

1 Protocol to increase accuracy and fidelity of Pollen Meta-genomic
2 Barcoding using Angiosperms353; Case study using Corbiculae
3 loads from wild Bumble Bees

4 Reed Clark Benkendorf^{1,2*}, Paul J. CaraDonna^{1,2,3}, Jeremie B. Fant^{1,2}, Jane E. Ogilive³,
5 Sophie Taddeo^{1,2,4}, Emily J. Woodworth^{1,2}

¹Chicago Botanic Garden, 1000 Lake Cook Road, Glencoe, Illinois 60022, USA

²Plant Biology and Conservation, Northwestern University, Evanston, Illinois 60208, USA

³ Rocky Mountain Biological Laboratory, P.O. Box 519, Crested Butte, Colorado 81224, USA

⁴ The University of California at San Diego, 9500 Gilman Dr, La Jolla, California 92093, USA

5 **Abstract**

6 1) DNA Barcoding has been successful for the rapid identification of species in complex ecological
7 assemblages. However, metabarcoding in the plant kingdom has been limited due to a lack
8 of universal gene regions that works across all taxa, limiting the applications of eDNA and
9 metagenomics in ecology.

10 2) To circumvent these limitations, we propose using more holistic approach which combines a
11 multiple gene approach that incorporates existing plant occurrence databases, species distribution
12 models, and phenology analyses to generate a list of likely candidate species to increase accuracy.
13 Such an effort can help guide plant sample collection and library construction, particularly for
14 projects with limited resources or local taxonomic knowledge. In addition, we demonstrate that
15 building a custom candidate plant list based on known occurrences, ecological requirements, and
16 phenology can improve computing efficiency, speed, and accuracy of metabarcoding. We tested this
17 approach to identify the plant species contributing to the corbiculae pollen loads of bumble bees.
18 To validate the efficiency of this approach we compare results of the DNA barcoding to expert-led
19 pollen identification and field observations of plant-pollinator interactions that generated these

*Correspondence: reedbenkendorf2021@u.northwestern.edu

21 pollen loads.

22

23 3) We show that the Angiosperms 353 probes, developed for phylogenomics, and which are currently
24 being used in the largest ever plant systematic endeavor, offer significant promise to metagenomic
25 approaches around the globe. The DNA barcoding of bumble bee corbiculae pollen loads was
26 most accurate when combined with knowledge of what plant species were flowering in the plant
27 community when they were collected.

28

29 4) Supplementing DNA barcoding data with ecological context is most accurate and powerful.

30 1 | INTRODUCTION

31 Large scale species loss (Joppa *et al.* (2011); ‘The biodiversity of species and their rates of extinction,
32 distribution, and protection’ (2014)) and biotic homogenization (Olden *et al.* (2004), Wang *et al.* (2021)),
33 and the impacts of these processes on ecosystem functions and human well-being (Cardinale *et al.* (2012)),
34 have inspired numerous calls for a more consistent monitoring of ecosystems and their diversity (Pereira
35 & Cooper (2006); Pecl *et al.* (2017)). Monitoring ecological communities structure and interaction will be
36 critical to informing and prioritizing conservation efforts (REF), particularly considering the large proportion
37 of known species currently threatened or likely to be threatened in the near future (‘The biodiversity of species
38 and their rates of extinction, distribution, and protection’ (2014); Pereira *et al.* (2010)). However, ecosystem
39 monitoring remains inconsistent in time and space (Yoccoz *et al.* (2001)) as it requires substantial resources,
40 thereby leaving several regions, ecosystems, and even species under-observed (Pereira *et al.* (2012); Collen
41 *et al.* (2008); Meyer *et al.* (2015); Ruete (2015)).

42 Species interactions are a critical component of ecosystem stability and therefore biodiversity conservation
43 (Soltis *et al.* (2019), Futuyma & Agrawal (2009), Voje *et al.* (2015), Weber *et al.* (2017), Hembry & We-
44 ber (2020)). Species interactions, which span from positive (e.g., symbiosis, mutualistic) to negative (e.g.,
45 predation, competition), shape species distribution (Bascompte (2009); Wisz *et al.* 2013) and tolerance
46 to environmental stressors (REF). They impact ecosystem stability and diversity (Agrawal *et al.* (2007),
47 Valiente-Banuet *et al.* (2015), Bascompte *et al.* (2006), and the relationships between biodiversity and
48 ecosystem functions (Duffy *et al.* (2007)). However, species interactions are challenging to observe (REFs).
49 They are typically measured through shorter-term observational studies yet are highly variable in time and
50 space (Liu & Gaines (2022)). Studying species interactions also requires extensive taxonomic expertise,
51 especially those from diverse clades (Hebert *et al.* (2003)), further challenging their frequent assessment.

52 As a result, few ecological networks have been fully mapped (REFs). Improving our capacity to efficiently
53 study species interactions consistently across several ecosystems, landscapes, and management context is
54 thus pivotal to enacting conservation interventions (Tylianakis *et al.* (2010)). This is particularly urgent as
55 increased biological introductions and climate change are impacting species interactions at a rapid rate, thus
56 calling for more frequent and consistent monitoring (REFs).

57 Recently barcoding (the identification of a sample from a single organism *e.g.* a piece of leaf), and metabar-
58 coding (the identification of a sample containing a mix of organisms *e.g.* soil), have shown considerable
59 promise in all Kingdoms of Life (Ruppert *et al.* (2019)). The ability to identify species from a fragments of
60 organisms (*e.g.* hair, scat, leaf tissue) increases our ability to understand the interactions of not only entire
61 ecosystems but also a focal, generally rare and hence difficult to detect, organism with their surrounding; al-
62 lowing for the most precise allocation of conservation decisions and funds *e.g.* those for restoration processes
63 (Banerjee *et al.* (2022), Johnson *et al.* (2023)). With plants the identification of members of certain clades
64 to species using barcoding has been successful in some groups (Kress (2017)), whereas many other clades
65 have proven more problematic (Liu *et al.* (2014), Group *et al.* (2011), Coissac *et al.* (2012)). Metabarcoding
66 incurs additional challenges to those which exist for the currently implemented barcodes (Li *et al.* (2015),
67 Kress & Erickson (2007), Group *et al.* (2009), Coissac *et al.* (2012)). Particular limitations of the tradi-
68 tional high copy number barcodes (*e.g.* ITS2, *rbcL*, *matK*, *trnH-psbA*) include the utilization their rates of
69 divergence, gene tree conflict, and hybridization (Coissac *et al.* (2016), Fazekas *et al.* (2009)). Currently,
70 most plant metabarcoding endeavors only allow the identification of material to the level of family or genus.

71 [Table 1 about here.]

72 Currently the largest plant systematic endeavor ever undertaken, by the Royal Botanic Gardens Kew, the
73 Plant and Fungal Tree of Life (PAFTOL) is approaching completion (Baker *et al.* (2021a)). This data set will
74 contain hybridization capture (Hyb-Seq) data from at least one species in each genus of the plant kingdom,
75 14,000+ represented species, using the popular Angiosperms353 (A353) probes (Baker *et al.* (2021a), Johnson
76 *et al.* (2019)). Data from the 10kP project, which seeks to develop reference genomes from a phylogenetically
77 diverse suite of 10,000 plant species, will contribute more species by 2030 (Cheng *et al.* (2018)). These
78 publicly available data serve to provide a phylogenetically comprehensive backbone for plant metabarcoding.
79 Similar projects, focused on whole genome sequencing of all organisms in certain geographic areas, such as
80 the ‘Darwin Tree of Life’, which aims to sequence all described taxa in Britain and Ireland, will provide
81 geographically dense data sets to seed projects globally (Life Project Consortium *et al.* (2022), Lewin *et al.*
82 (2022)). These novel genomic data sets will promote the ability to apply metabarcoding to resolve a diverse

83 array of questions relevant to theoretical and applied ecology (Kress (2017), Hollingsworth *et al.* (2016)).
84 However, the application of metabarcoding still faces challenges relating to the enormity of the genomic data
85 sets and the computational power required to process sequence data.

86 To address these challenges, we are proposing the generation of a smaller, and more appropriate reference
87 list for identifying candidate taxa from barcode data. By reducing the list of likely species in a study site
88 it can increase the accuracy and efficiency of metabarcoding results in plants, which will better inform data
89 collection for projects with limited access to taxonomic expertise. Specifically, we propose using current
90 distribution data to generate a list of candidate species relevant to the study area and its ecological charac-
91 teristics (Bell *et al.* (2022)). Initially such a species list can be leveraged to identify a target collections list
92 of plant tissues to be included in genetic studies. It can subsequently help to reduce the size of a reference
93 sequence database, which increases computation speed and efficiency, allowing for the use of genomic data
94 on personal computers. This approach can significantly reduce processing time, increasing project efficiency,
95 particularly as most next-generation sequence data is deposited as raw-sequence reads. We test this metage-
96 nomic and informatics approach by examining the plant species composition within corbiculae pollen loads of
97 bumble bees as a case study. We use DNA barcoding to identify the plant species in corbiculae pollen loads
98 collected from overwintered queen bumble bees, and compare this approach with direct observations of bee
99 flower visits and microscopically-examined pollen slides, which has shown some incongruity in several flo-
100 ral visitation networks involving smaller bodied fauna (Olesen *et al.* (2011), Barker & Arceo-Gomez (2021),
101 Zhao *et al.* (2019), Alarcón (2010)). The assessment of the plant species in pollen is a commonly desired,
102 across several potential applications, despite numerous potential complications (Poronon *et al.* (2017), Bell
103 *et al.* (2017a), Sickel *et al.* (2015), Bell *et al.* (2019), Suchan *et al.* (2019), Johnson *et al.* (2021)).

104 [Table 2 about here.]

105 [Figure 1 about here.]

106 2 | METHODS

107 To improve metabarcoding reliability and efficiency, we suggest creating a regional list of candidate species
108 using digital collections gleaned from herbaria, survey work, and community science (Figure 1). This list
109 can further be refined using species distribution models and temporal filtering to limit the impact of spatial
110 and taxonomic biases in the species list and account for spatial variations in niche availability throughout
111 the study area. The final list is then used to inform collection of plant samples to create a library and inform

112 metabarcoding . We apply this methodological framework to the metabarcoding of corbiculae pollen loads
113 of bumble bees and compare the accuracy of our metabarcoding approach both prior and after applying a
114 spatial and temporal filtering to pollen identification conducted by experts and field observations (Pereira
115 & Cooper (2006),)

116 Create Species List

117 Survey Databases

118 The first step of our methodological approach consists in collating existing species occurrence data to create a
119 preliminary list of candidate species likely to occur in the study area. Such occurrence data can be retrieved
120 from databases including the Botanical Information and Ecology Network ‘BIEN’ (Maitner (2022)) or the
121 Global Biodiversity Information Facility (GBIF; REF) which aggregate occurrence records from various
122 sources including community science datasets (e.g., iNaturalist) and herbarium records. These databases
123 can be queried to extract a list of ecologically relevant vascular plant species with occurrence records in
124 the administrative (e.g. county, region) or ecological unit (e.g. ecoregion) including the study site or within
125 a search radius surrounding the cite. Focusing on a broader geographic region than the study area alone
126 can account for inconsistent sampling effort and undetected species, which are both common in herbarium
127 records and community science datasets

128 Distribution modeling to see which taxa are likely to be within his area

129 While increasingly used in biodiversity assessments (REFs), datasets like BIEN and GBIF can nonetheless
130 include occurrence records with erroneous taxonomic identification, geographic inaccuracies, or historical
131 records (Freitas *et al.* (2020), SMITH????) Hence additional filtering might be needed to develop a more
132 robust list of candidate species. Once a regional list of potential species is established, Species Distribution
133 Models can be used to further reduce the species list to only keep species with ecological requirements
134 found within study sites (Figure 1). This step can also help filter for species adapted to the local ecological
135 condition of the study sites, which can be particularly useful in highly heterogeneous landscapes in which
136 a simple survey of databases may overestimate the list of potential species or when limited taxonomic
137 expertise is available to filter the preliminary species list. Species Distribution Models examine relationships
138 between known species occurrences data and a set of environmental variables to identify potentially suitable
139 habitats with similar ecological conditions elsewhere in the landscape (Guisan & Zimmermann (2000), ...
140). They are often used to inform conservation strategies, evaluate invasion risks, or estimate species richness

141 (Thuiller *et al.* (2006); Pineda & Lobo (2009)). Environmental variables used as predictors in SDMs typically
142 describe the fundamental ecological niche of the species (e.g., climate, elevation, soil types) or constraints
143 to its distribution (e.g., land uses/land covers, canopy openness) and commonly rely on publicly available
144 geospatial datasets (see Table 1 for examples). Once relationships between known occurrence records and
145 these environmental variables are established, SDM algorithms then score remaining portions of the study
146 area on a suitability scale ranging from 0 (highly unsuitable) to 1 (highly suitable) on the basis of their
147 similarity to habitats currently sheltering species populations. Once SDMs are generated for the entire pool
148 of species identified through a database search (section 2.3), these suitability maps can be aggregated to
149 refine the list of species likely to occur in study sites.

