

HerbVar Project Manual & Field Protocols

The Herbvar Steering Committee

2024-05-03

Table of contents

Preamble	3
1 What is HerbVar?	4
1.1 Motivation	4
1.2 HerbVar Goals	4
I New Collaborators: Start Here	6
2 First Steps	7
3 Contact Information	8
II Research Workflow	9
4 HerbVar Workflow Overview	10
4.1 Collecting new HerbVar Data	10
4.2 New analysis with Herbvar Datasets	10
4.3 Protocol	22
4.3.1 Data Recording Tips	23
4.3.2 Examples of different types of damage to reproductive material	24
4.4 Overview	34
4.5 Protocols	34
4.5.1 Option 1 – Comprehensive Patch Census	34
4.5.2 Option 2 – Walking Transect	36
4.5.3 Option 3 – Comprehensive Census of Subset	36
4.6 Overview	41
4.7 Objectives	41
4.8 Protocol – Tree Seedlings & Saplings	41
4.9 Protocol – Mature Trees	41
4.10 Selecting Leaves for Herbivory Estimates	42
4.11 Github Access	51
4.12 Making Edits	51
4.13 Rebuilding the Website	52

5 The Data Portal	53
5.1 Your Job After Someone Submits Data via the Portal	53
5.2 Updating the Portal	53
5.3 Deleting Old Data	55
5.4 Service Account FAQ Background Information	58
5.5 Tutorial	59
6 Google Drive Structure	60
6.1 Phase I Data Wrangling	60
6.2 Phase II Data Wrangling	61
6.3 Phase II Completed Surveys Versions	61
6.4 Phase II Metadata	61
6.5 Phase II Raw Data	62
6.6 Phase II Wrangled Data -	62
6.7 Miscellaneous Other Files	62
7 Wrangling Repository	63
7.1 Misc. Non-Manuscript Subset Scripts	63
7.2 Phase 1 Scripts	63
7.3 Phase 2 Scripts	64
7.4 Script Archive	64
7.5 Singleton Tasks	64
7.6 Manuscript Subsetting Scripts	65
8 To-Do List for Manual	66
8.1 Data Use Agreement	66
8.2 Authorship Guidelines Conversation	67
8.3 Herbivore Protocol Gray Area	67
References	68
III Appendix	69
HerbVar Publications	70
HerbVar Presentations	71

Preamble

This book is a manual for researchers involved the the [HerbVar Project](#). It includes a check-list for new collaborators, guidelines on accessing, adding, and using project data, tutorials for using the RStudio Project Templates for conducting analyses and preparing manuscripts, protocols for field work, and guides for administering the HerbVar Network resources (*e.g.*, website, data portal).

This guide is a [Quarto Book](#) hosted on the [HerbVar Network's Github site](#), so any team members can edit the text, add new sections, or make suggestions for improvement either by [pull request](#) or by [posting an issue](#) on the HerbVar Manual's repository. A tutorial on getting started with Quarto and RStudio can be found [here](#).

1 What is HerbVar?

1.1 Motivation

Published studies and personal observations suggest the distribution of herbivore feeding damage among individual plants within a population is often highly skewed such that most plants experience relatively low levels of damage, and a small fraction of plants experience disproportionately high levels of damage. Theory suggests that such variability can have dramatic ecological and evolutionary consequences. For example, variability among plants can lead overall herbivore population size to be greater or less than expected based on average plant quality and asymmetric fitness surfaces can lead to over-investment in defensive traits. *Surprisingly, despite the theoretical importance and potential generality of variability in herbivory, it has received little empirical attention, limiting our fundamental understanding of how plants and herbivores interact.* We are a global collaboration to quantify the distribution of herbivory for diverse plant species in multiple ecosystems across the world.

1.2 HerbVar Goals

The goals of HerbVar are (1) to assess if variability in herbivory is indeed a common feature of plant–herbivore interactions, and (2) to examine how the amount of variability and skew varies among different types of plant species, herbivore communities, and ecosystems. Quantifying general patterns in the distribution of herbivore damage within populations would be a major contribution to our fundamental understanding of herbivory. In addition, identifying the factors that correlate with variability in herbivory would provide the field with a new paradigm for describing plant–herbivore interactions and allow us to generate novel hypotheses about the ecology and evolution of plant–herbivore interactions. Among the factors under consideration are:

- Plant species identity,
- Plant functional traits (from literature),
- Plant ecology (e.g., rarity),
- Herbivore species,
- Herbivore functional groups,
- Ecosystem type,

- Latitude,
- ...and many others (e.g., seasonality, precipitation).

The HerbVar Website can be found [here](#).

Part I

New Collaborators: Start Here

2 First Steps

1. Add info to Collaborator Contact Information file
 - make sure you have:
 - orcid id
 - github account
1. Get added as “Contributor” to the HerbVar Shared Drive (see if this is still best way)
2. get added to slack
3. get added to herbvar zotero group
4. request access to github
5. Review Authorship Guidelines
6. Sign Data Use Agreement

3 Contact Information

Communicating with HerbVar collaborators:

Want to email the Steering Committee? Here's their contact info in a format that easily copy/pastes into an email "To" field

Need to convert this to a list-serve

Nora Underwood nunderwood@bio.fsu.edu, Brian Inouye binouye@bio.fsu.edu, Susan Whitehead swhitehead@vt.edu, Phil Hahn hahnp@ufl.edu, Lee Dyer nolaclimber@gmail.com, Emilio Bruna embruna@ufl.edu, Ivalu Cacho ivalu.cacho@gmail.com, Karen Abbott kca27@case.edu, Will Wetzel william.wetzel@montana.edu

Part II

Research Workflow

4 HerbVar Workflow Overview

HerbVar relies heavily on version control with Git/Github to facilitate data validation and management (Goodman et al. 2014, Yenni et al. 2019, Kim et al. 2022), promote collaboration (Boland et al. 2017, Braga et al. 2023), and streamline data access, analysis, and archiving (Champieux et al. 2023). Below we describe the workflow for (a) collecting data for HerbVar and uploading it for integration with the global data set and (b) analyzing some or all of the global HerbVar data set and preparing a manuscript based on these analyses.

4.1 Collecting new HerbVar Data

1. Review and Select Protocols
2. Collect Data
3. Create a repository for cleaning the data using the `new_dataset_template`.
4. Upload the clean version of the data using the portal

4.2 New analysis with Herbvar Datasets

1. Identify Questions, Submit for Review to make sure no overlap
2. Create a repository for harvesting, organizing, and analyzing the data to be used in the analyses using the `new_analysis_and_paper_template`.
3. This template can also be used to write the manuscript in markdown using a template such as `papaja` or one of the 'rticles' templates.

**Originally
pub-
lished:
Up-
dated:**

**Originally
pub-
lished:
Up-
dated:**

```
{=html}  
<!--  
quarto-file-metadata:  
eyJyZXNvdXJjZURpciI6InByb3RvY29scyIsImJvb2tJdGVtVHlwZSI6  
-->  
# Primary  
Protocol  
{#secondary}  
primary}  
“““{.quarto-  
title-  
block  
template='/Applications/quarto/share/projects/book/pandoc/title-  
block.md'}
```

editor: markdown: wrap: 72

- TODO: add links to Below the Field #3 below

Below, we provide a straight-forward and broadly applicable protocol to achieve project goal of quantifying variability in plant-herbivore interactions. This is HerbVar's Primary Survey Protocol. In brief, 30 randomly selected plant individuals in a site (\~population) are surveyed for herbivore damage and (possibly) herbivore abundance. Data are also collected on the nearest conspecific neighbor of each plant (for a total of N = 60 plants). These methods yield estimates of variability, skew, and spatial patterns (e.g., autocorrelation) in herbivore damage. For more information on the broader motivations and goals of the project see [here](#).

****Note on Alternate Protocols:**** The HerbVar Primary Survey Protocol is designed to work for many common plant growth forms and contexts, so we expect most surveys to use this protocol. The primary protocol, however, will not work for every plant growth form or context, so HerbVar has multiple alternative survey protocols. These include protocols for surveying the following:

- Reproductive damage: @sec-repro
- Low density/abundance populations: @sec-low_density
- Cacti and other succulents: @sec-succulents
- Mature trees: @sec-tree *(surveys of immature trees (i.e., seedling/saplings) use the Primary Protocol below)*
- Rhizomatous geophytes: @sec-rhizo
- Insect herbivores, galls, and mines: @sec-herbivore

If the primary protocol is not feasible for a species or site, then we suggest one of these alternative protocols. If none of these alternative protocols fits the situation, then collaborators may deviate from the primary protocol. We trust collaborators to decide how to adapt the primary protocol in ways that work for their systems. We suggest, however, that collaborators strive to follow the spirit of the protocol below: randomly select at least 30 plants from a site and census them and their nearest neighbors for herbivory and herbivore data. For a dataset to be usable in the overall study, it will have to be comparable to data collected using this protocol. Collaborators who deviate from the HerbVar protocols should carefully record their methods.

```
## Guidelines for Picking Plant Species
```

```
### Ideal Focal Species Abundance
```

The primary protocol works best for sites with at least \~90 plant individuals, such that it makes sense to sample individuals randomly. If your site has fewer than \~90 individuals of your plant species, then please consider comprehensively censusing all individuals within the site as suggested in the Low Density/Abundance Protocol. A comprehensive census, when feasible, would be even better than the protocol below. If plants are far enough apart, please take GPS coordinates for each plant. If a comprehensive census is not feasible, then please modify the primary protocol or the low-density/low-abundance guidelines to work efficiently with your species and site. Please reach out to the HerbVar coordinators if you have questions or want to check that your modifications will lead to adequate data.

Before the Field

1. **Please review HerbVar's Damage Estimation Training Guidelines in @sec-damage.** Note that information on precision of estimates and acceptable binning is contained in this document. This document contains valuable information on how to estimate percent damage on various leaves.\
2. **Please spend some time training and testing yourself and anyone working with you in the field using the [ZAX Herbivory Trainer] (<https://zaxherbivorytrainer.com/>)**. This web-based application, created by Dr. Angela Moles and Zoe Xirocostas, provide a risk-free environment for testing oneself on per-leaf damage estimation. Note that the app prompts you to assess damage to the nearest percent while our protocol is slightly coarser (see Table 1 of the Damage Estimation Guidelines). The app has two stages, one in which you assess damage and are immediately told how close you are to correct and a second where you assess 50 leaves, and the results are given to you at the end. Please feel free to focus on the first part of the app until you are confident (though you are of course welcome to do the second if you want extra training).
3. **Download the template datasheet for this protocol**. There are both digital (see siteData, densityData, and plantData sheets) and printable versions to facilitate standardized data entry. If you have a question on this, please feel free to reach out to [herbvar@gmail.com] (mailto:herbvar@gmail.com)

Primary Protocol

1. Pick a plant species (see "Guidelines for Selecting Plant Species" below)
2. Pick a site (see "Delineating a Site" below for advice)
3. Pick a time to sample (see "When to Sample" below for advice)
4. Determine a radius for your circular quadrats
 - a. **If you are sampling one of the HerbVar focal species** (*_Taraxacum officinale*, *Plantago major*, *Plantago lanceolata*): Use a radius of 0.4 m for your quadrats. This will standardize across surveys of these same species. Note that if your populations are sparse, you may use a larger radius following the other process or pre-calculated values (Table 1).
 - b. **If you are using any other species**, use the following process to determine a quadrat radius
 1. Estimate mean density of plant/m²
 2. Count the number of plants in 1 m² at 10 random locations within the site
 3. Calculate mean density (*D*)
 4. Use *D* to calculate a circular quadrat radius (*r*) that would on average contain 4 plants $*r* = \sqrt{4/*D*}$.
 5. This approach seeks an optimal, intermediate quadrat size that balances the costs

