Manuscript Restructuring Notes

People

RCB, Brianna Wieferich, Chris Woolridge?, Jeremy Foster?, Jeremie Fant?

Overview

Consider both eSTZwriter (EW) and safeHavens (sfH) as a continuum. EW for non-empirically developed strategies for bio-banking. sfH as a software used for developing empirical seed zones. POSSIBLY new functionality for EW to select eSTZs for collection prioritization.

Both tools are used to: - Prioritize the collection of a minimal number of collections to meet conservation and/or restoration targets. - sfH focuses on neutral genetic variation - Purely computational approach, using pre-existing data sets. - EW on adaptive genetic variation - Common garden based approach to infer adaptive patterns for restoration across time points

Software todo - safeHavens

- Use data product for environmental variables which includes predicted timepoints, e.g. 2050 under SSP X.
- Ensure we are using spatial cross validation for the glmnet folds

Possible testing data for manuscript

- Comparison of how existing biobank collections cover the scenarios generated by sfH.

  - i.e. fraction of all identified collection areas which are covered under realized collection goals.

- Comparison of how each of the sfH methods would capture adaptive variation as found from all existing eSTZ data products. ~30 taxa

- i.e. fraction of all identified stzs which would be 'collected' in under a total target number of STZs

- IS THERE A WAY TO COMBINE POP GEN NEUTRAL DATA SETS TO SEE HOW MUCH ALLELIC VARIATION MAY BE COVERED ACROSS A SPECIES RANGE??

Introduction Structure

1) General Background
2) Most of sampling well determined from synthesized population genetics data and forestry
3) space for improvements remain in most sectors as higlighted with the number of individuals, populations etc.
4) spatial tools can help provide a measurable framework for achieving collection goals.
5) Here we present such tools.

The Earth is experiencing a mass extinction event with estimates suggesting 45% of flowering plant species are at risk of extinction (@bachman2024extinct, @isbell2023expert). Two major recent responses to mitigate the loss of taxonomic and genetic diversity include the preservation of plant genetic materials (hereafter germplasm) in *ex situ* collections, the restoration of natural areas to enhance the preservation of germplasm *in situ*, and combinations thereof (@heywood2017plant, @volis2010quasi, @westwood2021botanic). Long-term *ex situ* conservation is underway for species that have had, or are expected to experience, the entire loss of their habitats, or a deterioration in the ecological integrity of their habitats beneath thresholds that can support their populations continued persistence (@mounce2017ex, @union2012integrated, @TADDEO). Recent advances in ecological restoration allow for the enhancement of degraded natural habitats, supporting the persistence of both relatively abundant species and species threatened with extinction; however, the development of native plant propagules for restoration requires the collection, and on farm amplification of genetically representative material (SHRIVER et al. 2026). Hence, both major responses require effective sampling of germplasm across geographic and environmental conditions (@wambugu2023role).

Central to attaining germplasm collection goals is the collection of adequate amounts of genetic diversity for the conservation of a species for *ex situ* or of populations or groups of populations for restoration (@walters2021unique, @SHRIVER). Approaches for collecting germplasm from wild species for *ex situ* conservation and *in situ* restoration projects in use by the nearly 1750 seed banks globally often follows a set of minimal rule-of-thumb guidelines developed from agricultural and population genetics studies (@hay2013advances, @westwood2021botanic). Currently, guidance exists for the number of propagules that can be collected to safeguard the persistence of natural populations, the minimum number of individuals that need to be sampled per population, the number of populations to sample per species to reach genetic diversity targets, and guidance on the geographic structure of where to move restoration collections to and from to avoid using ma-adapted phenotypes at the restoration site (@CITE, @CITE, @CITE).

However, a variety of goals exist at the institutional, programmatic, and bio-regional organizational levels, as well as opinions among curators, which have allowed for a natural evolution of collection goals above minimal rule-of-thumb guidelines (@guerrant2014sampling). Further, these minimal guidelines do not guarantee alignment or achievement of goals for all observed species and population genetic structures, or population vital rates, because of nuances in the biology and ecology of individual species (@bucharova2025assessing, @hofner2025spatial).

… However tooling and integration required to prioritize targets, reduce the possibility of duplicated efforts, and use funding for field operations most effectively (… … A variety of systems for collection of materials have been proposed formally in the literature, or are in use (@havens2015seed, @hoban2020taxonomic,) … Tools for sampling can help design collection goals and priorities (@wambugu2023role)… Here two tools are detailed, one for biobank curators to help maximize their coverage of a species allelic diversity `safeHavens` and a second to assist quantitative geneticists to produce easily usable empirical seed transfer zone products. In this manuscript we provide a quick overview of both software packages, describing core functionality.

We further make comparisons to show when / where the methods for general collections vary, by 1) Describing the structure of XXX biorepository collections relative to the sampling scenarios. 2) For both common and rare taxa determining how well neutral genetic variation would represented under the various sampling criteria. 3) For common taxa, comparing how well each of various sampling schema would cover the identified adaptive variation documented via empirical seed transfer zones.