150 **Temporal filtering**

151 When studying species interactions, a potentially helpful additional step is to query the candidate species
152 list to retain plant species with phenology that aligns with the study species preferences. For example, when
153 studying plant-pollinator interactions, considering that plant flowering varies over space and time, the project
154 may want to focus only on the species flowering during the pollinators' active period,. These contrasts in
155 the phenological periods can thus provide an additional filter for identifying material in certain types of
156 metagenomic samples (Janzen (1967), Newstrom *et al.* (1994), Thompson (1994), CaraDonna *et al.* (2017)).
157 Community science efforts focusing on plant phenology can provide helpful information on the phenology
158 of commonly observed species. Examples of such datasets include: the National Phenology Network in the
159 United States (Denny *et al.* 2014), SeasonWatch in India, PlantWatch in Canada, and the Pan European
160 Phenology Project. Local Floras can also provide useful information on the phenological characteristics of
161 species.

162 **Collecting representative samples for reference library**

163 GOES HERE

164 **Using baits to pull out specific regions (barcode) of the genome**

165 For barcoding the environmental samples and leaf samples collected from candidate taxa, we are proposing
166 using a Next generation sequencing approach which incorporates a target capture probes (Jones & Good
167 (2016)). There are several approaches available for preparing genomic libraries that vary in cost, ease of use
168 (CITE), and the quality and quantity of sample DNA required that are all appropriate for target capture

169 approach. The choice of best library preparation approach will depend on the type of samples, the lab and
170 sequencing platform preferred. Once a library has been generated, and size selected to generate appropriate
171 sized fragments, and appropriately barcoded to distinguish individuals, a synthetic probes, often referred to
172 as baits, is used to preferably retain fragments within the library which contain sequences associated with
173 desired loci, while remainder of fragments which do not are washed away, maximizing sequencing efficiency.
174 There are a number of probe sets which have been developed, some are taxon specific, but a handful are
175 more general and appropriate for barcoding multiple taxa, including Angiosperms353 probe set (Johnson *et*
176 *al.* (2019)), which is appropriate for most flowering species, and goFlag (Breinholt *et al.* (2021)) might be
177 more appropriate for non-flowering species. These approaches capture data from multiple loci across genome,
178 which are both expressed, therefore have lower mutation rates, and are found to occur in single copies in
179 most taxa, making them appropriate for taxonomic studies.

180 Compare barcode to reference library

181 The data generated is filtered and trimmed using pipelines such as TRIMOMATIC Bolger & Giorgi (2014)
182 to identify and remove sequences added during library preparation process and not associated with sample.
183 These sequence fragments (contigs) are then compared back to the reference probe sequences to sort them by
184 loci of origin Johnson *et al.* (2016). For libraries which contain only a single taxa, the contigs from the same
185 loci then be aligned to generates a consensus sequence Katoh & Standley (2013). This consensus sequence
186 can then be used as a reference for which to compare the multiple species samples and identify taxa.

187 Case Study: Measuring bumblebee visitation based on corbiculae load

188 Bee-Flower Observations and Pollen Load Collection

189 To test this metagenomic and informatics approach, we used the identification of plant species within corbic-
190 ulae pollen loads of bumble bees collected at the Rocky Mountain Biological Laboratory (RMBL; 38°57.5”
191 N, 106°59.3” W (WGS 84), 2900 m.a.s.l.), Colorado, USA (Appendix 1 for site information). Bumblebees are
192 one of the only groups of insects with unequivocal quantitative evidence for numerous population declines.
193 They are also the most effective pollinators in many temperate montane ecosystems (Cameron & Sadd
194 (2020), Goulson *et al.* (2008b), Williams (1982), Colla *et al.* (2012), Bergman *et al.* (1996), Bingham &
195 Orthner (1998), Grixti *et al.* (2009)), which are experiencing steep changes in annual temperature under
196 climate change (Brito-Morales *et al.* (2018), Pepin *et al.* (2022)). As such, the effective management of
197 these ecosystems relies on the study of keystone species, such as bumble bees, that play a critical role in

198 maintaining their current diversity (Loarie *et al.* (2009), Dobrowski & Parks (2016)). The queen bumblebees
199 emerge early in the season to identify nesting sites. Floral resources are essential at these foraging stages
200 for increasing their weight before diapause and increasing their ovary weights while establishing their found
201 nests, and therefore this period represent potential demographic bottlenecks in bumble bee populations
202 (Sarro *et al.* (2022)).

203 **Species list**

204 **Spatial Analyses** The threshold distance under which a species would undergo species distribution mod-
205 elling was the median (25.009 km) of the logistic regression assessing the probability of occurrence of a
206 species record as a function of distance from the study area. A 2-sample test for equality of proportions
207 with continuity correction ($X^2 = 13.254$, df = 1, p-value = 0.000136, 95% CI 0.04-1.00) was used to
208 test whether more of the records located in the broad ecological sites present at the field station, between
209 the distance of the median (25.009 km) to the third quantile (ca 43.830 km) of the regression distance, were
210 true presences at the field station. Including these records would have resulted in modelling an additional
211 222 species distributions of which 30 are true presences, these taxa were not modelled.

212 Across the entire spatial domain of modelling, all ensembled models ($n = 968$) had an accuracy of 0.84
213 (95% CI 0.8356 - 0.8443), kappa = 0.68, p-value < 0.001, sensitivity = 0.80, specificity = 0.87, AUC =
214 0.92. The 493 ML ensembles accurately predicted the presence of 362 occurrences (65.3%), incorrectly
215 predicted the presence of 64 (11.6%), incorrectly predicted 34 true presences (6.1%) as being absent, and
216 correctly predicted the true absence of 33 (6.0%). Of the 554 vascular plants with biotic pollination, the
217 475 LM ensembles accurately predicted the presence of 286 (51.6%), incorrectly predicted the presence of
218 41 (14.3%), incorrectly predicted 93 true presences (16.8%) as being absent, and correctly predicted the
219 true absence of 55 (9.9%). The balanced accuracy of the ensembled models is 0.664 (Sensitivity = 0.573,
220 Specificity 0.754). Of the 117 plant species identified to the species level across the spatial extents of all plots
221 and duration of queen bee activity, the ML ensembles predicted the presence of 105 (89.7%) of them, and
222 LM ensembles 102 (87.2%). Of the missing species two (1.7%) are Orchids, six (5.1%) are non-native, and
223 one (0.85%) is of contested taxonomic standing, all of which (7.65%) are restricted from the initial query
224 database.

²²⁵ **Pollen and plant material collection**

²²⁶ Observations of visitation and bee corbiculae pollen collection was conducted at the Rocky Mountain Bio-
²²⁷ logical Laboratory (RMBL; 38°57.5" N, 106°59.3" W, 2900 m.a.s.l.), Colorado, USA (Appendix 1 for site
²²⁸ information). Pollinator observations of *Bombus Latreille* spp. (Apidae Latreille) were conducted from May
²²⁹ 29th – July 23rd of 2015 in six study sites as a part of a larger study (described in Ogilvie & CaraDonna
²³⁰ (2022)). Observations of *Bombus* foraging took place for one hour at each field site. Corbiculae pollens loads
²³¹ were non-lethally collected from queens encountered by capturing them in an insect net and transferring
²³² them into a restraining device (Kearns *et al.* (2001)). We then collected a single pollen load (i.e., from one
²³³ leg) from the bee and then released it. At weekly intervals at each site, we also recorded the abundances of
²³⁴ flowers visited by bumble bees within belt transects spread over three vegetation types (0.5 x 40 m transects
²³⁵ in each vegetation type, 60 m² total area per site). Using five years (2015-2020) of observational data on
²³⁶ *Bombus* queen interactions with flowering plants at these studies sites, we identified the plant taxa most
²³⁷ frequently visited by queens across all years. In order to capture more variability inherit in the 353 loci we
²³⁸ sampled the 12 most visited taxa twice using samples collected from one site within the Gunnison Basin
²³⁹ River Drainage and one individual collected from another more distal population. In addition, we included
²⁴⁰ a congener - or a species from a closely related genus to serve as an outgroup for all 12 taxa to confirm
²⁴¹ the efficiency of 353 baits for species level identification. We also sequenced another 15 taxa of plants com-
²⁴² monly visited by *Bombus* workers, based on the abundances, and immediate access to plant tissue, in the
²⁴³ aforementioned data set (Appendix 3). Plant collections were identified typically using a combination, of
²⁴⁴ dichotomous keys and primary literature as required (Flora of North America Editorial Committee (1993+),
²⁴⁵ Hitchcock & Cronquist (2018), Ackerfield (2015), Lesica *et al.* (2012), Cronquist *et al.* (1977+), Allred &
²⁴⁶ Ivey (2012), *Jepson flora project* (2020), Mohlenbrock (2002)).

²⁴⁷ **Validation process**

²⁴⁸ To precisely classify the contents of each corbiculae load, the sequences classified by molecular methods
²⁴⁹ were compared with the fieldwork which recorded the presence and absence of species and their duration
²⁵⁰ of flowering at a very fine resolution, . The quantitative counts of grains from microscopy were combined
²⁵¹ with the semi-quantitative sequencing results to estimate the abundance of each identified species in each
²⁵² corbiculae load.

253 **Validation using morphological pollen identification**

254 Five years of observational data on *Bombus Queen Bee* foraging at these studies sites Ogilvie & CaraDonna
255 (2022) was used as baseline of what we should expect for visitation within this system. This was supplemented
256 with the RMBL Vascular Plant Checklist (Frase & Buck (2007)). Microscopy was use to qualitatively identify,
257 and quantitatively note, the plant species present in corbiculae loads. A pollen reference library of know
258 taxa visited by Bumblebees and therefore expected to be present in the corbiculae loads was made using a
259 fuchsin-jelly stained grains from slides previously prepared by the authors (n = 21), and other researchers
260 (n = 38) (Beattie (1971), Brosi & Briggs (2013)). A total of 62 voucher slides for species were prepared and
261 imaged at 400x (Leica DMLB, Leica MC170 HD Camera, Leica Application Suite V. 4.13.0) from previously
262 prepared slides and both non-accessioned and RMBL herbarium collections to supplement the number of
263 species and clades covered (Appendix 3). To prepare the pollen slides from corbiculae, all corbiculae loads
264 were broken apart and rolled using dissection needlepoints to increase heterogeneity of samples. Circa
265 0.5mm² of pollen was placed onto a ~4mm² fuchsin jelly cube Beattie (1971) atop a graticulated microscope
266 slide, with 20 transects and 20 rows (400 quadrants) (EMS, Hartfield, PA). The jelly was melted, with
267 stirring, until pollen grains were homogeneously spread across the microscope slide. Slides were sealed with
268 Canada Balsam (Rublev Colours, Willits, CA) followed by sealing with clear nail polish to prevent oxidation;
269 all samples are noted in Appendix 4.

270 To identify the pollen present in corbiculae loads, light microscopy at 400x (Zeiss Axioscope A1) was used. In
271 initial sampling in three transects, each pollen grain was identified to morphotype and counted; an additional
272 two transects were scanned for morphotypes unique to that slide, if either transect contained a unique
273 morphotype then all grains in that transect were also identified and counted. We used clustering techniques
274 to supplement our subjective opinions of which plant taxa were distinguishable via light microscopy, and to
275 develop a dichotomous key to pollen morphotypes. Ten readily discernible categorical traits were collected
276 from each specimen in the image collection. These traits were transformed using Gower distances, and
277 clustered using Divisive Hierarchical clustering techniques (Maechler *et al.* (2022)). Using the cluster
278 dendrogram, elbow plot, and heatmaps Hennig (2020), of these results morphological groups of pollen which
279 could not be resolved via microscopy were delineated, and a dichotomous key was prepared (Appendix 6).
280 This key was then used to identify the pollen grains sampled from corbiculae loads to morphotypes in a
281 consistent manner.

282 Subsequent to the first round of sampling, non-parametric species richness rarefaction curves (Oksanen *et*
283 *al.* (2022)), and non-parametric species diversity rarefaction curves were used to assess the completeness of

284 sampling (Chao *et al.* (2014), Hsieh *et al.* (2020)). Slides not approaching the asymptote of the rarefaction
285 curve were then re-sampled, and analysed iteratively for up to a total of seven transects (Appendix 7 & 8).

286 To reclassify the sequence reads, these data were combined with the floral observation data, and mapped
287 by genus. If more than one species in the genus was flowering at that time and site, then the reads were
288 split evenly between the taxa. For sequence data which did not match at the genus level, a user subjectively
289 scored them based on the species composition and phenological activity at each site, the queen interaction
290 data, and pollen assignments. To estimate the abundance of each of these species in the corbiculae loads,
291 these data were combined with the microscopy data. For each morphotype detected in pollen, and each
292 classified sequence read which was not detected via microscopy, they were given a value of 0.5% to indicate
293 their trace presences. When more than a single species belonged to a morphotype group in a single sample,
294 the quantitative values from the morphological work were multiplied by the relative sequence abundance of
295 each species in the load. All final compositions were standardized to a sum of 100%, by adding or subtracting
296 the differences (induced by classifying records as ‘trace’) to all species with abundances > 1%.