Rather than calculating, you may also use this pre-calculated set of radii (Table 1) for non-focal species. ***Remember, for focal species, please use 0.4 m***

Number plants/m ² (Density D)	**Quadrat radius (**r**)**
D 0.1	3.6 meters
0.1 < D 0.25	2.9 m
0.25 < D 0.5	1.9 m
0.5 < D 1	1.35 m
1 < D 3	0.9 m
3 < D 6	0.55 m
6 < D 10	0.4 m
10 < D 20	0.3 m
20 < D	0.23 m

: Table 1. Pre-Calculated Quadrat Radii

5. Lay a transect through the middle of the site (Fig. 1) Record GPS coordinates of origin, length (m), and compass direction (degrees) of transect (need to pick a coordinate system and precision)
6. Select center points of circular quadrats (Fig. 1). Randomly select 40+ points in the site by selecting pairs of random numbers. One random number represents distance along the transect (0-length of transect); the other represents distance left or right of the transect (left=negative, 0=center, right=positive). These are the center points of quadrats.
7. For each quadrat:
 - a. Locate a quadrat center point using transect and measuring tape or stick (Fig. 1)
 - b. Count and record the number of focal plants within r meters of the center point (Fig. 1).
 - i. See above for explanation of how to calculate r(or use values in Table 1).
 - ii. It may be helpful to place a stick vertically in the center of the quadrat, attach
 - iii. **Note:** this includes only rooted focal plant species individuals in the quadrat
8. Record other quadrat-level data
 - a. Percent cover of focal plant species (ignore non-focal species).
Note: this includes both rooted and not rooted focal plant species individuals in the quadrat but hanging over the edge from above
 - b. Percent cover of all non-focal plant species (ignore focal species)
 - i. These 2 percent covers could total more than 100% if they overlap
 - ii. If surveying understory plants, ignore forest canopy when estimating percent cover

9. If the circular quadrat has 0 plants, record a zero and continue to the next quadrat
10. If the circular quadrat has ≥ 0 plants: Randomly choose 1 of the plants within the quadrat to survey A quicker alternative would be to choose the plant closest to the quadrat center. But this is recommended only if you think it will produce an unbiased sample of plants from your site. Be careful about over-representing large and/or isolated plants (which will be closer to more points relative to small plants in crowded patches).
11. For the selected plant (1 per quadrat) record plant life stage I.e., seedling, vegetative, flowering, fruiting Note that if multiple stages are present, record all relevant stages (i.e., a plant can be both flowering and fruiting)
12. For the selected plant record plant size. Use judgment to pick best measure for your species E.g., standing plant height (ground to tallest living part), stem length, foliage diameter, stem diameter
13. Perhaps most importantly: for the selected plant record herbivore damage in 3 ways (see Damage Estimation Training Guidelines) Note that "herbivore damage" includes damage caused by both vertebrate and invertebrate herbivores.
 - a. Presence/absence of leaf damage: Record both (A) the total number of leaves on the plant and (B) the number of leaves with $>0.5\%$ herbivory
 - i. If the plant has 60 leaves total, please record the true numbers
 - ii. If the plant has >60 leaves total, randomly (arbitrarily) choose 60 and record those values. Please also make a note that you stopped at 60 (see template datasheet).
 - iii. Note that we are no longer including undamaged leaves in the following step, so the
 - iv. If plants have reproductive parts (flowers/fruits/seeds) that could have been damaged
 - b. Estimated percent damage on 10 randomly (arbitrarily) chosen leaves with herbivory damage ($> 0\%$ herbivory)

- i. One estimate per leaf (for a total of 10 estimates).
 - ii. Please strive to sample in a way that selected leaves will be representative of all leaves.
 - iii. If desired, you may use an application to estimate damage (e.g., LeafByte, etc.). However, this is optional.
 - iv. **Note that all selected leaves should have > 0% damage** (this is a change from the previous version of the protocol).
 - v. For leaves with herbivore-built leaf shelters (rolls and ties), please carefully peer inside the shelter to estimate damage.
 - vi. If damage is estimated visually and leaves are visibly damaged but < 0.5% (i.e., damage is very small), then estimate visually again.
- c. **Estimated percent damage across the whole plant,** optionally also breaking apart damage by type or even species of herbivore if possible (e.g., sucking damage versus chewing damage, add columns as needed).
- i. E.g., If a plant has 4 equally sized leaves and 2 of those leaves are 50% eaten, then whole plant has 25% herbivory
 - ii. But take leaf size into account when leaves vary in size
 - iii. **If this measure is not feasible to collect, measure 30 leaves** instead of 10 in step 13.
14. Record presence of plant diseases (i.e., pathogens)
- Please also estimate your confidence in your pathogen estimate and include it as a note in the data.
15. Record number of leaf mines and galls per plant
- a. If there is reason to believe that galls or mines have accumulated through multiple years, estimate the number per plant.
 - b. If there are too many mines or galls to count individually, estimate the number per plant.
 - c. If serpentine/linear mines cannot be confidently recorded, instead count only blotch mines.
 - d. Optional: abundance of other externally-feeding herbivores (standardized approach; see Herbarium guide).
16. Distance to nearest conspecific neighbor (where the nearest neighbor is the plant with the closest above ground tissue to any aboveground tissue on the focal plant)
17. Data to record for the first nearest conspecific neighbor of selected plant: All the same data as focal plant except nothing for

neighbor's neighbor

18. Continue visiting the randomly selected points until 30 focal plants and 30 nearest neighbors have been surveyed

After the Field

1. Enter your field-collected data into the template Excel file. Refer to the data Dictionary sheet if column meanings are unclear
2. Use the data submission portal to upload your data. The portal has numbered steps to assist the upload process.
3. After uploading via the submission portal, check the Completed Surveys file to ensure that your data were uploaded successfully. Uploaded data will have your entries in the sidebar of the app as the bottom-most row of that file

![Fig. 1. A diagram of the sampling scheme described in the text. (1) Record plant density in 10 randomly located 1-m² areas to estimate average plant density D, which is used to calculate quadrat radius r. (2) A quadrat with one focal plant and its nearest neighbor (outside quadrat). (3) A quadrat with no focal plants. (4) A quadrat with 5 focal plants; plant 3 is randomly selected for data collection, and its nearest neighbor is plant 4. Diagram by Moria Robinson.] (images/img1.png)

Methods Notes

1. Modifications of this protocol may be necessary to adapt it to different systems (see "Note on Alternate Protocols" above). If the primary protocol won't work for your system, please first consult our alternative protocols (see protocol above and "Protocols - Phase 2" folder). If our alternative protocols do not solve the issues, then you may adapt the primary protocol as needed. **Whatever you do, please record methods carefully and strive to follow the spirit of the protocol and produce comparable data.**
2. In Phase 1, collaborators reported that one survey (~60 plants) took between 0.5 and 2 person days (4-16 hours) using the methods above (after a species and site have already been selected).

3. We select 40 quadrat center points (instead of 30) so that we have extra points ready in case some quadrats are empty. If you predict that many quadrats will be empty (e.g., in a very spatially clumped population of plants), then select more points (e.g., 60 points). (Remember the goal is to have 30 focal plants sampled, plus their nearest neighbors).
4. Sometimes, especially in small populations, a focal plant ends up being another focal plant's neighbor. This is fine. Just note and keep going. If you have time, you can add an extra focal plant at the end (but this isn't totally necessary).
5. For clonal plants, we have been calling stems "individuals" if they are not connected aboveground. When looking for above ground connections, we clear away detritus, but we do not dig or move soil. There is also a dedicated alternate protocol for surveying such species (see Rhizomatous Geophytes Protocol)
6. Please see our Damage Estimation Training Guidelines for guidelines on how to estimate herbivore damage. Here are two tips:
 - a. Sometimes discerning herbivore damage from physical damage (e.g., wind, trampling) is tricky. We do the best we can. We look at things like how jagged the cut edges are and if they travel past the missing area into the remaining leaf tissue (which would suggest the damage may have been physical).
 - b. Another challenge is old damage that occurred when leaves were still expanding. This could potentially make the area removed seem larger than it was. If we suspect something like this happened, then we try to bend the leaf back into shape to see if it seems like the missing area expanded over time.
7. We will accept surveys that only assess damage and do not identify herbivores. This will allow people without insect ID skills to participate in the study.

Guidelines for Picking Plant Species

From the patterns found in the first phase of data collection we have developed the following objectives going forward:

1. Surveys of the three focal species (**Taraxacum officinale**, **Plantago lanceolata**, and **Plantago major**), especially across broad environmental and/or geographic ranges
2. Surveys of species in the five focal families (Apocynaceae, Asteraceae, Fabaceae, Rubiaceae, and Solanaceae). We want surveys of new species within these families, especially species from new clades or with unusual growth forms. For repeat surveys of species within these families, we are prioritizing surveys from new regions, habitats, elevations, etc.
3. Surveys of damage to any species' reproductive tissues (e.g., flowers, fruits, etc.).

While we welcome all surveys, data that fall under one or more of these three guidelines is particularly valuable in addressing the current scope of HerbVar's research questions. Please refer to our more detailed HerbVar Species Selection Protocol for more information on species selection and how data contribution relates to authorship in papers utilizing those data.

Delineating a Site

We realize that defining the 'edges' of a site can be subjective and not easy. We search for an area where a given plant species occurs at a high enough density to easily select 30 focal plants and 30 unique neighbors with our method. This is usually a relatively dense patch. Walk around and see if you see the density drop off to well below the mean density that is used to calculate radius size. This is usually quite simple, e.g., when we walk out from the center of a "site" and don't see any individuals of the focal species within 5 m, we decide we're at the edge of a patch. In some systems, delineating a single, sampleable population simply might not be possible (e.g., where a species covers a vast area). In these cases, collaborators should simply do their best to select a reasonable, representative area to sample.

When to Sample

This will depend on the natural history of the system. We will accept data sampled at any time as long as there has been some herbivory. We

will use the sampling date to examine how herbivory changes seasonally (please note approximate dates for beginning and end of growing season for each survey, see siteData sheet in datasheet template). However, the most valuable surveys will be after enough time has passed for an ecologically meaningful amount of herbivory to accumulate. In strongly seasonal systems, this will be in the latter half of the growing season. But it could also be once leaves have reached maturity (e.g., for species in which most herbivory is on expanding leaves). In other systems, the best time to sample might be during or after a key life history stage (e.g., flowering). All that said, there is no perfect time to sample. Collaborators should use their knowledge to decide when to sample (and sample when is feasible; some data is better than no data!). And repeat sampling is acceptable.

```
## Common Garden Data
```

Common gardens are a powerful tool for studying plant-herbivore interactions. Several collaborators have proposed including them in HerbVar, and we would like to try if we can get enough data. To be applicable to this study a common garden's design would have to be random with respect to genotype. If a garden was somehow stratified with blocks containing repeated instances of, e.g., different levels of leaf toughness, then damage distributions will not be comparable to damage from wild populations. We may still be able to use such datasets, but only if we have enough to use them in a separate analysis. Please get in touch if you would like to contribute common garden data.

Originally published:\

Updated:

```
`<!-- quarto-file-metadata: eyJyZXNvdXJjZURpcii6InByb3RvY29scyJ9 -->`{=html}

`{=html}
<!-- quarto-file-metadata: eyJyZXNvdXJjZURpcii6InByb3RvY29scyIsImJvb2tJdGVtVHlwZSI6ImNoYXB0Zz
`-->

# Reproductive Damage {#sec-repro}

`{.quarto-title-block template='/Applications/quarto/share/projects/book/pandoc/title-'
`-->
```

author: 'Phil Hahn, Susan Whitehead, Paul Ode, & Will Wetzel'

This protocol aims to assess damage by insect herbivores to reproductive parts of plants (i.e., flowers, fruits, and/or seeds). This is a supplement to the Primary Protocol, which aims to randomly select and sample 30 plants, plus their nearest conspecific neighbor, within a population.

