

Report on the outcomes of a Virtual Mobility¹

Action number: CA18201

Grantee name: Adelaide Clemente

Virtual Mobility Details

Title: Construction of a mega database of European seed and gene banks - matching of plant names against the World Flora Online Taxonomic backbone data

Start and end date: 15/03/2022 to 15/06/2022

Description of the work carried out during the VM

One of the COST Action ConservePlants aims is to make a database on European seed bank accessions to evaluate the current status of *ex situ* conservation in Europe and to define a strategy for biodiversity conservation post-2020. Data from 113 institutions (>150,000 accessions) has already been reviewed and merged. Since taxonomic references differed among countries, the harmonisation of the taxonomy in this European database is crucial for further data analyses within this project. An automate match of the data against the World Flora Online Taxonomic backbone data, EuroMed+ and Plants of the World (from now on referred as WFO+EMed+POWO) has already been performed. The following tasks were then planned for the work:

- i) for unmatched *taxon* names identified by the automated procedure, identify and correct errors or misspellings;
- ii) for the remaining unresolved names, look up other references, consult experts of the local flora or the scientific literature in order to make a decision on the acceptance of specific *taxa*;
- iii) deliver a dataset ready to be used in further analyses.

A total of 1657 accessions with unmatched plant names were first classified into broad categories (genus name only, hybrids, doubtful names, and full *taxa* names). Full *taxa* names (1018 accessions) were then inspected, searched for and matched individually against WFO+EMed+POWO. Autonyms, which do not match against WFO+EMed+POWO using automate procedures, were categorized and names containing errors and/or misspelling were edited. Each edited name and autonym were searched again in WFO+EMed+POWO. The remaining unmatched names were individually searched for in on-line regional Floras (e.g. Flora Iberica, Tela Botanica, Panarctic Flora, Portale della flora d'Italia) and other Plant databases (e.g. TROPICOS, Catalogue of Life). When a match was found, the corresponding synonyms were searched for in WFO+EMed+POWO. Maintaining original full name or synonym was a

¹ This report is submitted by the grantee to the Action MC for approval and for claiming payment of the awarded grant. The Grant Awarding Coordinator coordinates the evaluation of this report on behalf of the Action MC and instructs the GH for payment of the Grant.

priority. When the synonymy was doubtful, the taxonomy was complex or names were unpublished (ined.), additional information was retrieved from Scopus or other scientific databases. Synonyms, reference sources and the fundamentals for selecting the accepted name were annotated in the file, together with relevant links to webpages.

Major difficulties, which precluded a confident decision on the accepted name, were found in unpublished names, *pro part* synonyms, full name or infraspecific *taxa* not found, unresolved and invalid names, several authors for a species name, and conflicting accepted names among the three sources. The decision criteria to be applied in such cases was discussed in meetings with WG2 co-leaders (Andreas Ensslin and Sandrine Godefroid) and, for the remaining unresolved names (283 accessions), 40 data contributors were enquired by e-mail to provide a synonym accepted in WFO+EMed+POWO or a reference for the original name.

Two additional datasets were imported into the European seed and gene banks database. One dataset was matched against WFO using automate procedures (88 accessions) and the other dataset was formatted for import (2079 accessions). The procedures described above were applied to unmatched names in both datasets.

Description of the VM main achievements and planned follow-up activities

The taxonomic homogenization of the dataset was achieved by providing a name accepted in at least one of the three taxonomic references (WFO+EMed+POWO) or an alternative reliable reference source for the original unmatched name;

The full dataset now made available is the basis for the following tasks of WG2 task 2.2 (Analysis of *ex situ* programmes throughout Europe), more specifically the group's objective 2: "Review seed bank databases about the quality of their collections and the utilization of the material for conservation measures". This is a contribution to one of the research coordination objectives stated in the MoU (Identify knowledge gaps in plant conservation and the sustainable use and promote best practices) and to deliverables 12 and 15;

Forty institutions participating in the Action as data providers have been contacted by e-mail, the great majority replying promptly and providing a solution for the name of over 200 accessions;

Data contributors corrected typos and misspelling or revised *taxa* names upon our queries, thereby improving the quality of the information in their own databases;

Prioritizing the acceptance of original names enabled the maintenance of many Red List *taxa* or *taxa* with local or regional conservation status in the European seed and gene banks database;

The progress of the work and criteria for accepting names were discussed in regular meetings with the WG2 co-leaders, and the fruitful discussions and exchange of information enhanced the progress of the work;

Given my knowledge on European seed and gene banks database characteristics, I will collaborate in future activities of WG2 task 2.2 (Analysis of *ex situ* programmes throughout Europe).