297 2.3 | Pollen Morphological identification

298 2.3.1 | Pollen Reference Library

299 To qualitatively identify, and quantitatively note, the plant species present in corbiculae loads microscopy
300 was used. A pollen reference library of fuchsin-jelly stained grains which may be present in corbiculae loads
301 of slides was assembled from slides previously prepared by the authors ($n = 21$), and other researchers ($n =$
302 38) (Beattie (1971), Brosi & Briggs (2013)). Using five years of observational data on *Bombus* Queen Bee
303 foraging at these studies sites (Ogilvie & CaraDonna (2022)), as well as the RMBL Vascular Plant Checklist
304 (Frase & Buck (2007)), an additional 62 voucher slides for species were prepared and imaged at 400x (Leica
305 DMLB, Leica MC170 HD Camera, Leica Application Suite V. 4.13.0) from non-accessioned herbarium col-
306 lections to supplement the number of species and clades covered (Appendix 3).

307 We used clustering techniques to supplement our subjective opinions of which plant taxa were distinguishable
308 via light microscopy, and to develop a dichotomous key to pollen morphotypes. Ten readily discernible cate-
309 gorical traits were collected from each specimen in the image collection. These traits were transformed using
310 Gower distances, and clustered using Divisive Hierarchical clustering techniques (Maechler *et al.* (2022)).
311 Using the cluster dendrogram, elbow plot, and heatmaps (Hennig (2020)), of these results morphological
312 groups of pollen which could not be resolved via microscopy were delineated, and a dichotomous key was
313 prepared (Appendix 6). This key was then used to identify the pollen grains sampled from corbiculae loads

³¹⁴ to morphotypes in a consistent manner.

³¹⁵ **2.3.2 | Preparation of Pollen Corbiculae Loads**

³¹⁶ To prepare the pollen slides from corbiculae, all corbiculae loads were broken apart and rolled using dissection
³¹⁷ needlepoints to increase heterogeneity of samples. *Circa* 0.5mm² of pollen was placed onto a ~4mm² fuchsin
³¹⁸ jelly cube (Beattie (1971)) atop a graticulated microscope slide, with 20 transects and 20 rows (400 quadrants)
³¹⁹ (EMS, Hartfield, PA). The jelly was melted, with stirring, until pollen grains were homogeneously spread
³²⁰ across the microscope slide. Slides were sealed with Canada Balsam (Rublev Colours, Willits, CA) followed
³²¹ by sealing with clear nail polish to prevent oxidation; all samples are noted in Appendix 4. To identify the
³²² pollen present in corbiculae loads, light microscopy at 400x (Zeiss Axioscope A1) was used. In initial sampling
³²³ in three transects, each pollen grain was identified to morphotype and counted; an additional two transects
³²⁴ were scanned for morphotypes unique to that slide, if either transect contained a unique morphotype than
³²⁵ all grains in that transect were also identified and counted. Subsequent to the first round of sampling, non-
³²⁶ parametric species richness rarefaction curves (Oksanen *et al.* (2022)), and non-parametric species diversity
³²⁷ rarefaction curves were used to assess the completeness of sampling (Chao *et al.* (2014), Hsieh *et al.* (2020)).
³²⁸ Slides not approaching the asymptote of the rarefaction curve were then re-sampled, and analysed iteratively
³²⁹ for up to a total of seven transects (Appendix 7 & 8).

³³⁰ **2.4 | Molecular Barcoding**

³³¹ **2.4.1 | Species reference list**

³³² **2.4.1.1 Spatial Analyses** We first generated a short list of potential candidate species. We downloaded,
³³³ from the Botanical Information and Ecology Network ‘BIEN’ (Maitner (2022)), all records adjacent to the
³³⁴ field sites to develop an ecologically relevant list of vascular plant species, with expected biotic pollination,
³³⁵ which may be present at the study area. To reduce the number of species to include in the reference sequence
³³⁶ databases, we then generated Species Distribution Models (SDMs) for these taxa to predict their distribution
³³⁷ throughout the study area.

³³⁸ To minimize the number of species for which SDM’s were to be generated, BIEN was queried at a distance
³³⁹ of up to 100km from our study area and all plant species records were downloaded. To account for the
³⁴⁰ stochasticity of botanical collecting and offset the number of records associated with the research station,
³⁴¹ this data set was bootstrap re-sampled 250 times, with 90% of samples selected, to create a testing data

342 set. The median of the logistic regression assessing the probability of occurrence of a species record as a
343 function of distance from the study area was used as a threshold distance, under which, to include species
344 as candidates for distribution modelling.

345 **2.4.1.2 Distribution Modelling** To determine which clades to include in the reference sequence database
346 we used Species Distribution Modelling. We used all occurrence records from BIEN ($n = 23,919$) within a
347 50km border of the ecoregion, Omernik level 3, which includes the study area (*No. 21 “Southern Rockies”*)
348 to construct the species distribution model (Omernik (1987)). These records were copied into two, initially
349 identical, sets, one for generating machine learning models (ML; Random Forest, and Boosted Regression
350 Tree’s), and the other for Generalised Linear (GLM) and Generalized Additive Models (GAM) (Barbet-
351 Massin *et al.* (2012)). Ensembled predictions have been shown to outperform their constituent models,
352 on average, and to reduce the ecological signal to the analytical noise of individual runs (Araujo & New
353 (2007)). No single method of producing SDMs has been shown to universally outperform others when faced
354 with a large and diverse number of applications, in our case a great number of species with different biology
355 and ecology (Elith* *et al.* (2006), Qiao *et al.* (2015)). In the spirit of these findings, multiple families of
356 models, which can be generated together as they have similar requirements regarding the number and ratios
357 of Presence to Absence records were ensembled together (Barbet-Massin *et al.* (2012)).

358 We then generated 4,029 absence points, locations where the focal taxon is anticipated missing, through a
359 random stratification of 19% of the land cover in the area and included them in (Land Management (2019)).
360 To achieve a larger absence data set, we generated 1,000 pseudo-absence records for each taxon by randomly
361 selecting coordinates located at least 10km away from an occurrence record. For ML models, these pseudo-
362 absences were reduced so that the ratio of presence to absence records were balanced (Barbet-Massin *et al.*
363 (2012)). To achieve this, we removed absence records inside of 10% of the mean sample value of any predictor
364 variable the presence records; the required number of absence records were then randomly sampled.

365 To predict the potential distribution of each species we used 26 environmental variables at 30m resolution,
366 six related to climate, five soil, four topographic, four related to cloud cover, with the remaining reflecting
367 assorted abiotic parameters (Wilson & Jetz (2016), Wang *et al.* (2016), Hengl *et al.* (2017), Robinson *et*
368 *al.* (2014)) (Appendix 2). These publicly available data sets, were selected as they pertain to a wide range
369 of variables interacting with plant physiology. For linear regression models these predictors underwent both
370 *vifstep* ($\theta = 10$, max observations = 12,500) and *vifcor* ($\theta = 0.7$, max observations = 12,500) to
371 detect highly correlated variables, and collinear features were removed leaving 16 variables (Naimi *et al.*
372 (2014)).

373 Modelling: Random Forest and Boosted Regression Trees, were sub sampled with 30% test and two replicates
374 each before weighted ensemble based on True Skill Statistics (tss) (Naimi & Araujo (2016)). Generalised
375 linear models (GLM) and Generalised additive models (GAM) with 30% sub sampling and three replicates
376 each were also ensembled using the tss (Naimi & Araujo (2016)). TSS was chosen as the ensemble criterion as
377 it has been shown to work across a wide range of species occurrences prevalence (Allouche *et al.* (2006)). The
378 results of these models were extracted on a cell-by-cell basis to a polygon feature derived from a minimum-
379 spanning tree which encompasses the study sites, and species from either ensemble with greater than 50%
380 mean habitat suitability across all cells were considered present for further purposes (Prim (1957)).

381 A total of 535 species were modelled using Generalized Linear Models and Generalized Additive Models and
382 534 species were modelled using Random Forest and Boosted Regression Trees. To evaluate the accuracy of
383 the species distribution models, additional presence records from GBIF ($n = 61,789$), and AIM ($n = 12,730$)
384 were used as test and training sets ($n = 74,519$) for logistic regression (Occdownload Gbif.Org (2021), Land
385 Management (2019)). Additional novel absence records were generated from the AIM data set to create a
386 data set where each species has balanced presence and absences. Eleven or more paired presence and absence
387 records were required for this testing, resulting in 334 species being included in the logistic regression (Mdn
388 = 110.0, $\bar{x} = 223.1$, max = 1568 record pairs used) with a 70% test split (Kuhn (2022)).

389 2.4.2 | Temporal Analyses

390 For assignment of reads to ecologically probabilistic species subsequent to BLAST, flowering time was used
391 as a filter. To estimate the duration of dates in which plant species were flowering Weibull estimates of
392 several phenological parameters all spatially modelled taxa were developed (Belitz *et al.* (2020), Pearse *et*
393 *al.* (2017)). Only BIEN records which occurred in the Omernik Level 4 Ecoregions within 15km of the
394 study area ($n = 5$ Level 4 Ecoregions, or conditionally 6 ecoregions if enough records were not found in
395 the nearest 5), and which were from herbarium records were included. To remove temporally irrelevant
396 herbarium records, i.e. material collected during times which flowering is impossible at the study area due
397 to snow cover, we used the SnowUS data set (Iler *et al.* (2021), Tran *et al.* (2019)) from 2000-2017 were
398 analyzed for the first three days of contiguous snow absence, and the first three days of contiguous snow cover
399 in fall. Herbarium records after the 3rd quantile for melt, and the 1st quantile for snow cover of these metrics
400 were removed. Species with > 10 records had their Weibull distributions generated for the date when 10%
401 of individuals had begun flowering, when 50% were flowering, and when 90% of individuals had flowered,
402 we used the initiation and cessation dates, respectively, as effective start and ends of flowering. These
403 estimates were compared to a long-term observational study of flowering phenology 1974-2012 (CaraDonna

404 *et al.* (2014)), and the floral abundance data from 2015, using Kendall's tau.

405 **2.5.2 | Barcode references library**

406 All lab work was carried out at The Daniel F. and Ada L. Rice Plant Conservation Science Center at the
407 Chicago Botanic Garden, Glencoe, Illinois, U.S.A.

408 **2.5.2.1 | Sampling Species for Barcoding** Using five years (2015-2020) of observational data on *Bombus*
409 queen interactions with flowering plants at these studies sites, we identified the plant taxa most frequently
410 visited by queens across all years. In order to capture more variability inherit in the 353 loci we sequenced the
411 12 most visited taxa twice using samples collected from one site within the Gunnison Basin River Drainage
412 and one individual collected from another more distal population. In addition we included a congener - or
413 a species from a closely related genus to serve as an outgroup for all 12 taxa. We also sequenced another
414 15 taxa of plants commonly visited by *Bombus* workers, based on the abundances, and immediate access to
415 plant tissue, in the aforementioned data set (Appendix 3). Plant collections were identified typically using
416 a combination, of dichotomous keys and primary literature as required (Flora of North America Editorial
417 Committee (1993+), Hitchcock & Cronquist (2018), Ackerfield (2015), Lesica *et al.* (2012), Cronquist *et al.*
418 (1977+), Allred & Ivey (2012), *Jepson flora project* (2020), Mohlenbrock (2002)).

419 **2.5.2.2 | Plant Genomic DNA Extraction** Plant genomic DNA was isolated from ~ 1 cm² of leaf tissue
420 from silica-gel dried or herbarium material using a modified cetyltrimethylammonium (CTAB) protocol
421 (Doyle & Doyle (1987)) that included two chloroform washes. DNA was quantified using a Nanodrop 2000
422 (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and Qubit fluorometer (Thermo Fisher Scientific).

423 **2.5.2.3 | Pollen Genomic DNA Extraction** To extract genomic DNA from pollen a CTAB based
424 protocol was modified from Lahlamgiah et al. and Guertler et al. (2014, 2014). A SDS extraction buffer
425 (350µL , 100mM Tris-HCl, 50 mM EDTA, 50 mM NaCl, 10% SDS v/v., pH 7.5) was added to the sampled
426 followed by vortexing to allow dissolution of corbiculae. Pollen grains were then macerated with Kontes Pellet
427 Pestles, and the tip of these washed with 130 µL of the SDS extraction buffer, samples were then incubated
428 for 1 hour at 30°C. This was followed by the addition of 10% CTAB solution (450ul, of 20 mM Tris-Cl pH.
429 8.0, 1.4 M NaCl, 10 mM EDTA pH 7.5, 10% CTAB, 5% PVP, ~85% Deionized water) and RNase (10 uL
430 of 10 mg/mL) and samples were incubated for 40 minutes at 37°C, on a heat block (Multi-Blok, Thermo
431 Fisher Scientific, Waltham Massachusetts) set to 40°C. After 20 minutes incubation, Proteinase K (15 µL of

432 20mg/ml) and DTT (12.5 μ L of 1M in water) were added, and the samples were further incubated at 60°C
433 for 1 hour. Samples were then incubated overnight at 40°C. 500 μ L of Phenol-Chloroform-Isoamyl alcohol
434 (25:24:1) were added, vortexed, and centrifuged at 10,000 rpm for 10 minutes and the aqueous phase was
435 pipetted to a 1.5 ml centrifuge tube.