Objectives: The goal is to measure the proportional damage to reproductive organs on each plant within the surveyed population. That is, for each individual plant we will record the number of damaged and undamaged reproductive organs. Ideally, these measurements should be taken as supplemental data for the same individual plants (focal and neighbor plants) for which leaf damage was taken for the primary protocol, although reproductive damage measurements are welcome from other plant populations from which no leaf damage measurements have been taken.

4.3 Protocol

1. **Select a species to survey:** We are hoping to get broad taxonomic and geographic coverage of damage to reproductive organs. Therefore, any species could be surveyed. However, to ensure that the data are comparable across sites/species/families/etc, the plants should have the following characteristics:
 - A. At least half of the individuals at your site should possess reproductive material. If most of the plants are in a vegetative stage, you probably won't be able to survey enough reproductive individuals to get a decent sample size. Ideally, >30 of the sampled plants will have reproductive damage data
 - B. Each individual plant should produce enough flowers/fruits/seeds so that you can survey between 15-30 reproductive units per plant. These 'units' could be flowers, fruits, or seeds, whichever will give you enough things to count.
 - i. E.g., if your plants have just a few fruits, try opening fruits and counting a random sample of seeds, which could get you higher numbers. If some plants have fewer units (~5-10 flowers/fruits/pods/seeds), that is okay, as long as most have at least 15.
 - ii. If your species typically has only one or a few flowers, we have provided a modified protocol below
 - C. Most of the plants should be in a similar phenological stage. If there is a mixture of flowering and fruiting plants within the population, it might be difficult to get a large enough sample

size for one organ type. Additionally, different phenological stages will likely be attacked by different insects

D. If your plant does not meet these requirements, please skip measuring damage to reproductive organs. That's okay. Or get in touch if you have questions

2. Use the Primary Protocol to establish a transect, pick/calculate a quadrat radius, and randomly select focal/nearest neighbor plants

3. Identify the reproductive tissue you will be surveying, e.g., flowers, fruits, pods, seeds, etc.

4. For each plant then record (see Fig 1):

A. Phenology of the plant (i.e., budding, flowering, fruiting, etc.)

B. Total number of units examined

C. Herbivore damage per unit: Either percent or count of injuries

D. Pathogen damage per unit

E. Unknown damage per unit (i.e., any damage not attributable to either herbivores or pathogens)

F. Ideally at least 30 reproductive units will be examined per plant. For plants that have much more than 30 units, examine a haphazard subsample of 30 units per plant. These should be sampled from different parts of the plant to obtain a good representative sample of damage levels to the plant. Damage measurements will be highly variable among species, depending on the types of flowers or fruits produced as well as the type of damage that is most common in the population. Below we give some general guidelines.

G. There likely will need to be modifications for some species, but we trust collaborators to do the best they can in their systems while maintaining the overall spirit of the protocol. For example, if your species has only one or a few flowers/fruit, you can estimate the damage as a percentage. Keep in mind that the goal is to capture the variability in damage rates among plants within a population, so you will want to choose a measure of flower/fruit/seed damage that best captures this variability.

4.3.1 Data Recording Tips

Record the following data in the provided printable and digital datasheets.

1. Start by surveying the type of damage that is most common on the reproductive structures of plants in the population. Look for damage by insects that may chew on developing flowers (e.g. katydids, beetles), insects that bore into flower heads or seeds (e.g. larval weevils, leps, or flies), or true bugs that may probe/pierce into the seeds or fruits (looks like little black dots on the fruits). In many cases you will need to tear open the seed head/fruit to look for boring insects inside the seeds/fruit.
2. Look for signs of chewing damage inside the fruit, such as destroyed seeds and insect frass.
3. The best measurements of damage will depend on the type and extent of damage present. If the plant species experiences damage to multiple organs (e.g., petals, stamens, etc.), focus on the damage to the primary reproductive parts if it is not feasible to measure multiple organs.
4. In the “reproUnit” column of the reproData sheet in the template Excel datasheet, please record what organs you are recording damage on (e.g., stamens, petals for flowers). If present and identifiable, record the number and identity of the florivore, frugivore, seed predator
5. If your species has only one or a few flowers per plant and it’s possible to record the percent damage to a flower or fruit, you can record these percent damages in the percR# columns. 15 such columns are provided but please add more if you are able to record the per-unit damage of more than 15 reproductive structures
6. Indicate whether you are recording ‘count’ or ‘percent’ damage in the ‘damageUnit’ column.

plantID	reproUnit	phenology	damageUnit	numRepro	numReproHerb	percReproHerb	numReproPath	numReproChew	numReproPierce	numReproUnkDmg
1	fruit	fruiting	count	19	17		5	10	0	2
2	fruit	fruiting	count	26	25		10	5	8	0
3	fruit	fruiting	count	21	2		0	2	0	0

Figure 4.1: Example reproData information

Note 1: numReproHerb can be less than the sum of the counts of types of damage because a single reproductive unit can have multiple types of damage.

Note 2: In the damageUnit column, record ‘count’ or ‘percent’ depending on whether you are recording the number of damage reproductive units or the percent damage to a single unit

4.3.2 Examples of different types of damage to reproductive material



(a) *Abronia umbellata* with chewing damage to two corolla (one is completely chewed and one is partially chewed). In this example, each floret would be scored for damage (2 damaged, 13 organs examined). Photo: Eric LoPresti.

(b) *Abronia turbinata* with damage and frass from a moth caterpillar (*Neogrotella macdunnoughi*). Note chewing damage to petals/corollas and caterpillars at base of florets. In this example, each of 30 florets would be scored for damage. Photo: Eric LoPresti.



Figure 4.3: *Vicia americana* seed pod. Notice boring holes in the upper-left seed and frass in the pod. In this example, each of 30 seeds would be scored for damage. Photo: Phil Hahn.



(a) *Monarda fistulosa* seed head ripped open to reveal a weevil larvae. Notice the damage holes to the floral tubes. In this example, each flower head would be recorded as damaged (or undamaged) and ideally 30 flower heads would be assessed per plant. Often this species does not produce 30 flowers per plant, so smaller numbers would be acceptable. If most plants have fewer than 10 flowers, this would not be a good population to survey for reproductive damage. Photo: Phil Hahn.

(b) *Lonicera* fruit with chewing damage. In this example, each of 30 fruits would be scored for damage. Photo: Susan Whitehead.

Aquilegia shockleyi with chewing damage to fruits (from *Heliothis phloxiphaga*). In this example, each fruit could be scored for damage, or each fruit could be opened and seeds could be counted (infer missing, fully-consumed seeds from pod features). Photo: Eric LoPresti.

Originally
pub-
lished:
Up-
dated:

author:
>-
Her-
bVar
Dam-
age
Esti-
ma-
tion
Sub-
group
(Ian
Pearse,
Zoe
Getman-
Pickering,
Julie
Davis,
Bastien
Castag-
ney-
rol,
and
Will
Wet-
zel)

Overview

So, you've figured out how to find the random set of plants to score damage on (e.g., using `sample`)

****In this document, we provide detail for four key steps from the Primary Protocol:****

1. Estimating plant size (and determining what tissue counts).
2. Counting number of leaves and number of damaged leaves (up to 60 max).
3. Estimating percent damage on 10 randomly chosen leaves.
4. Estimating percent damage across the whole plant

::: {.callout-note}

If you are just looking for quick tips on estimating percent damage, skip to "Estimating Percent Damage Across the Whole Plant".

:::

****Two related documents in the "Damage estimation training" folder in the Shared Drive are****

1. A field guide to types of plant damage, which has photos of different types of damage
2. An illustrated guide to amounts of percent damage, which shows images of leaves with different levels of damage.

****Objectives: ****First off, it is good to keep in mind the goals of these estimates. Ideally, we want to estimate the amount of damage to a plant.

::: {.callout-tip}

What kinds of plants will this work for?

This document is written with a focus on vegetative tissue on leafy plants 2 m in height. It does not cover fruiting or flowering plants.

:::

****If your species is a tree or large shrub, then you have two options: ****First, we recommend using a diameter at breast height (DBH) to estimate the volume of the plant.

****If your species is a cactus or other succulent,**** then please see the Cactus and Succulent section.

****If your plants have reproductive tissues**** (flowers, fruits, seeds) and have had them long enough, then please see the Reproductive Tissues section.

Estimating Plant Size (and Determining What Tissue Counts)

In many cases, it will be clear what an individual plant is, but in cases of clonal plants, we need to determine which clonal plant is being measured.

In most cases, herbivores do not eat all plant biomass. Therefore, it will be useful to note the amount of biomass lost due to herbivory.

```
! [Lepidium] (images/img9.png){#fig-lepidium fig-align="center" width=80%}
```

Once you have decided the extent of plant you will be surveying, you can measure plant size.

2. Counting Number of Total and Damaged Leaves (up to 60)

The first damage assessment step in the HerbVar Primary Survey Protocol is estimate the proportion

If you have a plant that has a small number of large leaves (e.g., 1-3), then the proportion

If your plant has < 60 leaves, we encourage you to quickly count and scan all of the leaves

If you are restricted to examining a subsample of leaves within plants (because your plant has > 60 leaves), then we encourage you to use one of the following methods:

Subsampling Method 1 - Nose-Pointing

Ian Pearse's nose-pointing method: For large plants, I like to choose four positions around the plant. Then, I turn facing the plant, open one eye, and I choose whatever leaf I am pointing to (or closest to my nose).

```
::: {#fig-nose layout=ncol=2}
```

```
![Is that creep trying to pick his nose with his thumb? No. This is the nose pointing method]
```

```
![He then turns to the plant, opens a single eye, and chooses the leaf that he is pointing to]
```

```
:::
```

Subsampling Method 2 - True Randomization

True randomization: Assign all leaves, seeds, etc. on the plant a random number, draw N numbers from that pool.

Subsampling Method 3 - Arbitrary Sampling

Arbitrary sampling. That sneaker-word, arbitrary! This is basically to say "I really tried to

Subsampling Method 4 - Make Your Own

Design (and make notes of!) your own subsampling scheme. Can you choose every seventh (or random) leaf?

! [legend needed] (images/img12.png){#fig-random width=70%}

::: {.callout-tip}
Philosophical thoughts about counting damaged leaves

This method works well (we think) as an estimate of overall herbivory for plants with small numbers of leaves.

! [legend needed] (images/img13.png){#fig-arbitrary width=70%}

:::

3. Estimating Percent Damage on 10 Randomly Chosen Leaves

The next damage assessment step is estimating percent damage on 10 randomly chosen leaves. We will assume that the first 10 columns (percLf10) come after the column for whole plant percent damage (percHerbPlant) in the HerbVar table.

Randomly choose 10 leaves:

If your plant has 10 or fewer leaves, then please examine them all; if your plant has more than 10 leaves, then randomly choose 10 leaves.

Estimate percent damage on each leaf

Finally, the main event--this is probably the main reason you're reading this document. There are many ways to estimate percent damage on a leaf.

Visual estimation

Visual estimation is essentially as simple as it sounds. You look at a leaf and eyeball what percentage of the leaf is damaged.

1. How to estimate damage

Record estimates at a high resolution. We usually record at a resolution of 2.5% (Table 1). This means that even with considerable error--high resolution estimates will likely be closer to the true value than low-resolution estimates.

