The genomic uses of a 200 year-old herbarium – Pitfalls and potentials

Camille Christe[1], Carlos G Boluda[1], Aina Randriarisoa[1], Tina Kiedaisch[1], Zeynep Toprak[1,2], Fred Stauffer[1], Yamama Naciri[1] and Mathieu Perret[1]

Major herbaria, such as the one hosted by the botanical garden of Geneva (G) have played a central role in the development of plant systematics over the last 200 years. Today, advances in high throughput sequencing technologies (HTS) together with the development of targeted capture, where DNA extracts are enriched for preselected loci using hybridization probes prior to sequencing, have considerably improved the use of herbaria as a source of genetic data, opening new avenues in the study of plant biodiversity.

Keywords
Hyphaene, *Sapotaceae*, *Silene*, Targeted capture

Addresses of the authors
[1] Conservatoire et Jardin botaniques de Genève & Department of Plant Sciences, University of Geneva, Chemin de l'Impératrice 1, Chambésy, 1292 Geneva / Switzerland
[2] University of Dicle, Faculty of Sciences, Department of Molecular Biology and Genetics, Diyarbakır / Turkey

Contact
camille.christe@ville-ge.ch

Accepted
15. 1. 2023

DOI
https://doi.org/10.12685/bauhinia.1360

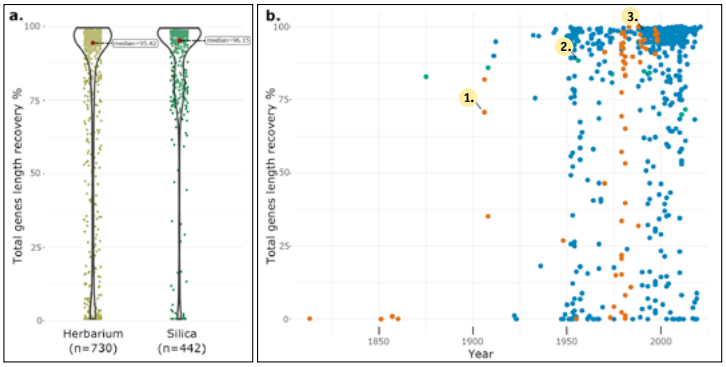

Fig. 1 a. Total genes length recovery for samples extracted from herbarium sheet and silica dry leaves. **Fig. 1 b.** Total genes length recovery versus time of collection for samples extracted from herbarium sheets only. Blue points correspond to *Sapotaceae* samples, orange to *Silene* samples and green to *Arecaceae* ones. Numbers correspond to example herbarium sheets: **1.** *Silene lagunensis* (Fig. 2), **2.** *Hyphaene* sp. (Fig. 3) and **3.** *Capurodendron nanophyllum* (Fig. 4).

Since 2016, research conducted at Conservatory and Botanical Garden of Geneva using HTS approaches on herbarium specimens were mainly focused on three taxonomical groups. First, the genus *Silene* in the Caryophylaceae family was investigated with the aim of defining the relationship and species boundaries in the section Italicae. Specimens were mainly from the Mediterranean region, with a total of 133 samples, with 56 % of herbarium origin (oldest 1813, mean 1970, SD 37.4 years). The kit used targeted 256 regions for a total of 650 000 bp. Secondly, the Sapotaceae family was investigated with the aim of refining generic and species circumscriptions in the family. Specimens were collected from tropical regions, mostly from Madagascar. They are the results of years of collections that evidence the

Fig. 2. *Silene lagunensis*, collected in 1906 in the Canary Islands (recovery 70.7 %; labelled as 1 in Fig. 1b).

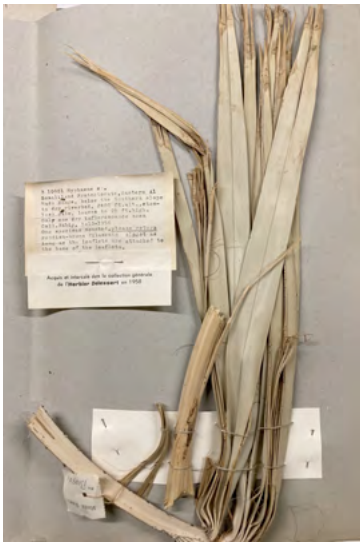

Fig. 3. *Hyphaene* sp. collected in 1956 in Somalia, a region difficult to explore today but key to understand the diversification of this genus (recovery 88.4 %; labelled as 2 in Fig. 1b).

Fig. 4. *Capurodendron nanophyllum*, the species with the smallest leaves in the Sapotaceae. Described in 2018 as already critically endangered due to deforestation (recovery 99.4 %; labelled as 3 in Fig. 1b)

dramatic loss of biodiversity due to deforestation. From the 995 samples, 70 % were extracted from herbarium sheets (oldest 1911, mean 1993, SD 21.4). The kit used was specially design for Sapotaceae (Christe et al. 2021) and targeted 792 regions for a total of >870 000 bp. Lastly, the genus *Hyphaene* in the Arecaceae family was investigated in order to define its position within the sub-family Coryphoideae and to address species delimitation issues within the genus. Only nine herbarium specimens out of 124 samples were used (oldest 1875, mean 1959, SD 50.5) but they represent valuable samples. The kit used targeted 916 regions for a total of >1 500 000 bp (Loiseau et al. 2019).

Results deriving from these studies helped us gaining experience with the use of herbarium specimens for HTS, and their chance of success in terms of maximum age and sequence recovery. We found that the total maximum base pairs (bp) recovery for herbarium specimens was on average very good and highly similar compared to silica gel preserved samples (median 95.42 and 96.15 bp, respectively; Fig. 1). The oldest sample with a good recovery rate, i.e. the percentage of the maximum genes length recovery for each taxonomical group, (83 %) is almost 150 years old. Unfortunately, we could not achieve such rates for older samples. Globally, the recovery rate is correlated with the collection year (Spearman's $\rho = 0.37$, $p = 4.66e^{-25}$), meaning that the older the sample is the less chance we have to obtain a good recovery rate.

We conclude, based on our experience with three taxonomic groups spanning different climates and collection times, that fragmented DNA does no more represent an absolute limit in using herbarium material. We are now ready to genetically explore the herbarium at a higher scale than before, with some prospective applications such as the discovery of undescribed diversity, or the monitoring of regional flora. However, the relatively high quantity of DNA that is needed for HTS, the destructive nature of the sampling, and the reduced chance of success with specimens older than 50 years, request a wise selection of the samples to be sequenced. Despite these pitfalls, we foresee that the use of herbaria as a source of genetic data will play a central role for the completion of the plant tree of life and the accurate quantification of plant diversity at all taxonomical levels.

References

Christe C, Boluda CG, Koubínová D, Gautier L & Naciri Y (2021) New genetic markers for Sapotaceae phylogenomics: More than 600 nuclear genes applicable from family to population levels. Mol Phylogenet Evol 160: 107123.

Loiseau O, Olivares I, Paris M, et al (2019) Targeted Capture of Hundreds of Nuclear Genes Unravels Phylogenetic Relationships of the Diverse Neotropical Palm Tribe Geonomateae. Front Plant Sci 10: 864