436 To precipitate the DNA, chilled Isopropyl alcohol & 3 mM Sodium acetate (5:1) equivalent to $\frac{2}{3}$ of the volume
437 of sample were added, with 1 hour of chilling at -20°C, followed by 10 minutes of centrifuging at 13,000 rpm.
438 The supernatant was pipetted to a new 1.5 ml centrifuge tube, and 70% EtOH (400 μ L) were added before
439 chilling at -20°C for 20 minutes followed by centrifugation at 13,000 rpm for 10 minutes. Both tubes were
440 then washed with 75% EtOH (400 μ L), inverted, centrifuged at 13,000 rpm for 4 minutes, and the solution
441 discarded, then washed with 95% EtOH (400 μ L), inverted, centrifuged at 13,000 rpm for 4 minutes, and the
442 solution discarded. Pellets were dried at room temperature overnight before resuspension in nuclease free
443 H₂O. Extractions were assessed using a Nanodrop 2000 (Thermo Fisher Scientific) and Qubit fluorometer
444 (Thermo Fisher Scientific). DNA extracts were then cleaned using 2:1 v./v. Sera-Mag beads (Cytiva, Little
445 Chalfont, UK) to solute ratio following the manufacturer's protocol, eluted in 0.5x TE, and the eluent allowed
446 to reduce by half volume in ambient conditions. DNA was quantified using a Qubit fluorometer.

447 **2.5.2.4 | Fragmentation, Library Preparation & Target Enrichment** Sequence library preparation
448 was performed using the NEBNext Ultra II FS-DNA Library Prep Kit for Illumina (New England BioLabs,
449 Ipswich, Massachusetts, USA) using slightly modified manufacturers recommendation. Fragmentation was
450 performed at $\frac{1}{2}$ volume of reagents and $\frac{1}{4}$ enzyme mix for 40 minutes at 37°C, with an input of 500 ng cleaned
451 DNA. Adapter Ligation and PCR enrichment were performed with $\frac{1}{2}$ volumes, while cleanup of products
452 was performed using SPRI beads (Beckman Coulter, Indianapolis, Indiana, USA) and recommended volumes
453 of 80% v./v. ethanol washes. The exception was the herbarium specimens which were not fragmented and
454 only end repaired, with similar library preparation of all samples. Products were analysed on 4% agarose
455 gels, and a Qubit fluorometer. Libraries were pooled and enriched with the Angiosperms 353 probe kit V.4
456 (Arbor Biosciences myBaits Target Sequence Capture Kit) by following the manufacturer's protocol and
457 Brewer et al. 2019. Sequencing was performed using an Illumina mi-Seq with 150-bp end reads, (NUSeq
458 Core, Chicago, Illinois).

459 **2.6 | Computational Processes and Analyses.**

460 **2.6.1 | Reference Library Data Processing** Sequences were processed using Trimmomatic, which
461 removed sequence adapters, clipped the first 3 bp, discarding reads less than 36 bp, and removing reads

462 if their average PHRED score dropped beneath 20 over a window of 5 bp (Bolger & Giorgi (2014), Tange
463 (2021)). Contigs generated were mapped to a reference with HybPiper with using target files created by
464 M353 (Johnson *et al.* (2016), McLay *et al.* (2021)).

465 **2.6.2 | Sequence Identification** A custom Kraken2 database was created by downloading representative
466 species indicated as being present in the study area by the spatial analyses from the Sequence Read Archive
467 (SRA) NCBI (Wood *et al.* (2019)). These sequences were processed in the same manner as our novel
468 sequences. The Kraken2 database was built using default parameters. Kraken2 was run on sequences using
469 default parameters (Appendix 9). Following Kraken2, Bracken was used to classify sequences to terminal
470 taxa (Lu *et al.* (2017)). Finally all reads which could be classified by these databases were passed to a local
471 BLAST database.

472 **2.6.3 | Identification of Sequence Matching Loci** A local NCBI database was built using the same
473 processed novel and downloaded sequences as the previous database (Camacho *et al.* (2009)).

474 **2.7 | Integrated Observational, Molecular, and Palynological Corbiculae**

475 To precisely classify the contents of each corbiculae load the sequences classified by molecular methods were
476 compared with the fieldwork which at a very fine resolution, recorded the presence and absence of species
477 and their duration of flowering, and was interpreted ala the computer derived temporal and spatial data sets.

478 The quantitative counts of grains from microscopy, were combined with the semi-quantitative sequencing
479 results, to estimate the abundance of each identified species in each corbiculae load.

480 To reclassify the sequence reads, these data were combined with the flora observation data, and mapped
481 by genus. If more than one species in the genus was flowering at that time and site, than the reads were
482 split evenly between the taxa. For sequence data which did not match at the genus level, a user subjectively
483 scored them based on the species composition and phenological activity at each site, the queen interaction
484 data, and pollen assignments. To estimate the abundance of each of these species in the corbiculae loads,
485 these data were combined with the microscopy data. For each morphotype detected in pollen, and each
486 classified sequence read which was not detected via microscopy, they were given a value of 0.5% to indicate
487 their trace presences. When more than a single species belonged to a morphotype group in a single sample,
488 the quantitative values from the morphological work were multiplied by the relative sequence abundance of
489 each species in the load. All final compositions were standardized to a sum of 100%, by adding or subtracting
490 the differences (induced by classifying records as ‘trace’) to all species with abundances > 1%.

491 **3 | RESULTS**

492 [Table 3 about here.]

493 [Figure 2 about here.]

494 [Table 4 about here.]

495 **Spatial Analyses**

496 [Table 5 about here.]

497 [Table 6 about here.]

498 The threshold distance under which a species would undergo species distribution modelling was the median
499 (25.009 km) of the logistic regression assessing the probability of occurrence of a species record as a function
500 of distance from the study area. A 2-sample test for equality of proportions with continuity correction
501 (χ^2 = 13.254, df = 1, p-value = 0.000136, 95% CI 0.04-1.00) was used to test whether more of the
502 records located in the broad ecological sites present at the field station, between the distance of the median
503 (25.009 km) to the third quantile (ca 43.830 km) of the regression distance, were true presences at the field
504 station. Including these records would have resulted in modelling an additional 222 species distributions of
505 which 30 are true presences, these taxa were not modelled.

506 Across the entire spatial domain of modelling all ensembled models (n = 968) had an accuracy of 0.84 (95%
507 CI 0.8356 - 0.8443), kappa = 0.68, p-value < 0.001, sensitivity = 0.80, specificity = 0.87, AUC = 0.92.

508 Across the entire spatial domain of modelling, all ensembled models (n = 968) had an accuracy of 0.84
509 (95% CI 0.8356 - 0.8443), kappa = 0.68, p-value < 0.001, sensitivity = 0.80, specificity = 0.87, AUC =
510 0.92. The 493 ML ensembles accurately predicted the presence of 362 occurrences (65.3%), incorrectly
511 predicted the presence of 64 (11.6%), incorrectly predicted 34 true presences (6.1%) as being absent, and
512 correctly predicted the true absence of 33 (6.0%). Of the 554 vascular plants with biotic pollination, the
513 475 LM ensembles accurately predicted the presence of 286 (51.6%), incorrectly predicted the presence of
514 41 (14.3%), incorrectly predicted 93 true presences (16.8%) as being absent, and correctly predicted the
515 true absence of 55 (9.9%). The balanced accuracy of the ensembled models is 0.664 (Sensitivity = 0.573,
516 Specificity 0.754). Of the 117 plant species identified to the species level across the spatial extents of all plots
517 and duration of queen bee activity, the ML ensembles predicted the presence of 105 (89.7%) of them, and

518 LM ensembles 102 (87.2%). Of the missing species two (1.7%) are Orchids, six (5.1%) are non-native, and
519 one (0.85%) is of contested taxonomic standing, all of which (7.65%) are restricted from the initial query
520 database.

521 **Temporal Analysis**

522 The first date of modeled snow melt in the Gothic area ($n = 17$, $\bar{x} = 137.9$, Mdn = 135, 3rd quartile = 151),
523 and the first date of a consistent winter snow base ($n = 17$, $\bar{x} = 299.9$, Mdn = 300, 1st quartile = 291) from
524 2000-2017, were used as delimiters for the inclusions of herbarium records in modelling. Of the 439 species
525 predicted likely present in the area via logistic regression, 332 species (64.4%) with more than 10 records in
526 the focal level 4 ecoregions ($\bar{x} = 35.016$, Mdn = 35, max = 96) had Weibull estimates calculated, an additional
527 56 species (11.2%) with enough contributing records from the “Sedimentary Mid-Elevation Forests”, a large
528 ecoregion generally just beneath the elevation bands occupied by the five ecoregions around the study area
529 had Weibull estimates also calculated ($\bar{x} = 13.868$, Mdn = 13, max = 24). We could only compare 58 of
530 these 388 species to plot based observational data from the long term (1974–2012) data set (CaraDonna *et*
531 *al.* (2014)), but this revealed high accord between the long-term ground truthed data set and the modelled
532 species. There was very strong evidence that the Weibull estimates were positively associated with the
533 observed onset. There was very strong evidence that the Weibull estimates were positively associated with
534 the observed onset ($p < 0.0001$, tau = 0.61), peak ($p < 0.0001$, tau = 0.65), and cessation of flowering
535 ($p < 0.0001$, tau = 0.49). There was moderate evidence that the Weibull estimates had a weak positive
536 association with the observed duration of flowering ($p = 0.58$, tau = 0.17).

537 Of the previous 58 species compared, 47 of these could be compared to transect based data from the six
538 sites observed in 2015. There was very strong evidence that the Weibull estimates were positively associated
539 with the observed onset ($p < 0.0001$, tau = 0.58), and cessation of flowering ($p < 0.0001$, tau = 0.40).

540 [Figure 3 about here.]

541 [Figure 4 about here.]

542 **Molecular analysis**

543 **Plant data**

544 Need some text here... how well did plant capture go?

545 **Corbiculae loads**

546 The 54 corbiculae loads had DNA extracted and underwent various steps towards hyb-seq, in the end a total
547 of 44 corbiculae samples were sequenced, 7,752,353 reads were recovered from sequencing. The number of
548 reads per sequence varied widely (range = 76 - 508,795, $\bar{x} = 176,189.8$, Mdn = 138,395). Of the possible 353
549 loci, the number which were recovered from each sample, and informative to BLAST were range = 24 - 353,
550 $\bar{x} = 305.5$, Mdn = 331. The number of reads per loci from across all samples had a range of 178 - 506,653,
551 $\bar{x} = 20,688$, Mdn = 12,616 (Appendix 11). After trimming 7,865,680 sequences remained. 10,682,538 reads
552 were matched using Kraken, of the reads classified by Kraken 10,160,768 reads were matched using Bracken,
553 of the reads classified by Kraken 0 reads were matched using BLAST. Based upon subjective review of
554 the three classifiers (Appendix 12) BLAST was chosen as the classification method which yielded the most
555 probable results by the field ecologist, and its values were used for all subsequent analyses.

556 **Metabarcoding Pollen Identification**

557 content should be here

558 **Validation**

559 In 2015 the six sites were surveyed once weekly from May 27-July 27 for a total of 52 hours. A total of 723
560 overwintered queen-pollen foraging interactions were observed (range per bee species by week range = 1-18,
561 $\bar{x} = 3.46$, median = 2), with a range of total observed interactions per bee species across this time period
562 (minimum = 1, $\bar{x} = 59.08$, median = 19, max = 184). Plants varied widely in the number of interactions
563 which they partook in with each species of bee (range per plant species by week minimum = 1 - 20, $\bar{x} = 3.51$,
564 median = 2), with a range of total observed interactions per plant species over this time period (minimum
565 = 1, $\bar{x} = 20.26$, median = 4, max = 141). The number of plant species which bees were observed interacting
566 with varied more narrowly (range = 1 - 18, $\bar{x} = 8$, median = 6), interactions were observed with a total of
567 36 plant species.

568 A total of 66 corbiculae loads were collected from bees, 64 of them from queens.

569

[Figure 5 about here.]

570 **Microscopic Pollen identification**

571 Using the fuchsin jelly preparation and light microscopic analyses of grains and scoring of 10 character states
572 resulted in the establishment of 28 morphotypes which grains could be classified into (Appendix 6). From the
573 37 samples that were counted and based on rarefaction we identified substantial amounts of the abundance
574 and morphotype richness of the samples (morphotype richness, $\bar{x} = 4.5$, median = 4, min = 1, max = 9)
575 (Appendix 7 & 8). The number of counted pollen grains in each sample range from (514 - 19924, $\bar{x} = 3319$,
576 median = 1891).

577 [Figure 6 about here.]

578 **3.3 | Metabarcoding Pollen Identification**

579 The initial classification of sequences which were made by BLAST were reviewed programmatically, using
580 predicted presence of the species (from spatial modelling), modelled flowering time (from temporal mod-
581 elling), and taxonomy (from existing sources). A sequential process was utilized which reassigned sequences
582 based on binary combinations of the factors above (Appendix 15). Given the relative sparsity of the number,
583 and relatedness, of species represented in the sequence database this was performed to: 1) Identify locally
584 present species represented by surrogates in the DB 2) Reduce false classifications of focal species 3) Identify
585 high confidence sequence matches. Of the top ten taxa which were identified by BLAST for the 680 distinct
586 records, 55.4% of the reads were classified to a species representing 48.3% of all classified reads, 41.9% of the
587 reads were classified to genus representing 48.3% of all classified reads, and 0% of the records were classified
588 to family.