Table 1. Recommended resolution for recording percent herbivory

Percent	
Meaning	
0%	No herbivory
0.5%	A trace amount of herbivory
1%	~1% herbivory
2.5%	~2.5% herbivory
5%	~5% herbivory
7.5%	~7.5% herbivory
So forth	So forth
Up to 100% (e.g., everything removed except base of leaf petiole)	Up to 100% (e.g., everything removed except base of leaf petiole)

2. When you first look at a leaf, do a quick mental calibration before estimating damage. We

3. When it is time to do the actual herbivory estimate, one strategy that works well for refining from a coarse estimate to a precise estimate. For example,

- If ~12.5% of a leaf were damaged, then...
 - Mentally cut the leaf into quarters
 - See that less than a quarter (25%) is damaged
 - Mentally cut the quarter with damage in half, yielding eighths (12.5%)
 - See that the area damaged is equal to an eighth and record 12.5%

- If ~30% of a leaf were missing, then...
 - Mentally cut the leaf in half
 - See that less than half is damaged
 - Mentally cut the leaf into quarters
 - See that more than a quarter (25%) is damaged
 - Take mental note of the 25% damaged, and then focus on estimating how much more than 25% is damaged
 - Mentally halve the quarter of the leaf with the excess damage above 25%, yielding eighths
 - See that the damage above 25% is a little less than half of one of those eighths, which is 6.25% plus a little less than 6.25% comes close to 30%, record it!

4. If your leaf has more than one area of damage, try mentally consolidating each area of damage.
5. An acetate grid can be a very helpful tool. Some people use them to help guide their estimates.
6. For complexly pinnate leaves (e.g., Apiaceae), it is probably best to divide the leaf into smaller sections.
7. If damage is very high and very little leaf tissue remains, take a large and small leaf and compare them.
8. If you have marginal damage on leaves with non-smooth margins: If you draw an entire margin, it will be difficult to estimate the amount of damage. Instead, it is easier/more accurate to imagine/draw a straight line than margin teeth.
9. Piercing-sucking damage, when visible, should be mentally consolidated and estimated similarly to other types of damage.

! [Legend Needed.] (images/img14.png){#fig-location}

10. For leaves with herbivore-built leaf shelters (rolls and ties), please carefully peer into the shelter to see what is causing the damage.
11. Through all of this, make sure you are correctly identifying what is herbivore damage versus other types of damage.
12. For internal feeding insects (e.g., hackberry psyllids, below):
 - Count discrete units: count either the number of insects or the number of galls or mines.
 - Mines should be included in percent damage and counted as discrete units.
 - Galls should only be counted, not included in percent damage because galls are actually plant structures.
 - Keep an eye out for signs of stem-boring insects. Sometimes these can be counted.

! [Internal feeding insects] (images/img15.png){#fig-internal}

How to make sure you're doing a good job

1. Be conscious that most people overestimate low levels of tissue damage (Johnson 2016). Try to estimate from the outside in.
2. Invest time in practicing, calibrating, and validating estimates. Especially do this before the field season begins.
3. Standardize among observers you work with or have a single observer for all estimates.
4. Print out An illustrated guide to amounts of percent damage. Study it while practicing, and keep it nearby.
5. Take our online herbivory estimation training quiz (in development, we will add the link later).

- Finally, ground truth a subset of your damage estimates using a digital method. When doing this, consider the following:
 - LeafByte: This is an app developed recently by scientists at Cornell (including our very own Dr. Jennifer Dillaha-Taylor) that allows you to scan leaves and estimate damage.
 - Scan leaves and estimate damage with Image J. For this, I usually collect leaves into a zip file and upload them to a cloud storage service like Google Drive or OneDrive, then use Image J to open the files and estimate damage.

! [Three steps in measuring damage with Image J.] (images/img16.png){#fig-imagej}

4. Estimating Percent Damage Across the Whole Plant

The final damage assessment step is estimating percent damage across the whole plant, or as many plants as possible.

Tips for visually estimating damage across the whole plant:

Effective methods will vary a lot based on the size of plants, size of leaves, and architecture.

We often find it helpful to pick a reference leaf size on which to base mental calculations.

An important tip to remember for speed is that when plants have more than ~9 leaves, leaves will likely be damaged. For larger plants and plants with many small leaves, it is impractical to scan each individual leaf.

Please let us know if you have additional tips, suggestions, or guidelines we can add to this page!

::: {.callout-important}

If you're in an area with tick-borne diseases, don't forget to check for ticks after!

:::

**Originally published: **

**Updated: **

```
`<!-- quarto-file-metadata: eyJyZXNvdXJjZURpcii6InByb3RvY29scyJ9 -->`{=html}
```

```
```{=html}
```

```
<!-- quarto-file-metadata: eyJyZXNvdXJjZURpcii6InByb3RvY29scyIsImJvb2tJdGVtVHlwZSI6ImNoYXB0ZD-->
```

```

```

```
Low Density/Abundance Plants {#sec-low_density}
```

```
-----{.quarto-title-block template='/Applications/quarto/share/projects/book/pandoc/title-
```

```

```

```
author: 'Danielle Salcido, Lee Dyer, Will Wetzel'
```

```

```

## 4.4 Overview

This protocol outlines three methods for surveying sites where the focal plant occurs at low density or low abundance. The Primary Protocol was designed to work for many plant species, growth forms, and contexts, but it requires sites with enough focal plants at a reasonably high density for efficient random sampling using our transect/sub-transect method. If the focal plants at your site are at very low density, then sampling them with our primary method will be very time-consuming due to the large distances between plants. If they are at low abundance, such that there are fewer than about ~90 plants in the site, then it does not make sense to draw a random sample of 30 plants + 30 neighbors from such a small population. If none of the methods below work well for your species and site, we encourage you to think of a comparable alternative. Feel free to get in touch if you have questions. Regardless of what you decide, please make sure to carefully document your methods.

## 4.5 Protocols

We provide three separate protocols for surveying sites with low density and/or abundance of focal plants: (1) Comprehensive Patch Census, (2) Walking Transect, (3) Comprehensive Census of Subset.

### 4.5.1 Option 1 – Comprehensive Patch Census

The best method, when feasible, is to census all of the individuals in a patch. This will work when there is a well-defined patch with a reasonable number of plants (e.g., < ~90). If this is possible, it is better in many ways than the Primary Protocol because it describes the whole distribution of herbivory at the site—there's no risk of missing the tails of the distribution if every plant has been included! Also, depending on the context, this could be faster than our Primary Protocol because setting up transects and quadrats to randomly sample plants is time consuming. So comprehensively examining all of a patch's 90 plants, for example, might

be faster than randomly sampling 60 plants (30 focal plants and their 30 neighbors) from a larger population. For this method, we stress that you should strive to survey every plant.

Record spatial information via one of two methods:

- 1) **Option 1A - Record GPS position of each plant:** If you do this, you will not need to record nearest neighbor information because we can reconstruct it (and more!) easily from the geographic coordinates. This of course means that you will need a GPS sensitive enough to differentiate the locations of your plants. If your plants are on average more than 2-3 m apart, then almost any modern GPS will be precise enough. If, however, your plants average less than ~1 m apart, then you would need a very precise GPS to describe the relative locations accurately. If you don't have such a precise GPS (or if you don't like how slow a precise GPS can be), we recommend the second method.
- 2) **Option 1B - Relative spatial coordinates:** You can measure the relative coordinates of your plants using two tape measures or a tape measure and a meter stick. This sounds similar to the primary protocol but it's much quicker because you're not using the tapes to select plants, just to record their locations.
  - Lay a tape measure through your patch. For each plant, record spatial coordinates as how far along the tape measure and how far from the tape measure. You can situate the tape either along the edge of the patch or through the middle of the patch.
  - If your tape measure is through the middle of the patch, remember to record the distance left of the tape as negative and right of the tape as positive. The start of your tape will have the coordinate (0,0).
  - After recording this information, follow the Primary Protocol as closely as possible

### Other useful information

1. Record `popDiameter1` and `popDiameter2` as the approximate extents of your patch/census area
2. For focal plant percent cover (`focalPlantCover`) and non-focal plant percent cover (`otherPlantCover`), please follow the Primary Protocol methods for estimating population density and calculating a quadrat radius size, if feasible. You can then center a quadrat on each focal plant in the census to define an area around each focal plant for recording focal and non-focal percent cover, as well as the number of focal plants in quadrat (`numPlantsinQuad`).
3. In comprehensive surveys the ~60 plants will all be focal plants and there is no random selection; hence the nearest neighbors (all the ".1" plant IDs in datasheet template) become focal plants. As described above, a quadrat is centered around each plant and the three quadrat-level variables (`focalPlantCover`, `otherPlantCover`, and `numPlantsinQuad`) are recorded for each plant. The nearest neighbor distance (`NNdist`) is still recorded,

but since that neighbor is treated as a focal plant, please record the unique plantID of that nearest neighbor as well (e.g. add NNplantID column as needed in datasheet).

4. If you cannot estimate population density (e.g., because your species is too sparse), then please pick an arbitrary quadrat radius. You can use that to define an area around each focal plant in your census for estimating percent cover variables. A 1-m radius might be a good choice for many plants, but go bigger for bigger plants. Remember to record your choice!

#### **4.5.2 Option 2 – Walking Transect**

Another alternative if you have widely dispersed plants that do not form a well-defined patch (or the patch is too large for a comprehensive search) is a walking transect.

1. Randomly pick distances (e.g., paces) along a transect and from a transect.
2. Pace out the distance along the transect, then turn orthogonally to pace out the distance from the transect.
3. Survey the closest plant within some reasonable distance (if no plant is reasonably close, then go back to transect and keep going).
4. Repeat until you have 30 plants and 30 neighbors.

This is similar to the Primary Protocol except pacing (rather than measuring with a tape) can make large areas more feasible to survey. Consider recording spatial coordinates for each plant, especially if plants are far from your randomly identified points. And try to survey neighbors for each plant.

#### **4.5.3 Option 3 – Comprehensive Census of Subset**

This method is similar to the comprehensive census of a patch (#1 above), but it applies when there is no well-defined patch and individuals are widely dispersed over a large area. There are two ways to do this but for both of these methods, record spatial coordinates for each plant and see other notes for method #1 above.

1. **Option 3A - Comprehensive survey of all plants along a transect:** With this method, you are doing a comprehensive survey of a linear subset of the whole population
  - a. Start by randomly picking a transect starting point and direction
  - b. Walk the transect and survey every plant that crosses your path *or* every plant within a reasonable distance of your path (e.g., 2 m)
  - c. Keep going until you get at least 60 plants

**2. Option 3B: Comprehensive survey radiating out from a random starting point within a population.** With this method, you are doing a comprehensive survey of a roughly circular (or blobby) area within the whole population. The cons of this approach are that if your plants are close together there could be high spatial autocorrelation such that you fail to capture the range of herbivory levels in the population. Of course, this is always a risk; it's just especially acute when the sampling extent is an arbitrary area rather than a biologically significant "patch".

- a. Explore outwards from your random starting point, surveying every plant you encounter until you get to at least 60 plants. *We do not recommend doing this unless your plants truly are all widely dispersed.*

\_\_\_\_\_  
**Originally  
pub-  
lished:**  
**Up-  
dated:**

\_\_\_\_\_  
author:  
Judie  
Bron-  
stein  
and  
Vic-  
toria  
Luizzi  
(Uni-  
ver-  
sity  
of  
Ari-  
zona)

## ## Overview

This document discusses issues relevant for quantifying herbivory on cacti and outlines a hope

**\*\*Unique Context:** \*\* With regard to quantifying herbivore damage, cacti are special: (a) they

## ## Sampling Issues within Architectural Categories

Cacti can be thought to consist of one or more (usually spiny) tubes with flowers usually located at the top. That's what they have in common - with these tubes having a diversity of spatial relationships that's what makes them different from each other. There are at least five categories, each with its own sampling issues:

- 1. Single, unbranched tube stuck in the ground (e.g., in the American Southwest, a barrel or a pencil cactus).

Entire structure should be scrutinized for herbivore damage. When these cacti form a cluster, sample all individuals.

2. Set of unbranched tubes connected underground (e.g., a hedgehog cactus (small), or a senita cactus).

Either the whole thing can be scrutinized for damage, or a subset of units could be (please make a note of which). This applies to both small and large sets of tubes.

3. Tube that starts to branch above the ground as it ages (e.g., a saguaro cactus).

Same method as #2 (again, please make a note of whether you examined the whole plant or a subset).

4. Large set of tubes connected at distinct joints (e.g., a cholla cactus). New tubes are added as the plant grows.

Subsampling within individuals will usually be necessary because individuals often have many tubes.

5. Large set of tubes connected at distinct joints but flattened into pancakes (e.g., a prickly pear cactus).

Same method as #4.

## ## Expected Types of Herbivory

Bites that remove chunks of flesh.

This is quite obvious for the fifth category of cacti (flattened pads), because the pads have a very thin skin.

Scarring of the surface of the cactus.

It can be very hard to know what causes this - some of this damage may be attributable to herbivores, others to physical weathering.

Colonies of sucking insects.

In particular, cochineal bugs live in colonies and are exciting to see. They are covered with a waxy coating.

## ## Other Notes About Cacti

Some genera have species with extrafloral nectaries (EFNs). Most (not all) barrel cacti have

It seems likely that the newest, tenderest units (particularly in categories 4 and 5 cacti) a

The buds and young fruits of some cacti get very heavily attacked, and these should be includ

## ## Protocol

We designed this with prickly pear (*Opuntia* spp.) in mind, but it should work essentially the

The gist of the protocol involves following the Primary Protocol except for a subsampling of

## ## Pre-Census Tasks

Pick a species to census.

Choose a site, ideally with at least 90 well-defined individuals that you can randomly sample

Decide on a maximum number of pads per plant to census.

We recommend focusing only on young cactus pads. But you should decide if this will do a good job. Older pads can be many years old, thus integrating herbivory over a much longer time than happens with young pads.

Practically, it's very hard to determine and quantify what is herbivory vs physical damage or disease. Physically, it can be hard and dangerous to access older pads on spiny plants!

If you think focusing on young pads will not be good for your species, then please modify the protocol. Take detailed notes.

Ideally, investigate the major types of damage you may see, potentially making a cheat sheet. See how long the protocol outlined below would take, then modify as necessary.

Please take detailed notes on any modifications made

## ## Census

Record site characteristics (e.g., date, site, plant ID, etc.)

Decide how you will define an individual plant. Past populations we surveyed had many very large plants.

See above discussion of architectural categories of cacti/succulents

If you have a site with >90 plant individuals, follow the Primary Protocol from the beginning.

Pick transect and subtransect distances that will encompass your site and lay the transect there.  
Estimate the density of plants in the population.

Use the estimated plant density to calculate a quadrat radius to use for the survey.

Randomly generate x,y points, visit them, and set up a quadrat centered on each random point

Once the first plant is selected, survey it for herbivory. For vegetative herbivory, we recommend quickly scan all of the terminal pads across the entire plant and visually estimate percent loss.

Randomly select 20 of the terminal pads. Record the number of pads you examined and the number

Randomly select 10 terminal pads, and record a visual estimate of the percent herbivory on each.

If your plant has reproductive organs, please randomly (arbitrarily) select up to 20 units of

Note presence of pathogens.

— — — —

**\*\*Originally published: \*\***

**\*\*Updated: \*\***

