

Finding the lost parents (Topic 1)

Aim/problem/goal

The idea for this topic resulted from research questions related to hybridization as a driver for plant speciation. For predicting the outcome (e.g. introgression, speciation, polyploidization) of plant hybridization it is important to know what the parental taxa are, and what their relatedness is. There is no single resource to discover what the parent species of a hybrid are. This is a particular problem in botanical research as a large proportion of plant taxa are of hybrid origin (Wissemann 2007). The goal was to compose an as long as possible, standardised checklist of hybrids and their parents taxa that can eventually be used for incorporation into other taxonomic resources, and to develop a workflow that automatically detects hybrids and their parents in publications.

Method

The parents of hybrids were parsed from literature, taxonomic checklists and Wikidata. This list was annotated with higher taxa hierarchies obtained from Global Biodiversity Information Facility (GBIF). Natural Language Processing tools - mostly rule based finite state automata - and named-entity recognition using gazetteer approaches applied to the text were used to deliver the annotations. A subset of articles from the Swiss Institute of Bioinformatics Literature Services (SIBiLS) were thus annotated.

Results

A checklist of 20,999 accepted hybrid names and a prototype of a tool for detecting hybrids and their associated parents from the literature.

Conclusion

Hybrids are ignored by Catalogue of Life and GBIF, yet during just one week of looking for hybrids and their parents, the team found a total of plant hybrids in the same order of magnitude as the number of species in the Asteraceae family (24,000), the largest family in the angiosperms. Knowledge about hybrids and their parentage is important to research in fields as diverse as evolutionary biology and the impacts of alien species. During the hackathon we have been linking literature and infrastructures such as GBIF to generate a list of hybrids and their parents (Fig. 4). By doing so, we came up with recommendations for extracting hybrid names from the literature for TreatmentBank (TB), improving the documentation of three Darwin Core terms, and amending the International Code of Nomenclature for algae, fungi and plants to standardise hybrid names.

References

- Wissemann, Volker. 2007. 'Plant Evolution by Means of Hybridization'. *Systematics and Biodiversity* 5 (3): 243–53. <https://doi.org/10.1017/S1477200007002381>.