589 Of the 0 classifications which were assigned to genera without any species predicted by spatial analyses, were
590 investigated by hand after post-processing steps. These were all assigned via post-processing conditions (:
591 , Appendix 15). These were manually assigned to a variety of ranks, occasionally to genus - 0, and species
592 - 0, by consultation of the alpha-taxonomic literature (Sadeghian *et al.* (2015), Sennikov & Kurtto (2017),
593 Rabeler & Wagner (2016), Pusalkar & Singh (2015), Moore & Bohs (2003), Weber (1998)).

594 To determine at which level species in pollen loads could be detected the results of light microscopy were
595 compared to the molecular results. The pollen samples contained three morphotypes which could readily be
596 identified via microscopy. Two of these mapped to the clades (Boraginaceae & Heliantheae Alliance), and one
597 to a Asteraceae less the Heliantheae Aliiance (Sunflower Tribe) (Boraginaceae grains were detected in 92.3%
598 of samples where the proportion of target grains were between 0.01-1 (n = 13 Mdn = 0.663). Asteraceae type

599 1, non-helianthoids, were detected in 50% of samples where the proportion of target grains were between
600 0.001-0.01 ($n = 4$ Mdn = 0.001) Asteraceae type 2, Helianthoids, were detected in 33.3% of samples where
601 the proportion of target grains were between 0.001-0.01 ($n = 6$ Mdn = 0.005); however, Asteraceae were
602 detected in 80% of samples where the proportion of target grains were between 0.001-0.01 ($n = 10$ Mdn =
603 0.003). Both morphotypes of Asteraceae pollen were detected in 100% of samples where the proportion of
604 target grains were between 0.01-1 ($n = 2$ Mdn = 0.338).

605 [Figure 7 about here.]

606 To detect whether the sequencing reads were semi-quantitative the subset of all pollen morphotypes distin-
607 guishable by microscopy were compared to the sequence reads. In all instances sequence reads were pooled
608 to the highest taxonomic rank associated with the morphotype, e.g. if both species of *Mertensia* Huth, or
609 one species and read only classified to genus were present in a sample, the reads were summed. The total
610 percentage of the ten most abundant grains per sample were then were then relativized to constitute the
611 entire sample.

612 The relationship between the number of pollen grains in a sample and the number of sequence reads is roughly
613 linear, where grains which are present in trace amounts are overestimated by sequence counts, while grains
614 present in high amounts are underestimated. This is likely due to the proportion of high false positives which
615 occur in the classification process with next-generation sequencing (Bell *et al.* (2021)). There was evidence
616 of a strong correlation between the proportion of grains per morphotype and the number of sequences per
617 group (0.426, $p < 0.0001$, $n = 32$).

618 A number of suspected false positives were returned by the sequence classified, in all instances these were
619 species present in low abundances at the field station which were also returned with a common species in the
620 same family. To ascertain the extent to which records of multiple species in a family, which were suspected
621 to be sampling artefacts occurred in molecular samples an index of similarity, ala Jaccard, the affinity index
622 was used to assess co-occurrence (Mainali *et al.* (2022), Mainali & Slud (2022)). Numerous taxa from the
623 family Ranunculaceae Jussieu (*Caltha* L. sp., *Thalictrum* L. spp., *Trollius* L. sp., *Aquilegia* L. spp.), had α
624 scores which indicated that they are only present when a more common confamilial taxa *Delphinium barbeyi*
625 (Huth) Huth *nuttallianum* Pritz. were recorded. A similar relationship was observed in the Hydrophyllaceae
626 R.Br. with samples placed in *Nemophila* Nutt., which only occurred when the more abundant *Hydrophyllum*
627 L. species were present. The flower size and sparsity of *Nemophila breviflora* A. Gray make it unlikely to
628 be visited by Bumble Bees, and it is a false positive. The floral morphology and orientation of flower of
629 *Thalictrum* spp. also makes them unlikely to be visited, and while evidence of visits to *Caltha* and *Trollius*

630 are lacking, due to the association between the reads these results appear unlikely.

631 4 | DISCUSSION

632 We have demonstrated how the Angiosperms533 hyb-seq probes may be used for plant barcoding in a
633 metagenomic context (Johnson *et al.* (2019), Hollingsworth *et al.* (2016)). This was exemplified in an
634 ecologically relevant scenario, where the results have immediate implications for natural history guided
635 fundamental science and land management. The test pollen loads contained a number of closely related
636 taxa, some in notoriously morphologically difficult clades with rapid rates of diversification (e.g. *Mertensia*,
637 *Lupinus* L.), at naturally occurring proportions (Nevado *et al.* (2016), Nazaire & Hufford (2014)). We
638 incorporated spatial and temporal approaches for creating custom sequence databases an approach which
639 is readily applicable to any lab group with the capacity to perform next-generation sequencing across the
640 entirety of multiple continents, and which we expect to be highly beneficial in many study areas. By
641 combining insights from these novel approaches with an extensive observational field based study we show
642 how these methods may be applied to test a variety of hypotheses related to ecological interactions.

643 The SDM's which we generated, with relatively few occurrence records and few modelling iterations, per-
644 formed beyond expectations, likely due to the utility of the predictor variables and strong alignment of
645 vegetation by orographic precipitation in the study area. However, we had difficulties in evaluating our
646 predictions in an operational context. We utilized the database query approach, to only model species with
647 a high probability of not being dispersal limited to the focal area, and focused on a relevant subset of many
648 of these species ranges to reduce the contributions of range wide adaptions on habitat (Sork (2018), Joshi
649 *et al.* (2001)). While the models worked well compared to both test, and validation with external point
650 data, moving from points to polygon features was more difficult. We were able to compare our results to
651 1) a Flora, 2) lists of plants used by Bumble Bees at plots; the former inappropriate in that it contained
652 a great number of species which we sought to use modelling to reduce *e.g.* all strictly alpine species, and
653 the latter inappropriate in that it contained only species relevant to *Bombus* but had no official 'absence'
654 data. Further given the, size of the minimum spanning tree which we extracted points to, a formal floristic
655 inventory would still be a time intensive process. Accordingly, we expect the real results of our data lay
656 somewhere in between these two evaluations; with an excess of species predicted present (Dubuis *et al.*
657 (2011), Calabrese *et al.* (2014), Pinto-Ledezma & Cavender-Bares (2021)), but few enough that they lend
658 themselves to metabarcoding. We observe that our models seemed very capable of effectively identifying
659 alpine species and removing them in binomial contexts.

660 In regards to the case study, our results indicate the overall information gleaned from observations of queen
661 Bumble Bee foraging and analysis of pollen records are largely congruent. Relaxing concerns regarding dif-
662 ferences between the broad insights gleaned from observational, as compared to data derived from the pollen
663 records (Barker & Arceo-Gomez (2021), Zhao *et al.* (2019), Alarcón (2010)). In general when interaction
664 networks are considered at coarse levels, such as the duration of a season, our perceptions regarding the
665 generality of interactions at smaller time scales may be inflated relative to the actualized interactions within
666 them, e.g. a week (CaraDonna & Waser (2020)). These results indicate a possibility that at even finer levels
667 *Bombus* display high amounts of floral fidelity within foraging bouts, an observation which implies that part
668 of the reason for the high efficiency of *Bombus* as a pollinator might partially be related to their lack of
669 movement of hetero-specific pollen (Broosi & Briggs (2013), Ashman & Arceo-Gómez (2013), Galloni *et al.*
670 (2008), Broosi (2016)). The mechanisms behind this observed fidelity are likely related to pollen nutritional
671 values, specifically high concentrations of protein, and the absence of particular amino acids required for
672 larval development in other flower more commonly used by workers (Genissel *et al.* (2002), Tasei & Aupinel
673 (2008), Goulson *et al.* (2005), Goulson *et al.* (2008a), Hanley *et al.* (2008)).

674 [Table 7 about here.]

675 Also regarding the case study, some foraging preferences of *Bombus*, at this field site and across several
676 localities globally emerge. These suggest the need for land managers to maintain relatively high amounts
677 of members of the plant families Fabaceae, Boraginaceae, and Ranunculaceae, in Western North American
678 montane landscapes (Goulson *et al.* (2005), Goulson (2010), Liang *et al.* (2021), Bontsutsnaja *et al.* (2021)).
679 Numerous historic practices reduce the ability of many landscapes to support stable populations of *Bombus*.
680 Historic livestock grazing involved the removal of many species known to have compounds toxic to cattle.
681 In particular, the removal of locoweeds (*Astragalus* & *Oxytropis*) and larkspurs (*Delphinium*) were common
682 across public lands administered by the U.S. Forest Service (Ralphs & Ueckert (1988), Aldous (1919), Ralphs
683 *et al.* (2003)). Further actions, generally initiated by early settlers, involved the channelization and incising
684 of streams, culling of beavers, and leaving cattle concentrated on higher order stream banks, processes which
685 lowered water tables and reduced the extent of stream-associated wetlands and the mesic meadows fringes
686 which provide habitat for many species of ‘tall’ *Mertensia* (e.g. *M. ciliata*), to an extent *Delphinium barbeyi*
687 and many species of native *Trifolium* (Dahl (1990), Naiman *et al.* (1988), Belsky *et al.* (1999), Cooke &
688 Reeves (1976)). Fire suppression resulted in the succession of many Aspen (*Populus tremuloides*) groves
689 to Conifer stands, decreasing the mosaic of age structured habitats in many landscapes, adversely effecting
690 habitat for tall *Mertensia* species and several species of *Delphinium* (Brewen *et al.* (2021), Keane (2002)).

691 Finally the effects of Nitrogen deposition, given the West's rapidly growing population still pose adverse
692 effects on the abundance of species of Fabaceae at urban-rural interfaces (see Stevens *et al.* (2018), Fenn
693 *et al.* (2003)). Current solutions to the above issues, involve targeted burns, reintroduction of beavers
694 and beaver habitat analogs, and the possibility of re-seeding a variety of 'locoweeds' and 'larkspurs' in
695 areas now seldom used, or only used for early, grazing. The highly enthusiastic response of land managers,
696 and homeowners, to plant *Asclepias*, using genetically appropriate materials, to improve Monarch Butterfly
697 (*Danaus plexippus*) habitat provides an effective framework for the latter (Oberhauser *et al.* (2015), Basey
698 *et al.* (2015)).

699 [Table 8 about here.]

700 We have concerns regarding the number of persons training to become and practice botany, and grave
701 concerns regarding the funding mechanisms for floristic and field based botanical research and for centralized
702 authorities to produce consensus opinions on alpha taxonomy (Prather *et al.* (2004b), Kramer & Havens
703 (2015), Prather *et al.* (2004a), Crisci *et al.* (2020), Manzano (2021), Stroud *et al.* (2022)). To reduce
704 the effects of a low population density of botanists on the maintenance of and production of Flora's and
705 to foster meta-genomics across landscapes without field stations we utilized Species Distribution Modelling
706 to generate predictive species lists. In this proof-of-concept example we performed several iterations of
707 modelling runs, and several approaches (i.e. the 'linear models', and the 'machine learning'), which took
708 notable amounts of compute power. We suspect the possible deleterious nature of this endeavor may be
709 reduced by: 1) more field surveying by crews will reduce the need to generate as many species 2) fewer
710 runs of models, 3) only running machine learning models which do not require an explicit process to reduce
711 spatial autocorrelation. However, given the time required to perform all aspects of a study, even our amount
712 of computation was negligible. Further, we are very optimistic about the possibility for persons to perform
713 these tasks, as mentioned we utilized roughly only one quarter of the records which were digitally available
714 for presence, and we suspect others will have enough records to perform this process nearly anywhere else in
715 the temperate. In certain scenarios modelling of predicted species via more formally tailored S(tacked)-SDM
716 or J(oint)-SDM approaches may be beneficial (Wilkinson *et al.* (2021), Pinto-Ledezma & Cavender-Bares
717 (2021), Schmitt *et al.* (2017)).

718 Tandem to the lack of continued expertise required to generate and maintain species lists, is the expertise
719 required to continue tracking when major phenological events occur in many plant species at relatively fine
720 scales or under novel climates. Knowledge of these events is currently limited to general time periods of only
721 a handful of phenological events and groups of organisms (e.g. flowering initiation, or trees) (Prather *et al.*

722 (2004a), Li *et al.* (2016)). While many programs and initiatives exist to collect phenological information on
723 subsets of easily identifiable charismatic species to detect major trends in phenology, these capture only a
724 subset of the extent diversity (Betancourt *et al.* (2005), Havens *et al.* (2007)). In many instances it appears
725 that while landscapes respond similarly to environmental variables which predict phenological responses,
726 that individual species vary widely in their responses to similar environmental cues, or respond to different
727 cues (Augspurger & Zaya (2020), Xie *et al.* (2015), Xie *et al.* (2018), CaraDonna *et al.* (2014)). As can
728 be seen here, predictions of when a single, major phenological event occurs is already data limited. A more
729 promising approach for the tropics may lay in utilizing circular statistics (Park *et al.* (2022)).