```
`<!-- quarto-file-metadata: eyJyZXNvdXJjZURpcI6InByb3RvY29scyJ9 -->`{=html}
`-->`{=html}
<!-- quarto-file-metadata: eyJyZXNvdXJjZURpcI6InByb3RvY29scyIsImJvb2tJdGVtV
`-->
```

# Trees {#sec-tree}

```
-----{.quarto-title-block template='/Applications/quarto/share/projects/book/pandoc/title-block.html'}
```

卷之三

### Tree Sub

rice Subgroup (Bastien Castagné, Amy Brewster, Will Weller, Nelia Robinson)

## Protocol – Tree Seedlings & Saplings Protocol – Mature Trees Selecting Leaves for Herbivory Estimates

## **4.6 Overview**

Mature trees, though harder to study than smaller plants, are a key plant growth-form that could have their own characteristic patterns of interactions with herbivores. Therefore, we want to include enough surveys of mature trees in HerbVar's global sampling to allow us to compare patterns between trees and other growth forms. It is also important to include mature trees because there may be major shifts in tree-herbivore interactions with tree ontogeny, from seedling to sapling and sapling to adult.

## **4.7 Objectives**

Provide a protocol for sampling mature trees. Collaborators who do not have a special interest in working with mature trees should restrict their surveys to individuals < 2 m height (seedlings and saplings). That is, survey tree species, but focus on seedlings and saplings. Seedling-sapling surveys won't be representative of all the individuals in a population of a tree species, but these are key stages in tree ontogeny—perhaps the stages in which herbivory is most influential. We are taking a two pronged approach to including tree species in HerbVar.

## **4.8 Protocol – Tree Seedlings & Saplings**

Follow the Primary Protocol. This includes (but isn't limited to) the following data:

1. Leaf-level percent herbivory estimates for 10 randomly selected leaves
2. Counts of presence/absence of herbivory for up to 60 leaves per plant
3. A whole-plant visual estimate of herbivory
4. And of course please record the number of galls, mines, and other discrete damage types from sessile herbivores

Please note in the metadata that you surveyed only immature individuals (< 2 m) at your site. Such a note can be complemented by recording the height of the individuals in the plantSize columns

## **4.9 Protocol – Mature Trees**

Follow the Primary Protocol & Site Selection Protocol to pick a site, set up a transect, and randomly select focal trees or quadrats. Many tree species cover huge geographic areas, making it unreasonable to survey an entire “population” or define a discrete study “site.” If you are working with a widespread species, it is fine to choose a “representative” study area, which actually might just be a small part of a large stand of trees.

Once you have selected a representative study area, you will need to randomly select 30 trees to survey (plus their nearest conspecific neighbors). There are many ways to do this. Here are three methods in somewhat decreasing order of amount of work and rigor:

Follow the Primary Protocol exactly, establishing a transect, selecting 30 points randomly (distance along main transect and distance from main transect), and using a circular quadrat at each point to randomly select 1 individual of the tree species within the quadrat (as in the Primary Protocol). Survey each selected tree and its nearest neighbor.

Follow the Primary Protocol except skip the circular quadrat step, which could need to be prohibitively large in some tree populations: Establish a transect and randomly select 30 points (distance along main transect and distance from main transect). Then select and survey the individual nearest to each random point (plus nearest neighbor).

For trees that are at low density or low abundance, please consult our Low Density Protocol to select trees. For low-abundance plants, we recommend surveying every plant within some area. Take GPS coordinates for each plant. Try to get as close to 60 plants as possible. If you are taking GPS coordinates for each plant, then you do not need to measure distances to nearest neighbors because we can measure spatial relationships using the GPS data.

However you select trees, please make sure to take detailed notes on what you did

Note that some of the trees you select may be seedlings or saplings. We recommend doing whichever individual you randomly select, regardless of its age. This should yield a representative sample of all individuals at the site, across age classes

Randomly select 30 leaves for quantitative estimates of percent herbivory on each of the 30 leaves See Primary Protocol and Damage Estimation Training Document for guidelines on quantifying percent herbivory per leaf

Please also record the number of galls, mines, and other discrete damage types. Note that mines should be included both in percent damage (because they represent damaged surface area) and as counts.

Randomly select an additional 30 leaves to score for presence/absence of herbivory. Record the number out of 30 with herbivory.

Do not worry about estimating herbivory at the whole-plant scale for mature trees; we will estimate this using the 30 presence/absence leaves and the 30 percent herbivory leaves

## 4.10 Selecting Leaves for Herbivory Estimates

Trees... are tall, and we will not be able to reach top branches. We will therefore focus on low branches that can be reached from the ground with a pole pruner, and sample from multiple places around the circumference (see below). We provide some guidelines below, but you should choose an approach that makes sense for you and your species. Remember that we

are trying to acquire a random subsample of all leaves on the tree; this means avoiding any preference for/against particular leaves (e.g., young vs old).



Figure 4.5: Focus on low branches that can be reached from the ground with a pole pruner, and sample from multiple places around the circumference

Ideally, leaves will be sampled in proportion to their frequency on the tree. Here are two alternate methods for selecting random (or at least haphazard) leaves

**The easiest method:** if it would work for your trees, is to close your eyes, point at the tree, open your eyes, and take the leaf you were pointing at (“Ian’s nose pointing method” in the Damage Estimation Training Document).

**Perhaps the most rigorous but most time-consuming method:** is to haphazardly strip several times as many leaves as you need (e.g., >200 leaves). Place leaves individually

into a large bag. Mix them. Close your eyes and draw 30 leaves for percent herbivory and 30 leaves for presence/absence of herbivory.

Originally

pub-

lished:

Up-

dated:

author:

Yu-

val

Sapir

(Tel

Aviv

Uni-

ver-

sity)

bibli-

ogra-

phy:

..../Her-

b-

Var.bib

Clonal plants present an interesting challenge and opportunity within the HerbVar Network. F

\*\*Objectives:\*\* Provide a protocol for surveying herbivory on a rhizomatous plant species tha

## Background

In semi-arid and arid climates, a considerably large number of plant species are rhizomatous

## Protocol - Rhizomatous Geophytes

1. When first starting this for a new species or at a new site, we suggest spending time investigating the area to get a sense of the distribution and density of genets.
2. Follow the Primary Protocol & Site Selection Protocol to pick a site, set up a transect, and mark the quadrats.

### Calculate a custom radius for circular quadrats

Estimate mean density of genets per square meter by counting the number of plants in 1 m<sup>2</sup> at each quadrat.

- If genet area (clone/genet diameter) is >1 m and/or distances between genets are apparently greater than 1 m, randomly choose 1 of the genets to survey and record the data.
- If a quadrat has >0 focal plants, randomly choose 1 of the genets to survey and record the data.

### Genet life stage: seedling, vegetative, reproductive

Genet size, measured as the height of the tallest leaf for plants in vegetative stage, or height of the tallest flower spike for plants in reproductive stage.

### Herbivore damage in one of 3 ways:

1. Total number of leaf fans (ramets)

- For genets with >100 ramets, write "100" and make a note that your estimate was capped at 100. Estimated percent damage across the whole genet. Visually scan all the green areas of all ramets.

- If the plant has <10 ramets, sample all ramets. From each chosen ramet, pick the 2nd or 3rd largest leaf fan.

Note that tip of the leaf may be dry due to climate fluctuations in the arid regions. This applies to both vegetative and reproductive stages.

2. If a quadrat has 0 focal plants, record a 0 and move to the next quadrat

3. Record the same data for the first nearest conspecific neighbor (of a different genet) that is within the quadrat.

4. Continue visiting randomly select points until 30 focal genets and 30 nearest neighbor genets have been recorded.

! [Top left: Dense ("phalanx type") genet of *Iris atrofusca*; Top right: Sparse ("guerrilla type") genet of *Iris atrofusca*]

<!-- ### References -->

```

<!-- Harper, J. L. 1977. Population Biology of Plants. Academic Press, London. -->
<!-- Herben, T., and J. Klimešová. 2020. Evolution of clonal growth forms in angiosperms. Ne...
<!-- Not cited in document, EMB inserted where looked relevant -->
<!-- Sapir, Y., and A. Shmida. 2002. Species concepts and ecogeographical divergence of Oncoc...
<!-- Vallejo-Marin, M., M. E. Dorken, and S. C. H. Barrett. 2010. The Ecological and Evolution...
<!-- Not cited in document, EMB inserted where looked relevant -->
<!-- Wilson, C. A., J. Padiernos, and Y. Sapir. 2016. The royal irises (Iris subg. Iris sect...

