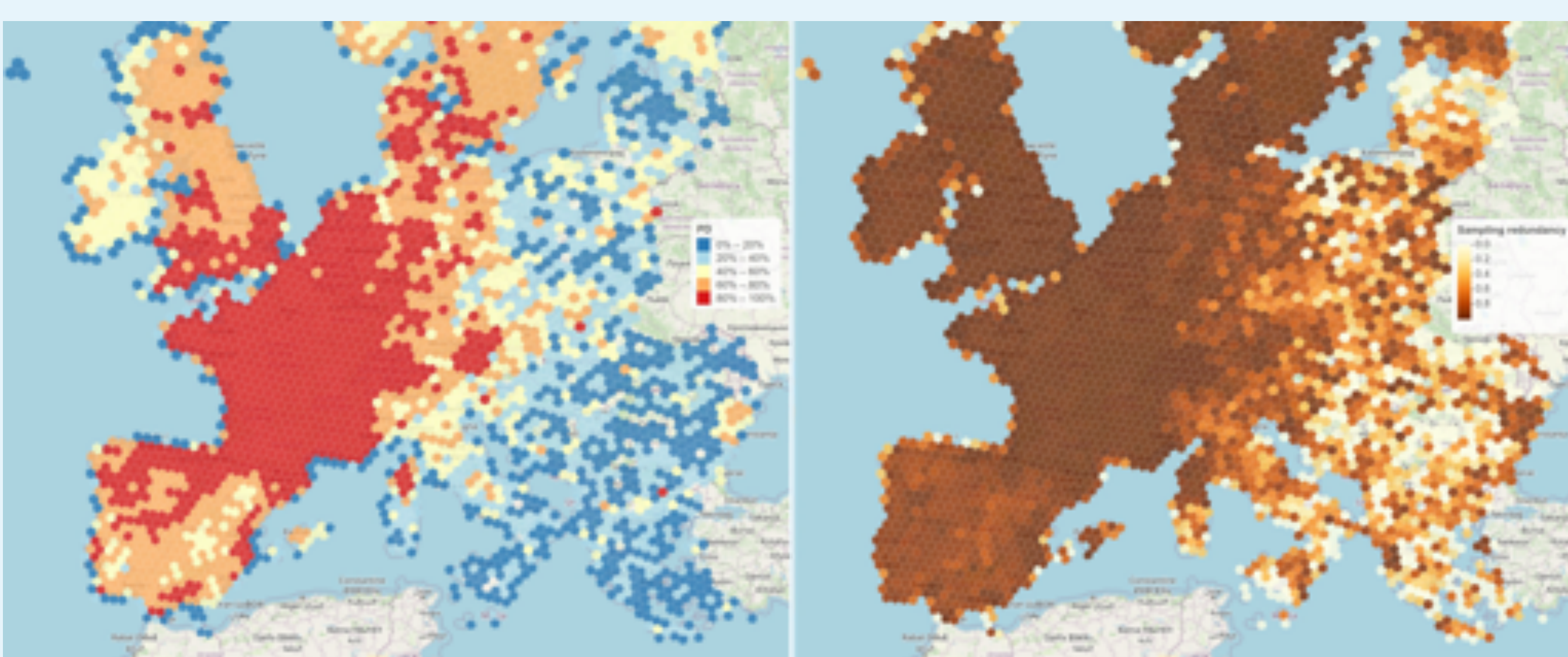
## Results

Data cleaning is an important step to exclude records with different geospatial issues

spatial outliers      species area

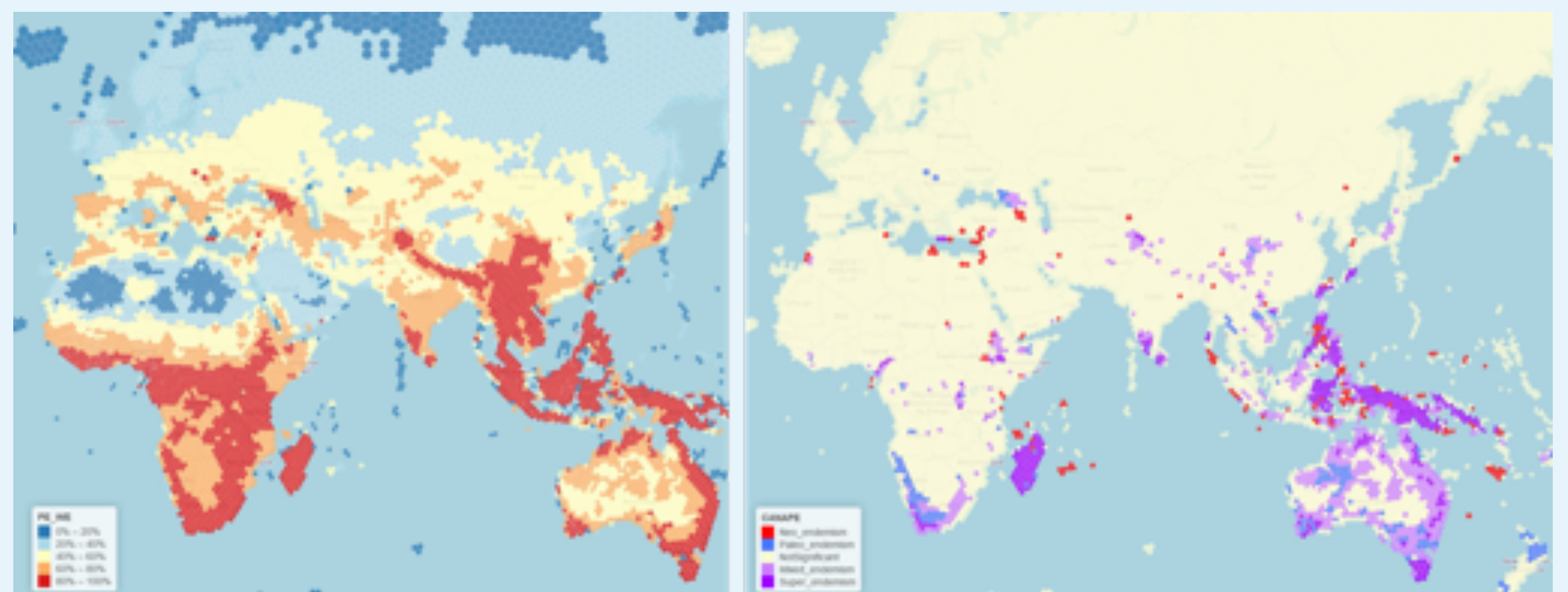


Phylogenetic diversity estimates depend on sampling effort, redundancy index could be used to mask undersampled grid cells



## Results

More than 350 metrics are available in **Biodiverse** (Laffan *et al.*, 2010) to capture multiple aspects of diversity.



Phylogenetic endemism and its types (paleo/neo)



## Acknowledgements

The work is supported by a grant "PD (Phylogenetic Diversity) in the Cloud" to GBIF Supplemental funds from the GEO-Microsoft Planetary Computer Programme. We thank Manuele Simi and nf-core community for helpful advices on the development of this pipeline.