730 The nearly complete Plant and Fungal Tree of Life (PAFTOL) will provide a comprehensive phylogenetic
731 backbone of the entire plant kingdom, and the inclusion of A353 probes with lineage specific probe sets is
732 common in producing massive genetic datasets (Baker *et al.* (2021b)). We predict that the A353 probes
733 which it is utilizing to work nearly immediately for DNA barcoding of whole plant material, and that more
734 elaborate validation studies in controlled metabarcoding settings, utilizing existing experimental designs,
735 will have favorable results (Bell *et al.* (2017b), Bell *et al.* (2019), Bell *et al.* (2021), Lamb *et al.* (2019)). In
736 particular the harvesting of loci with more variation in certain lineages, and or with more variable flanking
737 regions, will prove promising for identifying closely related plant material. We suspect that conserved reaches
738 of genes resulted in the high amounts of reads in somewhat obscure species. Given that the A353 loci are
739 nuclear, single copy, and a variety are present the possibility of identifying target loci for quantitative
740 purposes is high, without continual PCR enrichment is possible; this would align with relatively high efficacy
741 of WGS (Lang *et al.* (2019), Peel *et al.* (2019), Bell *et al.* (2021)). Recent evidence indicates that the
742 potential for identifying nearly cryptic taxa and even infra-specific inference, of either whole plant material,
743 and perhaps in metagenomic context are possible (Ottenlips *et al.* (2021), Wenzell *et al.* (2021), Loke et
744 al. in prep, Slimp *et al.* (2021), Beck *et al.* (2021)). We further believe that in synthetic phylogenetic
745 trees - with incorporation of NGS backbones - will allow in automatic reassignment of reads as a function of
746 phylogenetic distance with measures of uncertainty (Hinchliff *et al.* (2015), Smith & Brown (2018), Baker
747 *et al.* (2021a)).

748 5 | CONCLUSION

749 We believe that the combination of spatial and temporal models, united and guided by localized natural
750 history knowledge, provides the essential components of a framework for approaching the coarse elucidation
751 of ecological interactions using DNA Barcoding. Herein we crudely utilized this thinking via binary outcomes,

752 should a species predicted be predicted present or not? Is it unequivocally flowering or not? Myriad data
753 show biological systems and ecological interactions have more variance than can be reasonably discretely
754 parsed. We expect that within a framework developed from our preliminary works studies of pollinator
755 behavior may be enacted via this approach at a landscape level, e.g. the scale of an entire drainage basin
756 such as the Gunnison which is quickly becoming one of the worlds few model ecosystems. We hope that the
757 A353 probes as tools for metabarcoding play a role in these endeavors.

758 **AUTHOR CONTRIBUTIONS:** R.C.B conducted botanical collections, conducted all molecular lab
759 work, lead all analyses, and writing. J.E.O conceived, designed, and conducted ecological fieldwork, assisted
760 with analyses, and writing. P.J.C conducted ecological fieldwork, assisted with ecological analyses, and
761 writing. E.J.W. prepared, imaged, and collected trait data on pollen reference slides, and assisted with
762 analysis of trait data and writing a dichotomous key. S.T. assisted with spatial analyses and writing. J.B.F.
763 conceived, and designed lab work, analyses, and integration of approaches, assisted with writing, and secured
764 funding for molecular work.

765 **ACKNOWLEDGMENTS:** Nyree Zerega for assistance obtaining herbaria loans and accessioning our
766 collections at CHIC. Pat Herendeen for assistance with virtually all aspects of preparing pollen vouchers and
767 the identification process. Hilary Noble, Zoe Diaz-Martinez, Angela McDonnell, & Elena Loke for assistance
768 with genomic library preparation. Ian Breckheimer for sharing the SDM predictor variables. We thank
769 the curators at the following herbaria for supplying tissue: Ben Legler at Stillinger (ID), Charles (Rick)
770 Williams at Ray J. Davis (IDS), (B)Ernie Nelson at Rocky Mountain (RM); and the collectors: D. Knoke,
771 L. Brummer, J. Boyd, C. Davidson, I. Gilman, M. Kirkpatrick, S. McCauley, J. Smith, K. Taylor, & C.
772 Williams. David Giblin & Mare Nazaire for sharing relevant sections of an advanced draft of FNA V. 15.
773 The Bureau of Land Management is thanked as many plant specimens were collected by R.C.B as a partner
774 or contractor to the agency; Sarah Burnett and Lauren Price are thanked for sharing AIM data. Sanda and
775 New England Biotech are gratefully acknowledged for technical support and generously sharing samples.
776 T.C.H. Cole for sharing the Angiosperm Phylogeny 4 colour palette. The Program in Plant Biology and
777 Conservation is thanked for funding. The holdings of the following herbaria were essential for this project:
778 AK, ALTA, ASU, BABY, BC, BM, BMO, BOON, BRIT, CANB, CAS, CHSC, CM, CMN, CNS, COLO,
779 CONN, CS, CSU, DAV, DBG, DES, ENCB, F, FR, G, GH, GZU, IAC, K, KR, KSP, KSTC, KU, LD, LOB,
780 LSU, MA, MACF, MEL, MICH, MIL, MIN, MNHN, MO, MT, MW, NCSC, NSW, NY, O, OBI, PI, RBG,
781 RSA, SD, SDSU, SFV, TENN, TRT, UA, UAC, UAM, UAZ, UBC, UBC, UCR, UCS, UCSB, UMO, UNM,
782 UPS, US, USCH, USF, USU, UTEP, UWBM, V, VT, W, WSCO, WU, XAL, YPM, Z.

783 **CONFLICT OF INTERESTS** The authors declare no conflicts of interest.

⁷⁸⁴ **PEER REVIEW** The peer review history for this document is available at ...

⁷⁸⁵ **DATA AVAILABILITY STATEMENT** The queries required to download all data used in this project

⁷⁸⁶ are located in... All novel sequencing data are located at NCBI...

⁷⁸⁷ **ORCID**

⁷⁸⁸ Paul CaraDonna <https://orcid.org/0000-0003-3517-9090>

⁷⁸⁹ Jeremie Fant <https://orcid.org/0000-0001-9276-1111>

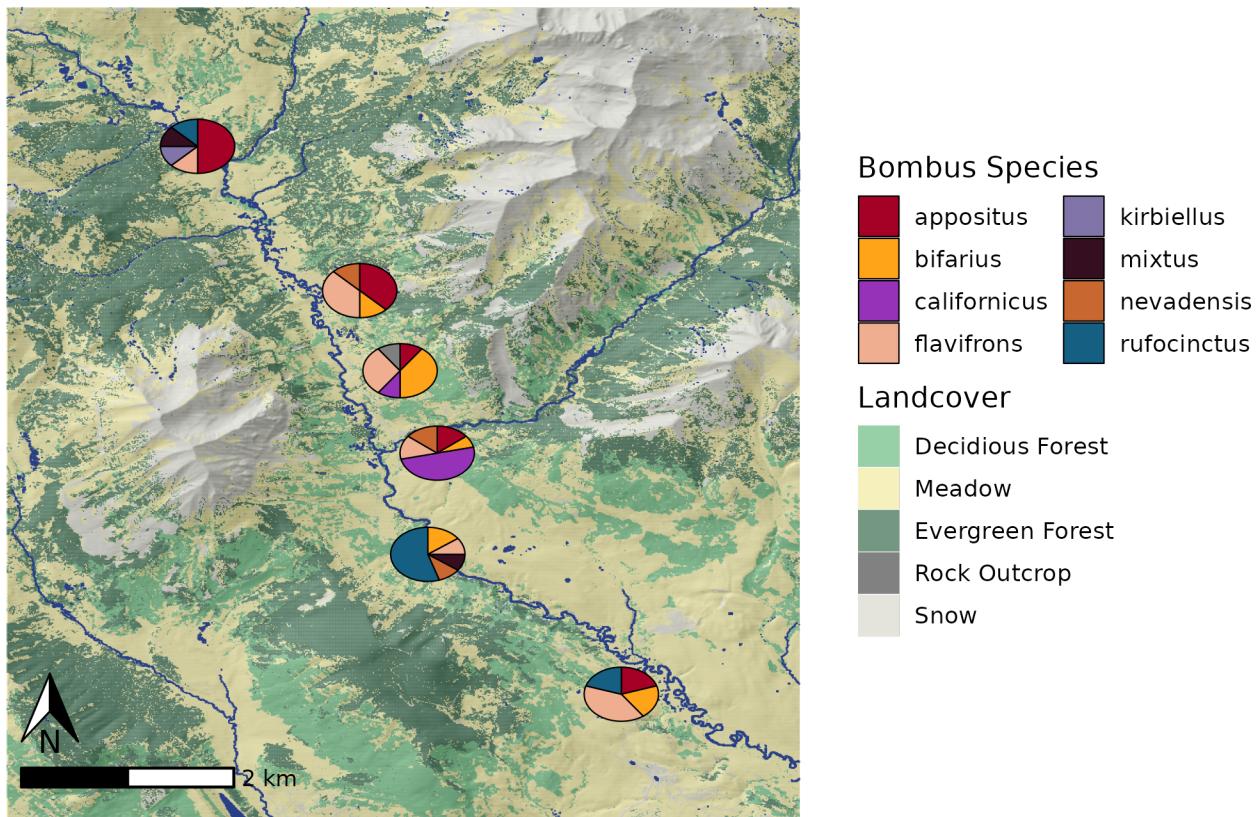
⁷⁹⁰ Jane Ogilvie <https://orcid.org/0000-0001-8546-0417>

⁷⁹¹ Sophie Taddeo <https://orcid.org/0000-0002-7789-1417>

⁷⁹² **References**

⁷⁹³ **Supporting**

Origins of Corbiculae Loads



Upper East River Valley, Colorado

796 Appendix 2 - Species Distribution Models Predictors

Layer	LM	Description	Source
1.	N	Mean annual cloudiness - MODIS	Wilson et al. 2016
2.	Y	Cloudiness seasonality 1 - MODIS	Wilson et al. 2016
3.	N	Cloudiness seasonality 2 - MODIS	Wilson et al. 2016
4.	Y	Cloudiness seasonality 3 - MODIS	Wilson et al. 2016
5.	N	Beginning of the frost-free period	Wang et al.
6.	N	Climatic moisture deficit	Wang et al.
7.	N	Degree-days above 5C	Wang et al.
8.	N	Mean annual precipitation	Wang et al.
9.	Y	Mean annual precipitation as snow	Wang et al.
10.	Y	Temperature seasonality	Wang et al.
11.	Y	2015 Percent Grass/Herbaceous cover - MODIS	(MOD44B)
12.	Y	2015 Percent Tree cover from Landsat 7/8	(GLCF)
13.	Y	Soil probability of bedrock (R Horizon)	SoilGrids
14.	N	Soil organic carbon (Tonnes / ha)	SoilGrids
15.	N	Surface soil pH in H ₂ O	SoilGrids
16.	Y	Surface soil percent sand	SoilGrids
17.	Y	Soil USDA class	SoilGrids
18.	N	Topographic elevation	EarthEnv DEM
19.	Y	Topographic elevation, moving window.	EarthEnv DEM
20.	Y	Topographic percent slope	EarthEnv DEM
21.	Y	Topographic wetness index	EarthEnv DEM
22.	Y	Topographic aspect	EarthEnv DEM
23.	Y	Annual potential solar radiation computed	r.sun
24.	N	Estimated actual (w/-cloud) solar radiation	r.sun / Wilson et al. 2016
25.	Y	Log-transformed distance to surface water	Global Surface Water Explorer
26.	Y	Percent surface water	Global Surface Water Explorer