**Originally published: **
**Updated: **

`<!-- quarto-file-metadata: eyJyZXNvdXJjZURpcii6InByb3RvY29scyJ9 -->`{=html}

```{=html}
<!-- quarto-file-metadata: eyJyZXNvdXJjZURpcii6InByb3RvY29scyIsImJvb2tJdGVtVHlwZSI6ImNoYXB0Z...
```

Herbivores, Mines, and Galls {#sec-herbivore}

Overview

Though a lower priority than the damage data, these data will permit us to pilot some more m...
::: {.callout-important}

For all plants, record the number of leaf mines and galls on the entire plant. If there are ...
Separate from counting mines and galls, please also collect insect herbivore data if you are ...
:::

Protocol

Deciding whether to sample "core" herbivores.**

Please use the following questions to help you decide whether to sample "core" herbivores.

```

1. \*\*Are you comfortable distinguishing the following 5 groups of herbivores? \*\*If not, prior

- Grasshoppers/crickets/katydid (Orthoptera).
- Caterpillar-like larvae (i.e., eruciform larvae). \*\*\_Note\_: \*\*this includes moth/butter
- Aphids (Aphididae)
- Hoppers (Hemiptera: Auchenorrhyncha). This includes planthoppers (Fulgoromorpha), leaf
- Non-Aphid Sternorrhynchans. This includes whiteflies, scale insects, and mealybugs

2. \*\*Are you confident that you can visually detect the herbivores on the selected plant species but be judicious of the added time required for sorting through a loaded beat-sheet!

3. \*\*Could you do another herbivory survey with the time required to conduct an herbivore sur

### Sampling insects beyond "core" herbivores.

1. In an effort to standardize the insect data we have included 5 groupings to use for tallying

2. Please indicate whether you are recording herbivores as a count or as presence/absence data

3. For both core and non-core insects, please herbivorous count insects whether or not they are

4. If you have more intimate knowledge of insect herbivores (e.g., can distinguish herbivorous

- To facilitate this, we have added "beetleHerbivore", "thysanopteraHerbivore", "gastropodHer

- We also recognize that herbivore surveys may differ dramatically among sampling sites and

- Please continue to record the 5 required insects (even when there are none please put a

- While mines/galls are recorded in the Primary Protocol, mine-/gall-forming insects should

## Herbivore Guide

### Mines & Galls Visual Guide

! [Leaf mines - Linear/Serpentine. Left/right = single mine. Center = multiple mines] (images/leaf\_mines.jpg)

How do you count multiple mines? It's a confusing picture but in that way is more likely some

This can get confusing, but do your best. Each count doesn't have to be exactly right; we sh

### ### Leaf Mines - Blotch Mines

! [Several examples of multiple blotch mines on single leaves] (images/img20.png){#fig-mines2}

### ### Galls - Leaf Galls

! [Leaf Galls] (images/img21.png){#fig-galls1}

### ### Galls - Stem/Branch Galls

! [Stem/Branch Galls] (images/img22.png){#fig-galls2}

## ## Insect Herbivore ID Visual Guide

Some groups of insects (e.g. Hemipterans, Coleopterans) include predatory, herbivorous, and c

### ### Grasshoppers / crickets / katydids (Orthoptera)

! [] (images/img23.png){#fig-herbivores1}

<!-- Grasshoppers (suborder: Caelifera); crickets/katydid (suborder: Ensifera) -->

<!-- May be confused with: mantids (left) or hoppers (right) -->

<!-- Synapomorphies & Identifying Marks: Have swollen "knee" joints of hind legs (femoro-tib

### ### Caterpillar-like (larval forms ONLY)

! [] (images/img24.png){#fig-herbivores2}

<!-- Lepidopteran caterpillars; sawfly larvae; herbivorous beetle larvae -->

<!-- May be confused with: syrphid larvae (left) or coccinellid larvae (right) -->

```

<!-- Synapomorphies & Identifying Marks: Have cylindrical bodies with a well-developed head -->

Hoppers (Hemiptera: Auchenorrhyncha)

! [] (images/img25.png){#fig-herbivores3}
<!-- Leafhoppers, treehoppers, planthoppers, and cicadas -->

<!-- May be confused with: true bugs (i.e., Hemiptera: Heteroptera) -->

<!-- Synapomorphies & Identifying Marks: Have piercing/sucking mouthparts with "beak" seeming -->

Aphids (Hemiptera: Aphididae)

! [] (images/img26.png){#fig-herbivores4}

<!-- May be confused with: whitefly -->

<!-- Synapomorphies & Identifying Marks: Have a pear-shaped body and may/may not have wings. -->

Non-Aphid Sternorrhynchans (whiteflies, mealybugs, scale insects)

! [] (images/img27.png){#fig-herbivores5}

<!-- Whitefly (Aleyrodidae); mealybug (Pseudococcidae); scale insect (Coccoidea) -->

<!-- May be confused with: aphids -->

<!-- Synapomorphies & Identifying Marks: Relatively distant relatedness among these three taxa -->

**Originally published: **
**Updated: **

::: {.quarto-book-part}

```

```
`<!-- quarto-file-metadata: eyJyZXNvdXJjZURpcii6Ii4ifQ== -->`{=html}

```{=html}
<!-- quarto-file-metadata: eyJyZXNvdXJjZURpcii6Ii4iLCJib29rSXRlbVR5cGUiOiJwYXJ0IiwiYm9va010Z
```

Admin Tasks

:::

`<!-- quarto-file-metadata: eyJyZXNvdXJjZURpcii6ImFkbWluX3Rhc2tzIn0= -->`{=html}

```{=html}
<!-- quarto-file-metadata: eyJyZXNvdXJjZURpcii6ImFkbWluX3Rhc2tzIiwiYm9va010ZW1UeXB1IjoiY2hhch
```

Potential Collaborators

Potential collaborators typically get email either Will or the HerbVar Admin gmail address.

1. Reply with the first email template (of 2) for prospective new members. Be sure to cc Will.

2. They respond and say:
 a) they're not interested/have issues with the expectations.
 - Send them over to Will and stay on top of the email chain between them so you can know how they respond.
 b) They say that sounds fine and they're still interested:
 i. Add them to the Collaborator Contact Information file
 ii. Add them as a "Contributor" to the HerbVar Shared Drive
 iii. Respond with the second email template (of 2)
```

It is important you send the email after doing the steps i and ii because the email template contains an onboarding document that you may need to reference.

```

`<!-- quarto-file-metadata: eyJyZXNvdXJjZURpcii6ImFkbWluX3Rhc2tzIn0= -->`{=html}

```{=html}
<!-- quarto-file-metadata: eyJyZXNvdXJjZURpcii6ImFkbWluX3Rhc2tzIiwiYm9va0l0ZW1UeXB1IjoiY2hhc
```

Editing the Website

`-----{.quarto-title-block template='/Applications/quarto/share/projects/book/pandoc/title.html'
---author: Nick Lyon

```

## 4.11 Github Access

The [HerbVar website](#) has been created in R. Ask to be added as a collaborator on [this repository](#).

Fork the [website's repository](#) to your computer. Nick Lyon initially forked the repository so that pushes would be preserved as “pull requests” and could be reviewed by Will before actually changing the website on the internet, but this may be unnecessary depending on your comfort with this type of coding.

## 4.12 Making Edits

Make whatever changes are asked for or required.

Each page of the website is saved as a separate .Rmd file and file names mostly correspond to website tab names so it should be relatively easy to identify which script(s) needs to be changed and implement those edits.

The existing scripts also include plenty of examples of heading formatting, font changes, and hyperlinks so use the existing pages to teach yourself how to do things you don’t already know how to do.

## 4.13 Rebuilding the Website

Once you've made the edits, go to the "Build" tab on R Studio and click "Build Website". This will take several minutes to process (there will be a running list of code as it processes through each .Rmd file) so feel free to grab a cup of coffee as this processes.

Once it completes, it will create a new tab and will pop up the new website in your browser **but you are not yet done!**

Once that is done, in the Git tab of R Studio, select all modified files (not just the scripts!) and commit/push them all.

Building the website may affect a lot of files in the "libs" folder deep in your system (you can tell how savvy I am about this, huh?) and these changes must also be included in the push for the website to successfully update.

Once you've pushed these changes (and if you're working in a fork, Will has accepted your pull request) the website on the internet should update within 10-15 minutes so double check your work after roughly that amount of time has passed.

# 5 The Data Portal

The data portal ([link](#)) is the preferred method for data submission for (at least phase 2). It is written in R Shiny and is built for the phase 2 template Excel file but will work (with some warnings) for the phase 1 Excel template. This is what you need to know to change and/or troubleshoot the app.

## 5.1 Your Job After Someone Submits Data via the Portal

The data portal puts submitted data in the “App Uploads - Phase 2” folder ([link](#)). The portal would be self-sufficient but I have added a step to require human involvement that I’ll describe here.

1. You need to move all files from that folder to the “Phase II Raw Data” folder ([link](#))
2. All phase 2 wrangling scripts will download raw data from that latter folder
  - a. I’ve set up all of the wrangling scripts to download raw data in an `if() else{}` framework that will print a message reminding you to move the data out of the app upload folder if you ever forget to/don’t see that new data have been uploaded
3. That’s it! The data wrangling scripts will work without issue now that you’ve moved the files to the correct folder

## 5.2 Updating the Portal

It may become necessary to edit the portal, especially if a user emails you indicating they had a problem and it seems like that problem is inside of the app rather than (not to be mean) user error.

1. All of the portal code is in the “Data-Portal” GitHub repository ([link](#))
  - a. The script from which the portal is created is called “app.R” and is the only file in the folder “Data Portal Actual”

- b. You will also need the “deployment-faq.R” script in the “Support Scripts” folder in order to deploy the app after you have made/tested any changes to the portal’s code. i. If it is of interest, the “Support Scripts” folder also includes my (Nick’s) incremental forays into the world of Shiny so you can see the first through eighth versions of the portal before getting to a version that was deployed.
- 2. Before changing the portal I strongly recommend asking the user who pointed out the issue for a screenshot of how they’ve filled the app out immediately before the error
  - a. The error is almost always (or at least has usually been) something to do with how the user filled out the app or attached their data. If that is the case, you may need only point that out to them (in a polite way) and go about your day
  - b. Also, the only time the app can break is when they click the “Submit Data” button. Prior to that, the app is not actually trying to do anything, so any app-breaking user error will not be apparent to them until they click that button
  - c. HOWEVER, some users who experience an issue actually create a larger error that will prevent them from uploading their data even after you point the app key issue out to them. To handle such cases, see the next subsection. The importance of this is also noted in part d of the next bullet 3. To change the data portal do the following:
    - 1. First, modify the app. R script as desired. Note that every Shiny App consists of three components: (1) the user interface, (2) the server that includes all the internal mechanisms for the portal, and (3) a ShinyApp() call that combines the UI and server.
      - i. If the app is not working, it will likely be in the server component
      - ii. If the app doesn’t look right but does function appropriately, modify the UI
      - iii. If the app does not collect some information that it should, you will need to change both the UI and server
      - iv. The ShinyApp call at the end never needs to be modified so don’t worry about that bit
    - 2. Second, test the app on your computer by running the app
      - i. In R Studio, the top right of the R script panel containing a ShinyApp has a “Run App” button to the right of a green ‘play’ button
      - ii. Pushing this button will create a local version of the portal that functions as the app will but does not deploy to the internet (yet).

- iii. I recommend submitting a test data file (see the folder of the same name for pre-built phase 2 data that you can use) from start to finish to ensure that everything works as desired.
- 3. Third, once you are satisfied with your changes, you can deploy the app to replace the old publicly-available version on the internet!
  - i. In the “deployment-faq.R” script, you will load the “rsconnect” library (line 12) and then use it to redeploy the app (line 18)
  - ii. Running the deployApp function will prompt you in the console to type a “Y” if you’re sure that you want to re-deploy the app
  - iii. After you type “Y” and hit return in the console, it will build your new portal, terminate the old one, replace the old with the new, and then activate the new one for all users
  - iv. You’ll know this is done when R automatically kicks you to a new tab in your web browser with the new portal open
- 3. Fourth, and this is crucial, if your changes to the app were because a user was having issues, you need to delete any files they successfully submitted
  - i. See the next sub-section for information on how to do this/why it needs to be done
- 4. Finally, notify the user that initially contacted you letting them know that you have resolved the issue on your end, thanking them for bringing it to your attention, and inviting them to reach out again if it still isn’t working for them

### **5.3 Deleting Old Data**

The data portal will fail if it tries to create two files of the same name.

- 1. The amount of information used to create the file name means that it is incredibly unlikely that two different users could accidentally create the same file name
- 2. BUT, as mentioned in the “Updating the Portal” section, it is entirely possible (and has happened previously) for the same user to try to submit data more than once and inadvertently create two files of the same name
  - a. This occurs when the following happens:
  - b. First, the user tries to submit data using the portal but something goes wrong so their data aren’t actually submitted (but a blank GoogleSheet of the user-supplied name is created)

- ii. Second, the user tries to re-submit data (possibly after you fix the issue in the portal and notify them) but the blank document they unknowingly created earlier now causes a different error (i. e. , that there are now two files of the same name)
  - iii. Unfortunately, because Shiny Apps are noninteractive (see the “Service Account FAQ” section) the user will never be provided with an informative error (neither will you) so you’ll need to diagnose this as part of your ‘fixing the app’ process
- 3. To resolve this, I (Nick) have created a second Shiny App that is not deployed
  - a. To be clear, it should never be deployed to prevent its accidental (mis)use by general HerbVar members
- 4. Justification and location of the second app
  - a. The second app is in the “Data-Portal-Maintenance” GitHub repository ([link](#))
  - b. The “Service Account FAQ” section below gives more context but in brief: the ‘app key’ file that the data portal makes users attach is actually activating a sort of Google robot with the authority to create Google Sheets and move them i. This is necessary because an online portal cannot send an authorization request to each user in the way that R/R Studio does when such code is run on a local computer (again, see the “Service Account FAQ” section for more details)
  - c. This ‘robot’ then is the true owner of all data files submitted through the app
  - d. The portal cannot submit data to a Shared Drive (due to issues with the R packages that connect R and Google that are outside of our control) so this is an unavoidable state
  - e. So, if a user accidentally creates a flawed data object of the same name as their real data they will be unable to submit their real data until the flawed one is deleted
  - f. HOWEVER, because the ‘robot’ owns those files, you cannot actually delete any of its files (when you “delete” a file you don’t own you actually just remove yourself as a collaborator with no effect on the original file)
  - g. Here is where we get to the need of a second app
  - h. The robot’s GoogleDrive cannot be accessed via a Graphical User Interface in the way that you would access any other Google Drive
  - ii. So, to truly delete these files so a user can re-submit their data successfully, you will need to use this second app
  - iii. If you fail to delete the bad data, the user will never be able to successfully submit data of the same name 1. In theory, you could ask them to re-name their file in some slightly different way (i. e. , by changing their site name), but that would still have this flawed data floating in the ether which is not desirable

5. Tutorial of the second app
    - a. To reiterate, this Shiny app should never be deployed.
    - b. You will see why, but for the moment, take my word on it that deploying this app has a non-zero potential of permanently deleting data files you actually want
  - ii. By keeping the script in GitHub and locking view access to only Will & the Data Scientist, we preserve its utility without opening Pandora's box of deploying it and possibly having an HerbVar member use it improperly
  - b. The second app is fully contained in the “check-service-acct-files.R” script (the only script in this project)
  - c. Open that script and click the “Run App” button in the top right of the R script pane of R Studio
  - d. As with updating the data submission portal, this will create a new tab in your web browser that contains a fully functional (but not available on the internet) version of the app
  - e. The app is divided into three columns that you will proceed through from left to right
  - f. First, download and attach the key for the service account that owns the files you want to look through (column 1)
  - g. For now we only have one service account for phase
6. I recommend creating a new service account for each subsequent phase to evade data storage limits and partition sources of error in a clean, behind-the-scenes sort of way
    - ii. See below for information on creating Service Accounts
    - f. Second, click the “Authorize” button to notify the app that it should attach the app key (column 1)
    - g. This may take a few seconds but should generate a full list of all files owned by the robot (i. e. , owned by the service account) g. Third, after looking at the list of files, click the “Extract File Names” button (column 2)
    - h. This just populates the third column so don't worry about the violence implied by the verb ‘extract’
    - i. Fourth, scroll through the drop down list (column 3) and select the file you want to delete (could be a test data file or the product of a specific user's failed attempt to upload their data)

- j. Fifth, above the dropdown menu, check the “Yes” option beneath “I am ready to delete a file”
- k. I recommend doing this after selecting a file to further mitigate the risk of deleting the wrong file j. Sixth, click the “Delete Selected File” button i. Because you attached the service account key in column 1, you are viewing and interacting with the robot’s files as the robot (rather than as yourself)
- ii. This gives you access to actually delete files rather than just—as mentioned before—removing yourself from seeing the file
- k. Seventh, once a dialogue has popped up below the “Delete Selected File” button confirming the file has been deleted, click the “Update List of Drive Contents” button
- l. This will update the dropdown menu with the new file list now that the file you marked for deletion has been erased l. Finally, scroll through the dropdown menu (or look at the list of files in column 2) to ensure that all problem files have been deleted 6. After you’ve gone through that process to delete the flawed file, you can notify the user that it is safe for them to resubmit their data
- m. This is all likely too much information for the user though so I suggest that you just tell them you have fixed the data portal and leave it at that
- 7. Also, I have written the app to work with any service account key that owns Google files so unless the structure of future phases’ data portals changes massively, this app should be sufficient for all issues involving service account-owned files in the future

## **5.4 Service Account FAQ Background Information**

The data submission portal accepts uploaded data locally and then (1) creates a Google Sheet version of the data and (2) moves that sheet into the designated folder in the HerbVar Admin Drive. However, the Shiny app is “non-interactive” (see *gargle*’s vignette) which means that a user cannot input a gmail or access token to tell Google Drive/Sheets who is creating/moving files. A “service account” is necessary to get around this.

A service account is essentially a robot that we pre-approve to (1) create google sheets, (2) move files, and (3) have access to the folder(s) we want those sheets made in/moved to. Side note: see the list of people with access to the folder the Shiny portal saves files to and you’ll see the service account I created in that list.

To create/manage a service account you need to use “Google Cloud Platform” as described below:

## 5.5 Tutorial

1. Sign into the herbvar@gmail.com Google Account
2. Visit the Google Cloud Platform ([link](#))
  - a. If there is a pale red/pink error saying you don't have sufficient permissions to view the page, select the herbvar@gmail.com account from the drop down in the top right of the screen
  - b. The page should then re-load to the dashboard
3. Don't get overwhelmed by the level of detail on this page!
4. In the left sidebar, click "APIs & Services" and within that menu click "Credentials"
5. Click the service account name in the "Service Accounts" list at the bottom of the screen