Table 1: samples used in creating the Reference Library

Taxon	Family	Accession	Pres.	Locality	Date Col.	GenBank	Dist. (km)
<i>Cirsium parryi</i> (A. Gray) Petr.	Asteraceae	CHIC tba	S	Colorado, Gunnison	20.IX.2020	tba	3.6
<i>Cirsium parryi</i> (A. Gray) Petr.	Asteraceae	CHIC tba	S	Colorado, Gunnison	20.IX.2020	tba	3.6
<i>Ericameria parryi</i> (A. Gray) G.L. Nesom & Baird	Asteraceae	CHIC tba	S	Colorado, Gunnison	20.IX.2020	tba	3.6
<i>Erigeron speciosus</i> (Lindley) De Cardolle	Asteraceae	CHIC tba	S	Colorado, Gunnison	18.VII.2020	tba	3.6
<i>Erigeron subtrinervis</i> Rydb. Ex Porter & Britton	Asteraceae	CHIC tba	S	Colorado, Gunnison	20.VII.2020	tba	3.6
<i>Helianthella quinquenervis</i> (Hook.) A. Gray	Asteraceae	CHIC tba	S	Colorado, Gunnison	18.VII.2020	tba	3.6
<i>Helianomeris multiflora</i> Nutt.	Asteraceae	CHIC tba	S	Colorado, Gunnison	18.VII.2020	tba	3.6
<i>Heterotheca villosa</i> (Pursh) Shinners	Asteraceae	CHIC tba	S	Colorado, Gunnison	20.IX.2020	tba	3.6
<i>Senecio sera</i> Hook.	Asteraceae	CHIC tba	P	Idaho, Idaho	26.VII.2020	tba	105.0
<i>Symplytrichum foliacium</i> (Lindl. Ex D.C.) G.L. Nesom	Asteraceae	CHIC tba	S	Illinois, McHenry	28.VII.2020	tba	1624.6
<i>Taraxacum officinale</i> F.H. Wigg.	Asteraceae	CHIC tba	S	Illinois, McHenry	28.VII.2020	tba	1624.6
<i>Mertenia ciliata</i> (James ex Torr.) G. Don	Boraginaceae	ID 1754185	S	Idaho, Valley	18.VI.2018	tba	979.3
<i>Mertenia ciliata</i> (James ex Torr.) G. Don	Boraginaceae	ID 169837	P	Idaho, Adams	10.VII.2014	tba	991.5
<i>Mertensia fusiformis</i> Greene	Boraginaceae	RMH 720522	P	Colorado, Gunnison	7.VI.1997	tba	44.8
<i>Campomanesia rotundifolia</i> L.	Boraginaceae	RMH 720600	P	Colorado, Gunnison	9.VII.1997	tba	38.9
<i>Lathyrus lanszwertii</i> Kellogg var. leucanthus (Ryd.) Dorn	Fabaceae	CHIC tba	S	Colorado, Gunnison	18.VII.2020	tba	3.6
<i>Lathyrus lanszwertii</i> Kellogg var. leucanthus (Ryd.) Dorn	Fabaceae	CHIC tba	S	Colorado, Gunnison	18.VII.2020	tba	3.6
<i>Lupinus argenteus</i> Pursh	Fabaceae	CHIC tba	P	Nevada, Pershing	29.V.2018	tba	971.2
<i>Lupinus argenteus</i> Pursh	Fabaceae	ISU 10387	P	Colorado, Gunnison	29.VI.2010	tba	0.2
<i>Lupinus bakeri</i> Greene	Fabaceae	ISU 10142	P	Colorado, Gunnison	15.VIII.2010	tba	2.6
<i>Vicia americana</i> Muhl. ex Willd.	Fabaceae	CHIC tba	S	Colorado, Gunnison	18.VII.2020	tba	3.6
<i>Vicia americana</i> Muhl. ex Willd. var. minor Hook.	Fabaceae	CHIC tba	S	Montana, Carbon	4.VII.2019	tba	10020.8
<i>Frasera speciosa</i> Douglas ex Griseb	Gentianaceae	RMH 721930	P	Colorado, Gunnison	20.VI.1997	tba	66.2
<i>Frasera speciosa</i> Douglas ex Griseb	Gentianaceae	RMH 719305	P	Colorado, Gunnison	7.VII.1997	tba	19.8
<i>Hydrophyllum capitatum</i> Douglas ex. Benth	Hydrophyllaceae	RMH tba	P	Colorado, Mesa	30.VI.2011	tba	64.6
<i>Hydrophyllum capitatum</i> Douglas ex. Benth	Hydrophyllaceae	RMH tba	P	Colorado, Delta	8.VI.2011	tba	65.3
<i>Hydrophyllum fendleri</i> (Gray) Heller	Hydrophyllaceae	ID 161100	P	Washington, Yakima	9.VI.2008	tba	1429.7
<i>Hydrophyllum fendleri</i> (Gray) Heller	Hydrophyllaceae	ID 164040	P	Idaho, Idaho	27.V.2009	tba	1014.4
<i>Agastache pallidiflora</i> (Heller) Rydberg	Lamiaceae	CHIC tba	S	Arizona, Coconino	17.VII.2020	tba	617.7
<i>Chamerion angustifolium</i> (L.) Holub	Lamiaceae	CHIC tba	S	Colorado, Gunnison	18.VII.2020	tba	3.6
<i>Delphinium barbeyi</i> (Huth) Huth	Ranunculaceae	CHIC tba	S	Colorado, Gunnison	18.VII.2020	tba	3.6
<i>Delphinium nuttallianum</i> Pritz.	Ranunculaceae	ID 166162	P	Idaho, Gem	15.VI.2011	tba	9825.5
<i>Delphinium nuttallianum</i> Pritz.	Ranunculaceae	ID 179376	P	Idaho, Gooding	29.IV.2017	tba	733.7
<i>Potentilla fruticosa</i> Pursh	Rosaceae	CHIC tba	S	Colorado, Gunnison	18.VII.2020	tba	3.6
<i>Potentilla fruticosa</i> Pursh	Rosaceae	CHIC tba	S	Colorado, Gunnison	18.VII.2020	tba	3.6
<i>Potentilla hippiana</i> Lehman.	Rosaceae	CHIC tba	S	New Mexico, Catron	15.VIII.2020	tba	573.8

(Continued on Next Page)

Table 1: samples used in creating the Reference Library (*continued*)

Taxon	Family	Accession	Pres.	Locality	Date Col.	GenBank	Dist. (km)
Potentilla pulcherrima Lehman.	Rosaceae	CHIC tba	S	Colorado, Gunnison	18.VII.2020	tba	3.6

^a Accession includes both Herbarium and Accession number

^b Pres. refers to Preservation method. 'S' denotes silica gel dried, 'P' denotes pressed

^c All Localities are in the United States of America

800 Appendix 4 - All Pollen Reference Slides Used to Establish Morphotypes

Table 1: All Pollen Voucher Slides Consulted

Taxon	Family	Locality	Accession	Type	Prepared by	Date
<i>Cymopterus lemnoides</i> (J.M. Coulter & Rose) Dorn	Apiaceae	CO, Gunnison, Gothic	None	Physical	P.J.C.	2014
<i>Heracleum sphondylium</i> L.	Apiaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Ligusticum porteri</i> J.M. Coulter & Rose	Apiaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Osmorhiza depauperata</i> Phil.	Apiaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Maianthemum stellatum</i> (L.) Link	Asparagaceae	CO, Gunnison, Gothic	None	Physical	P.J.C.	2014
<i>Achillea millefolium</i> L.	Asteraceae	CO, Gunnison, Gothic	None	Physical	P.J.C.	2014
<i>Achillea millefolium</i> L.	Asteraceae	CO, Gunnison, Washington Gulch	tba	Novo	E.J.W.	2021
<i>Acourtia wrightii</i> (A. Gray) Reveal & King	Asteraceae	NV, Clark, Gold Butte	tba	Novo	E.J.W.	2021
<i>Antennaria racemosissima</i> Hook.	Asteraceae	WY, Park, Ishawooa Mesa	tba	Novo	E.J.W.	2021
<i>Arnica latifolia</i> Bong.	Asteraceae	ID, Blaine, Wildhorse Canyon	tba	Novo	E.J.W.	2021
<i>Artemisia scopulorum</i> A. Gray	Asteraceae	CO, Hinsdale, Uncompahgre Peak	tba	Novo	E.J.W.	2021
<i>Canadanthus modestus</i> (Lindl.) G.L. Nesom	Asteraceae	ID, Idaho, Whiskey Creek	tba	Novo	E.J.W.	2021
<i>Chaenactis douglasii</i> (Hook.) Hook. & Arn.	Asteraceae	MT, Carbon, Pryor Mtn. Rd.	tba	Novo	E.J.W.	2021
<i>Erigeron corymbosus</i> Nutt.	Asteraceae	MT, Carbon, Pryor Mtn. Rd.	tba	Novo	E.J.W.	2021
<i>Erigeron flagellaris</i> A. Gray	Asteraceae	CO, Gunnison, Gothic	None	Physical	P.J.C.	2014
<i>Erigeron speciosus</i> (Lind.) DC.	Asteraceae	CO, Gunnison, Gothic	None	Physical	P.J.C.	2014
<i>Erigeron speciosus</i> (Lindl.) DC.	Asteraceae	CO, Gunnison, Washington Gulch	tba	Novo	E.J.W.	2021
<i>Helianthella quinquenervis</i> (Hook.) A. Gray	Asteraceae	CO, Gunnison, Gothic	None	Physical	P.J.C.	2014
<i>Heliotropium multiflorum</i> Nutt.	Asteraceae	CO, Gunnison, Gothic	None	Physical	P.J.C.	2014
<i>Heliotropium multiflorum</i> Nutt.	Asteraceae	CO, Gunnison, Washington Gulch	tba	Novo	E.J.W.	2021
<i>Heterotheca villosa</i> (Pursh) Shinners	Asteraceae	CO, Gunnison, Gothic	None	Physical	P.J.C.	2014
<i>Heterotheca villosa</i> (Pursh) Shinners	Asteraceae	AZ, Coconino, Lake Mary Rd. & 209	tba	Novo	E.J.W.	2021
<i>Hymenoxys hoopesii</i> (A. Gray) Bierner	Asteraceae	CO, Gunnison, Gothic	None	Physical	P.J.C.	2014
<i>Hymenoxys rusbyi</i> (A. Gray) Cockerell	Asteraceae	AZ, Coconino, Lake Mary Rd. & 209	tba	Novo	E.J.W.	2021
<i>Ionactis stenomeria</i> (A. Gray) Greene	Asteraceae	ID, Idaho, Marshall Mountains	tba	Novo	E.J.W.	2021
<i>Senecio hydrophilus</i> Nutt.	Asteraceae	ID, Custer, E. fl. Salmon River	tba	Novo	E.J.W.	2021
<i>Senecio integrerrimus</i> Nutt.	Asteraceae	CO, Gunnison, Gothic	None	Physical	P.J.C.	2014
<i>Senecio serrula</i> Hook.	Asteraceae	CO, Gunnison, Washington Gulch	tba	Novo	E.J.W.	2021
<i>Senecio wootonii</i> Greene	Asteraceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Solidago lepida</i> DC.	Asteraceae	ID, Idaho, American River	tba	Novo	E.J.W.	2021
<i>Sympotrichum foliacum</i> (Lindl. ex DC.) G.L. Nesom	Asteraceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Sympotrichum subspicatum</i> (Nees) G.L. Nesom	Asteraceae	ID, Custer, E. fl. Salmon River	tba	Novo	E.J.W.	2021
<i>Taraxacum officinale</i> F.H. Wigg	Asteraceae	CO, Gunnison, Gothic	None	Physical	P.J.C.	2014
<i>Taraxacum officinale</i> F.H. Wigg	Asteraceae	IL, McHenry, Barrington	tba	Novo	E.J.W.	2021
<i>Lappula squarrosa</i> (Retz.) Dumort.	Boraginaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Mertensia ciliata</i> (James ex Torr.) G. Don	Boraginaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Mertensia fusiformis</i> Greene	Boraginaceae	CO, Gunnison, Gothic	None	Physical	P.J.C.	2014
<i>Boechera</i>	Brassicaceae	NV, Washoe, Mt. Rose	tba	Novo	E.J.W.	2021
<i>Boechera stricta</i> (Graham) Al-Shehbaz	Brassicaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Cardamine cordifolia</i> A. Gray	Brassicaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Draba aurea</i> Vahl. Ex Hornem	Brassicaceae	CO, Gunnison, Gothic	None	Physical	P.J.C.	2014

(Continued on Next Page)

Table 1: All Pollen Voucher Slides Consulted (*continued*)