6. Keys can be managed in the "Keys" tab
7. In the event of a security breach (not sure what that would look like but still good to have the contingency), delete the existing keys and create a new one
8. Download that key and replace the one HerbVar members have access to with the new key file.
9. If creating a new key or service account prompts you to add permissions to the account be sure that it includes BOTH the GoogleSheets API AND the GoogleDrive API
  - a. Both are needed because the data portal uses the service account key to both create a GoogleSheet (using the eponymous API) and move that GoogleSheet (using the GoogleDrive API)

# 6 Google Drive Structure

This project has a lot of files coming and going so this is a brief description of all of those folders.

## 6.1 Phase I Data Wrangling

- HerbVar Phase I Data
- All Uploads
- All of the phase 1 raw data (and we've since moved to phase 2 so there should not be any more new data)
- Phase I
- Herbivore Data
- herbivoreData (from the eponymous Excel sheet) raw, tidied at plant-level (one row per plant) and tidied at survey level (one row per survey)
- Raw is the herbivore columns sliced off of the plantData sheet as part of the main wrangling script
- Phase I - Primary Productivity
- Primary productivity metadata to go along with phase 1 data
- Phase I - Reproductive Data
- Raw reproductive data columns sliced off of the plantData sheet as part of the main wrangling script
- Phase I - Richness Data
- Extracted native/invasive species richness from Ellis et al. 2012
- Phase I - siteData - Tidied version of siteData (from eponymous Excel sheet)

- Phase I - Survey Indices - Attempt (later abandoned) to split off just the site and plant-identifying information without any of the “actual” data. This would allow wide sharing of location information to enable all interested HerbVar members to harvest publicly-available metadata
- Phase I Data - Wrangled
- The “abiotic” (i.e., climatic) data for each site’s lat/long coordinates
- The tidied plant-level data (one row per plant) for all phase 1 data
- The tidied survey-level data (one row per survey) for all phase 1 data
- HerbVar Foliage Index.xlsx - Index describing whether each species (from each PI) is deciduous, evergreen, or annual (this information was integrated into the phase 2 data submission portal so no equivalent Google Sheet will exist for phase 2

## 6.2 Phase II Data Wrangling

- \_Data Submitted Via Email - I need to put it through the App - This is a clearinghouse for all raw phase 2 data that users send via email rather than using the submission portal. While discouraged, we don’t want to completely block data in such instances. So, once you receive data via email, drop it into this folder until you have the time to run it through the app (the handful of times this has occurred the data went through the portal fine, users just got frustrated from unrelated things)
- App Uploads - Phase 2 - Raw data submitted via the portal arrive here

## 6.3 Phase II Completed Surveys Versions

- All phase 2 wrangling scripts copy the completed surveys file with a time stamp to this folder for posterity. I can think of no direct utility of these backups but it doesn’t hurt to save them -

## 6.4 Phase II Metadata

- Rather than have separate folders for each metadata type (as was the case in phase 1) I have created this folder to contain them. Only abiotic (i.e., climatic) data have been retrieved so far but all should be placed here to keep the Drive folder hierarchy clean

## **6.5 Phase II Raw Data**

- All data submitted through the submission portal should be manually moved from the “App Uploads - Phase 2” folder to here. The wrangling scripts will prompt you to do this if you do not before running them.

## **6.6 Phase II Wrangled Data - ...**

- Each sheet of the Excel file has its own version of the above folder where the ellipses (...) is replaced by that sheet’s name. Where applicable (e.g., plantData, reproData, herbivoreData) there are survey-level (i.e., one row per survey) versions of the data

## **6.7 Miscellaneous Other Files**

*Note: this heading is not a folder name but refers to the random other files in the Drive.*

- This manual is unfiled in the Drive!
- The Data Management Plan (DMP) in graphical form
- Tutorials for one-off tasks you may need to explain to others

# 7 Wrangling Repository

The [Wrangling Repository](#) contains scripts for data wrangling for all phases of the project. It takes in raw data and outputs analysis/visualization-ready .csv files.

It will be the primary home for this Research/Admin Position (or at least it was for me) so it may help to give you a brief explanation of each of the main scripts. Here's the link.

## 7.1 Misc. Non-Manuscript Subset Scripts

So far, this only includes the script to separate out PlantPopNet members' data to make sharing that with PPN leadership (upon request) simpler

## 7.2 Phase 1 Scripts

`phase 1 abiotic wrangling.R`: Wrangles WORLDCLIM climatic data for phase 1 surveys.

`phase 1 herbivoreData wrangling.R`: Wrangles any information to do with herbivores from phase 1.

`phase 1 plant richness wrangling.R`: Extracts interpolated native/invasive species richness information from Ellis et al. 2012 shapefiles.

`phase 1 primary productivity.R`: Extracts primary productivity data from satellite data for phase 1 surveys.

`phase 1 shareable index.R`: Creates files of only location information for phase 1 sites to enable other HerbVar collaborators to harvest metadata without sharing “actual” data.

`phase 1 siteData wrangling.R`: Wrangles information from siteData sheet of template Excel file.

`phase 1 soil data wrangling.R`: Placeholder describing where soil data may someday be acquired. For now, the relevant R package does not work (though their team is aware of and working on this issue).

`phase 1 survey-lvl summarizing.R`: Summarizes the tidy plant-level (one row per plant) phase 1 data to survey-level (one row per survey).

`phase 1 wrangling.R`: Takes all the separate phase 1 raw data files combines and wrangles them to plant-level (one row per plant).

## 7.3 Phase 2 Scripts

`phase 2 densityData wrangling.R`: Wrangles eponymous sheet from template Excel file.

`phase 2 herbivoreData wrangling.R`: Wrangles eponymous sheet from template Excel file (at both plant-level and survey-level).

`phase 2 metadataabiotic.R`: Extracts WORLDCLIM climatic data from phase 2 site locations (requires tidy file from siteData wrangling script).

`phase 2 newColumns wrangling.R`: Wrangles eponymous sheet from template Excel file.

`phase 2 notes wrangling.R`: Wrangles eponymous sheet from template Excel file. `phase 2 plantData survey-lvl summarizing.R`: Summarizes tidy plantData to survey-level (i.e., one row per survey).

`phase 2 plantData wrangling.R`: Wrangles eponymous sheet from template Excel file (at ONLY plant-level).

`phase 2 reproData wrangling.R`: Wrangles eponymous sheet from template Excel file at plant-level only (survey level absent because insufficient raw data at this point).

`phase 2 siteData wrangling.R`: Wrangles eponymous sheet from template Excel file.

## 7.4 Script Archive

All “actual” scripts (i.e., those used in day-to-day wrangling) should have a consistent aesthetic and comment structure (as well as being primarily tidyverse-based). When others contribute code, duplicate the file and edit one version to match internal standards. The second version goes here to be preserved in its original form as a back-up

## 7.5 Singleton Tasks

Any scripts written to accomplish a ‘one-off’ task I thought unlikely to be repeated regularly are placed here. Some of them may include operations that could be useful in other contexts though!

## **7.6 Manuscript Subsetting Scripts**

Each script is dedicated for a single HerbVar manuscript and does the subsetting and/or column selection necessary to create a tidy data file of only what authors request to test their hypotheses

# **8 To-Do List for Manual**

(as of 2/1/22)

1. Edit the “Google Drive” chapter
2. Convert “data portal troubleshooting” to a flowchart/checklist
3. Edit Protocols:
  - a. Insert Photos
  - b. link to pdf versions for downloading by users
  - c. edit / correct formatting
4. review and expand the workflow chapter
5. review and expand the data analysis chapter; include tutorial for repo
6. review and expand the publications chapter; include tutorial for repo
7. References/multiple .bib files
8. herbvar publications, herbvar presentations
9. add tutorial for editing the manual and publishing with actions <https://quarto.org/docs/publishing/github-pages.html>

---

Unfortunately I (Nick) took a new position while there were still some loose ends left hanging but I can list them here so they don't completely fall between the cracks -

## **8.1 Data Use Agreement**

I wrote a data use contract for HerbVar members to sign upon joining the Network so that we can have them release rights to their data to us (and promise not to share the larger datafiles produced by the collaboration - The email conversation for getting the Planning Group's feedback and green light has the subject line “HerbVar Data Sharing Agreement Draft” - The relevant Google Drive folder is linked here and lives in the “HerbVar Management” Shared Drive -

## **8.2 Authorship Guidelines Conversation**

The Planning Group was in the midst of deciding on revised authorship guidelines to catch some edge cases that fell outside of the earlier authorship policy. - You can catch up on this in the email thread with the subject “HerbVar Authorship Criteria Questions”

## **8.3 Herbivore Protocol Gray Area**

The Herbivore protocol is written to prioritize taxonomic identification of insects. However, many researchers are also (and sometimes more) interested in functional identification of insects. Our current approach to the template Excel file allows people to add their own columns as they see fit but this makes later switching from functional to taxonomic (or vice versa) difficult. So, there is a draft email in the herbvar@gmail account to send when a panel of herbivore experts has been assembled to resolve this uncertainty

# References

A .bib file of all HerbVar publications is available for download [here](#).

- Boland, M. R., K. J. Karczewski, and N. P. Tatonetti. 2017. [Ten Simple Rules to Enable Multi-site Collaborations through Data Sharing](#). PLOS Computational Biology 13:e1005278.
- Braga, P. H. P., K. Hébert, E. J. Hudgins, E. R. Scott, B. P. M. Edwards, L. L. Sánchez Reyes, M. J. Grainger, V. Foroughirad, F. Hillemann, A. D. Binley, C. B. Brookson, K. M. Gaynor, S. Shafiei Sabet, A. Güncan, H. Weierbach, D. G. E. Gomes, and R. Crystal-Ornelas. 2023. [Not just for programmers: How GitHub can accelerate collaborative and reproducible research in ecology and evolution](#). Methods in Ecology and Evolution 14:1364–1380.
- Champieux, R., A. Solomonides, M. Conte, S. Rojevsky, J. Phuong, D. A. Dorr, E. Zampino, A. Wilcox, M. B. Carson, and K. Holmes. 2023. [Ten simple rules for organizations to support research data sharing](#). PLOS Computational Biology 19:e1011136.
- Goodman, A., A. Pepe, A. W. Blocker, C. L. Borgman, K. Cranmer, M. Crosas, R. D. Stefano, Y. Gil, P. Groth, M. Hedstrom, D. W. Hogg, V. Kashyap, A. Mahabal, A. Siemiginowska, and A. Slavkovic. 2014. [Ten Simple Rules for the Care and Feeding of Scientific Data](#). PLOS Computational Biology 10:e1003542.
- Kim, A. Y., V. Herrmann, R. Barreto, B. Calkins, E. Gonzalez-Akre, D. J. Johnson, J. A. Jordan, L. Magee, I. R. McGregor, N. Montero, K. Novak, T. Rogers, J. Shue, and K. J. Anderson-Teixeira. 2022. [Implementing GitHub Actions continuous integration to reduce error rates in ecological data collection](#). Methods in Ecology and Evolution 13:2572–2585.
- Yenni, G. M., E. M. Christensen, E. K. Bledsoe, S. R. Supp, R. M. Diaz, E. P. White, and S. K. M. Ernest. 2019. [Developing a modern data workflow for regularly updated data](#). PLOS Biology 17:e3000125.

# **Part III**

# **Appendix**

# HerbVar Publications

A .bib file of all HerbVar publications is available for download [here](#).

This will eventually have a list of all publications from the HerbVar Project

- Boland, M. R., K. J. Karczewski, and N. P. Tatonetti. 2017. [Ten Simple Rules to Enable Multi-site Collaborations through Data Sharing](#). PLOS Computational Biology 13:e1005278.
- Braga, P. H. P., K. Hébert, E. J. Hudgins, E. R. Scott, B. P. M. Edwards, L. L. Sánchez Reyes, M. J. Grainger, V. Foroughirad, F. Hillemann, A. D. Binley, C. B. Brookson, K. M. Gaynor, S. Shafiei Sabet, A. Güncan, H. Weierbach, D. G. E. Gomes, and R. Crystal-Ornelas. 2023. [Not just for programmers: How GitHub can accelerate collaborative and reproducible research in ecology and evolution](#). Methods in Ecology and Evolution 14:1364–1380.
- Champieux, R., A. Solomonides, M. Conte, S. Rojevsky, J. Phuong, D. A. Dorr, E. Zampino, A. Wilcox, M. B. Carson, and K. Holmes. 2023. [Ten simple rules for organizations to support research data sharing](#). PLOS Computational Biology 19:e1011136.
- Goodman, A., A. Pepe, A. W. Blocker, C. L. Borgman, K. Cranmer, M. Crosas, R. D. Stefano, Y. Gil, P. Groth, M. Hedstrom, D. W. Hogg, V. Kashyap, A. Mahabal, A. Siemiginowska, and A. Slavkovic. 2014. [Ten Simple Rules for the Care and Feeding of Scientific Data](#). PLOS Computational Biology 10:e1003542.
- Kim, A. Y., V. Herrmann, R. Barreto, B. Calkins, E. Gonzalez-Akre, D. J. Johnson, J. A. Jordan, L. Magee, I. R. McGregor, N. Montero, K. Novak, T. Rogers, J. Shue, and K. J. Anderson-Teixeira. 2022. [Implementing GitHub Actions continuous integration to reduce error rates in ecological data collection](#). Methods in Ecology and Evolution 13:2572–2585.
- Yenni, G. M., E. M. Christensen, E. K. Bledsoe, S. R. Supp, R. M. Diaz, E. P. White, and S. K. M. Ernest. 2019. [Developing a modern data workflow for regularly updated data](#). PLOS Biology 17:e3000125.

# HerbVar Presentations

A .bib file of all HerbVar presentations is available for download here.

This will eventually have a list of all presentations (eg talks or posters at meetings) by HerbVar Members.

- Boland, M. R., K. J. Karczewski, and N. P. Tatonetti. 2017. [Ten Simple Rules to Enable Multi-site Collaborations through Data Sharing](#). PLOS Computational Biology 13:e1005278.
- Braga, P. H. P., K. Hébert, E. J. Hudgins, E. R. Scott, B. P. M. Edwards, L. L. Sánchez Reyes, M. J. Grainger, V. Foroughirad, F. Hillemann, A. D. Binley, C. B. Brookson, K. M. Gaynor, S. Shafiei Sabet, A. Güncan, H. Weierbach, D. G. E. Gomes, and R. Crystal-Ornelas. 2023. [Not just for programmers: How GitHub can accelerate collaborative and reproducible research in ecology and evolution](#). Methods in Ecology and Evolution 14:1364–1380.
- Champieux, R., A. Solomonides, M. Conte, S. Rojevsky, J. Phuong, D. A. Dorr, E. Zampino, A. Wilcox, M. B. Carson, and K. Holmes. 2023. [Ten simple rules for organizations to support research data sharing](#). PLOS Computational Biology 19:e1011136.
- Goodman, A., A. Pepe, A. W. Blocker, C. L. Borgman, K. Cranmer, M. Crosas, R. D. Stefano, Y. Gil, P. Groth, M. Hedstrom, D. W. Hogg, V. Kashyap, A. Mahabal, A. Siemiginowska, and A. Slavkovic. 2014. [Ten Simple Rules for the Care and Feeding of Scientific Data](#). PLOS Computational Biology 10:e1003542.
- Kim, A. Y., V. Herrmann, R. Barreto, B. Calkins, E. Gonzalez-Akre, D. J. Johnson, J. A. Jordan, L. Magee, I. R. McGregor, N. Montero, K. Novak, T. Rogers, J. Shue, and K. J. Anderson-Teixeira. 2022. [Implementing GitHub Actions continuous integration to reduce error rates in ecological data collection](#). Methods in Ecology and Evolution 13:2572–2585.
- Yenni, G. M., E. M. Christensen, E. K. Bledsoe, S. R. Supp, R. M. Diaz, E. P. White, and S. K. M. Ernest. 2019. [Developing a modern data workflow for regularly updated data](#). PLOS Biology 17:e3000125.