Taxon	Family	Locality	Accession	Type	Prepared by	Date
<i>Draba spectabilis</i> Greene	Brassicaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Erysimum capitatum</i> (Douglas ex Hook.) Greene	Brassicaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Lepidium montanum</i> Nutt.	Brassicaceae	NM, Catron, Zuni Dry Lake	tba	Novo	E.J.W.	2021
<i>Smelowskia americana</i> Rydb.	Brassicaceae	ID, Blaine, Pioneer Mtns Crest	tba	Novo	E.J.W.	2021
<i>Thlaspi arvense</i> L.	Brassicaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Campanula rotundifolia</i> L.	Campanulaceae	CO, Gunnison, Gothic	None	Physical	P.J.C.	2014
<i>Campanula rotundifolia</i> L.	Campanulaceae	CO, Gunnison, Washington Gulch	tba	Novo	E.J.W.	2021
<i>Downingia</i>	Campanulaceae	CA, Nevada, Truckee Meadows	tba	Novo	E.J.W.	2021
<i>Lonicera involucrata</i> (Richardson) Banks ex Spreng.	Caprifoliaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Minuartia nuttallii</i> (Pax.) Briq.	Caryophyllaceae	ID, Blaine, Wildhorse Canyon	tba	Novo	E.J.W.	2021
<i>Stellaria longifolia</i> Muh. Ex. Willd.	Caryophyllaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Parnassia kotzebuei</i> Cham. ex Spreng	Celastraceae	ID, Lemhi, Terrace Lakes	tba	Novo	E.J.W.	2021
<i>Sedum lanceolatum</i> Torr.	Crassulaceae	ID, Lemhi, Terrace Lakes	tba	Novo	E.J.W.	2021
<i>Astragalus robbinsii</i> (Oakes) A. Gray	Fabaceae	ID, Custer, Lake Creek	tba	Novo	E.J.W.	2021
<i>Glycyrrhiza lepidota</i> Nutt.	Fabaceae	ID, Butte, Warm Springs Creek	tba	Novo	E.J.W.	2021
<i>Lathyrus eucomus</i> Butters & H. St. John	Fabaceae	NM, Catron, Zuni Salt Lake	tba	Novo	E.J.W.	2021
<i>Lathyrus lanszwertii</i> var. <i>leucanthus</i> (Rydb.) Dorn	Fabaceae	CO, Gunnison, Gothic	None	Physical	P.J.C.	2014
<i>Lathyrus lanszwertii</i> var. <i>leucanthus</i> (Rydb.) Dorn	Fabaceae	CO, Gunnison, Washington Gulch	tba	Novo	E.J.W.	2021
<i>Lupinus argenteus</i> Pursh	Fabaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Lupinus argenteus</i> Pursh	Fabaceae	CO, Gunnison, Washington Gulch	tba	Novo	E.J.W.	2021
<i>Lupinus argenteus</i> Pursh	Fabaceae	NV, Pershing, Star Peak Canyon	tba	Novo	E.J.W.	2021
<i>Lupinus crassus</i> Payson	Fabaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Lupinus sericeus</i> Pursh	Fabaceae	CO, Gunnison, Gothic	None	Physical	P.J.C.	2014
<i>Melilotus albus</i> Medik.	Fabaceae	NM, Catron, Hwy 159	tba	Novo	E.J.W.	2021
<i>Trifolium hybridum</i> L.	Fabaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Trifolium pratense</i> L.	Fabaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Vicia americana</i> Muhl. Ex Willd.	Fabaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Vicia americana</i> Muhl. Ex Willd.	Fabaceae	UT, Cache, Spawn Creek	tba	Novo	E.J.W.	2021
<i>Vicia americana</i> Muhl. Ex Willd.	Fabaceae	CO, Gunnison, Washington Gulch	tba	Novo	E.J.W.	2021
<i>Geranium</i>	Geraniaceae	NM, Catron, Jim Smith TH. Rd.	tba	Novo	E.J.W.	2021
<i>Geranium richardsonii</i> Fisch. Trautv.	Geraniaceae	CO, Gunnison, Washington Gulch	tba	Novo	E.J.W.	2021
<i>Phacelia</i> sp.	Hydrophyllaceae	NV, Nye, Toiyabe Crest	tba	Novo	E.J.W.	2021
<i>Iris missouriensis</i> Nutt.	Iridaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Agastache palliflora</i> (A. Heller) Rydb.	Lamiaceae	AZ, Coconino, Lake Mary Rd. & 209	tba	Novo	E.J.W.	2021
<i>Erythronium grandiflorum</i> Pursh	Liliaceae	CO, Gunnison, Gothic	None	Physical	P.J.C.	2014
<i>Linum lewisii</i> Pursh	Linaceae	CO, Gunnison, Washington Gulch	tba	Novo	E.J.W.	2021
<i>Zigadenus elegans</i> Pursh	Melanthiaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Anticlea elegans</i> (A. Gray) Zomlefer & Judd	Melanthiaceae	ID, Blaine, Wildhorse Canyon	tba	Novo	E.J.W.	2021
<i>Cistanthe/Calyptidium</i>	Montiaceae	NV, Nye, Toiyabe Crest	tba	Novo	E.J.W.	2021
<i>Chamerion angustifolium</i> (L.) Holub	Onagraceae	CO, Gunnison, Washington Gulch	tba	Novo	E.J.W.	2021
<i>Epilobium obcordatum</i> A. Gray	Onagraceae	ID, Lemhi, Bighorn Crags	tba	Novo	E.J.W.	2021
<i>Castilleja miniata</i> Douglas ex Hook.	Orobanchaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010

(Continued on Next Page)

Table 1: All Pollen Voucher Slides Consulted (*continued*)

Taxon	Family	Locality	Accession	Type	Prepared by	Date
<i>Castilleja sulphurea</i> Rydb.	Orobanchaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Pedicularis groenlandica</i> Retz.	Orobanchaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Pedicularis racemosa</i> Douglas ex Benth.	Orobanchaceae	CO, Gunnison, Washington Gulch	tba	Novo	E.J.W.	2021
<i>Corydalis aurea</i> Willd.	Papaveraceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Erythranthe guttata</i> (DC.) G.L. Nesom	Phrymaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Penstemon wilcoxii</i> Rydb.	Plantaginaceae	MT, Missoula, Mission Mtns vic.	tba	Novo	E.J.W.	2021
<i>Collomia linearis</i> Nutt.	Polemoniaceae	CO, Gunnison, Washington Gulch	tba	Novo	E.J.W.	2021
<i>Phlox condensata</i> (A. Gray) E.E. Nelson	Polemoniaceae	CO, Hinsdale, Uncompahgre Peak	tba	Novo	E.J.W.	2021
<i>Polemonium foliosissimum</i> A. Gray	Polemoniaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Polemonium occidentale</i> Greene	Polemoniaceae	ID, Custer, Lake Creek	tba	Novo	E.J.W.	2021
<i>Polemonium viscosum</i> Nutt.	Polemoniaceae	CO, Gunnison, Gothic	None	Physical	P.J.C.	2010
<i>Bistorta bistortoides</i> (Pursh) Small	Polygonaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Eriogonum</i> spp.	Polygonaceae	NV, Washoe, Hwy 445	tba	Novo	E.J.W.	2021
<i>Polygala barbeyana</i> Chodat	Polygonaceae	NM, Eddy, Yeso Hills	tba	Novo	E.J.W.	2021
<i>Polygonum polygaloides</i> L.	Polygonaceae	MT, Missoula, Mission Mtns	tba	Novo	E.J.W.	2021
<i>Androsace filiformis</i> Retz.	Primulaceae	ID, Custer, Bradshaw Creek	tba	Novo	E.J.W.	2021
<i>Aquilegia coerulea</i> E. James	Ranunculaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Aquilegia coerulea</i> E. James	Ranunculaceae	CO, Gunnison, Washington Gulch	tba	Novo	E.J.W.	2021
<i>Aquilegia elegantula</i> Greene	Ranunculaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Delphinium barbeyi</i> (Huth) Huth	Ranunculaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Delphinium nuttallianum</i> Pritz. Ex Walp.	Ranunculaceae	CO, Gunnison, Gothic	None	Physical	P.J.C.	2014
<i>Ranunculus alismifolius</i> Geyer ex Benth.	Ranunculaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Ranunculus glaberrimus</i> Hook.	Ranunculaceae	ID, Lemhi, Agency Creek	tba	Novo	E.J.W.	2021
<i>Ranunculus inamoenus</i> Greene	Ranunculaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Ranunculus</i> spp.	Ranunculaceae	NV, Washoe, Mt. Rose	tba	Novo	E.J.W.	2021
<i>Thalictrum sparsiflorum</i> Tuzc. Ex Fisch. & C.A. Mey.	Ranunculaceae	ID, Custer, E. fk. Salmon River	tba	Novo	E.J.W.	2021
<i>Dasiphora fruticosa</i> (L.) Rydb.	Rosaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Dasiphora fruticosa</i> (L.) Rydb.	Rosaceae	CO, Gunnison, Washington Gulch	tba	Novo	E.J.W.	2021
<i>Fragaria virginiana</i> Duchesne	Rosaceae	CO, Gunnison, Gothic	None	Physical	P.J.C.	2014
<i>Geum triflorum</i> Pursh	Rosaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Potentilla biennis</i> Greene	Rosaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Potentilla hippiana</i> Lehm.	Rosaceae	NM, Catron, Jim Smith TH.Rd.	tba	Novo	E.J.W.	2021
<i>Potentilla pulcherrima</i> Lehm.	Rosaceae	CO, Gunnison, Gothic	None	Physical	P.J.C.	2014
<i>Potentilla pulcherrima</i> Lehm.	Rosaceae	CO, Gunnison, Washington Gulch	tba	Novo	E.J.W.	2021
<i>Salix bebbiana</i> Sarg.	Salicaceae	ID, Custer, Lake Creek	tba	Novo	E.J.W.	2021
<i>Salix geyeriana</i> Andersson	Salicaceae	ID, Butte, Clyde	tba	Novo	E.J.W.	2021
<i>Mitella stauropetala</i> Piper	Saxifragaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Valeriana occidentalis</i> A. Heller	Valerianaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010
<i>Viola canadensis</i> L.	Violaceae	CO, Gunnison, Gothic	None	Image	H.M.B.	2010

* All Localities are in the United States of America

† Accession refers to whole-plant vouchers, all specimens are deposited at CHIC.

‡ Type Refers to whether both a physical and digital copy exist; 'Image' denotes only digital

§ Date refers to the Date of preparation.

⁸⁰⁴ Appendix 5 - Pollen Dendrogram

- 1a: Pollen shed in clumps (tetrads/polyads); grains generally triangular, with an annulus subtending the porate apertures (go 34)
- 1b: Pollen generally dispersed as single units (monads); grains seldom if ever with annulus.
- 2a: Apertures porate, always lacking colpi
- 3a: grain outline from equatorial view circular
- 4a: Pores distributed along the equator.
- 5a: Pores > 5 (stephanoporate)
- 6a: Ornamentation homobrochate (~ *MENTZELIA*)
- 6b: Ornamentation otherwise (~ *POLYGALA*)
- 5b: Pores < 5 (*CURRENTLY OPEN*)
- 4b: Pores +/- distributed across grain (pantoporate)
- 7a: Ornamentation with striate ornamentation (~ *POLEMONIUM*)
- 7b: Ornamentation otherwise
- 8a: Ornamentation, slightly irregular - without regularly repeating features (scabrate) (~ *STELLARIA*)
- 8b: Ornamentation forming regularly repeating (reticulate) cells of varying shapes.
- 9a: spacing between the grid cells large (lophate), the walls of the cells with another set of projecting ornamentation (~ *OPUNTIA*)
- 9b: spacing between cells small, the wall of the cells without projecting features.
- 10a: Pores extending beyond the reticulate grids (~ *ARENARIA*)
- 10b: Pores extending beyond the reticulate grids (~ *PHLOX*)
- 3b: Outline from equatorial view otherwise (usually slightly triangular)
- 11a: Outline elliptic (*CURRENTLY EMPTY*)
- 11b: Outline not elliptic, grains often with acute, if rounded, angles along sides (e.g., triangular, polygonal) (*EMPTY*)
- 2b: Apertures with colpi, occasionally also with pores in addition (coporate)
- 12a: Grains with bristles tapering to points (echinate), and tri-colporate.
- 13a: Grains uniformly echinate, less the apertures. (Asteraceae 1)
- 13b: Grains with echinate bristles on ridges of lophae (Asteraceae 2)
- 12b: Grains without echinate ornamentation - this lead includes projections with ornamentation with round tips.
- 14a: Grains with either less than 3 apertures, or with two distinct ornamentation types (generally $\frac{1}{2}$ psilate, $\frac{1}{2}$ reticulate).
- 15a. Grains apparently lacking any apertures. (~ *IRIS*)
- 15b. Grains aperturate
- 16a. Ornamentation on one face of grain psilate, the other homobrochate (~ *ZIGADENUS + ANTICLEA*)
- 16b. Ornamentation psilate across both faces of grain (~*ERYTHONIUM*)
- 14b. Grains with either 3 or more apertures, or with an elongated spiral like aperture
- 17a. Grain with spiral like colpi
- 18a. Spiral with deep well-defined furrows (~ *ERYTHRANTHE GUTTATA*, syn. obsolete. *MIMULUS*)
- 18b. Spirals without well-defined grooves, ornamentation evidently perforate (~ *RANUNCULUS ALISMIFOLIUS*)
- 17b. Grains with colpi these not forming irregular spiral motifs.
- 19a. Grains elliptic, essentially perfectly cylindrical along longest axis, except for minor inundations along equatorial region. Apertures, of two types (heteroaperturate). (~ *BORAGINACEAE*)
- 19b Grains shaped similar or not, but never heteroaperturate.
- 20a. From a polar view, grains notably polygonal (hexagonal), also evident when seldom seen from a equatorial view. (~ *PHACELIA*/ maybe *Hydrophyllaceae*, *Hydrophyllum* not sampled)
- 20b. From a polar view, grains not with 6 convex apices

- 21a. Grains elliptic, with a short colporate aperture on each psilate face, the edges of each face and the apices with a distinct (homobrochate) textured ornamentation. (~ POLYGONUM)
- 21b. Grains otherwise, not featuring a mix of ornamentations independent of the apertures.
- 22a. Ornamentation perforate, the three colpi very short, their longest axis parallel to the equator rather than perpendicular. These colpi often times almost appearing to be slightly raised on an annulus like feature (~ LONICERA)
- 22b. Grains not as described in all aspects of the above.
- 23a. Apertures colporate
- 24a. Outline of grain in equatorial view circular, ornamentation smooth. (~MORPHOTYPE A).
- 24b. Grains otherwise
 - 25a Grains distinctly triangular from polar view (go 26)
 - 25b Grains elliptic (go 27)
 - 26a Grains very large, clearly strongly triangular in cross section. (~ GERANIUM)
 - 26b Grains smaller (SIZE), weakly triangular in cross section (~ POTENTILLA/DASIPHORA in part)
 - 27a Grains elliptic to weakly circular (~MORPHOTYPE B)
 - 27b Grains elliptic, much longer pole to pole than across equator.
 - 28a Grains with evident protrusions of the pore, colpi short, scarcely noticeable (~ APIACEAE)
- 23b. Apertures colpate
 - 30a Ornamentation psilate (~MORPHOTYPE C).
 - 30b Ornamentation otherwise
 - 31a Ornamentation homobrochate (~ MORPHOTYPE D)
 - 31b Ornamentation otherwise
 - 32a. Ornamentation bacculate, grains large, (~LINUM)
 - 32b. Ornamentation otherwise
 - 33a. Ornamentation of minor cross-corrugated grooves (fossulate) (~CORYDALIS)
 - 33b. Ornamentation of scarcely perceptible irregular features (scabrate) (~RANUNCULUS IN PART)

34a: Annula subtending the apertures – making grains appear more or less triangular; Pollen often with viscin threads (ONAGRACEAE)

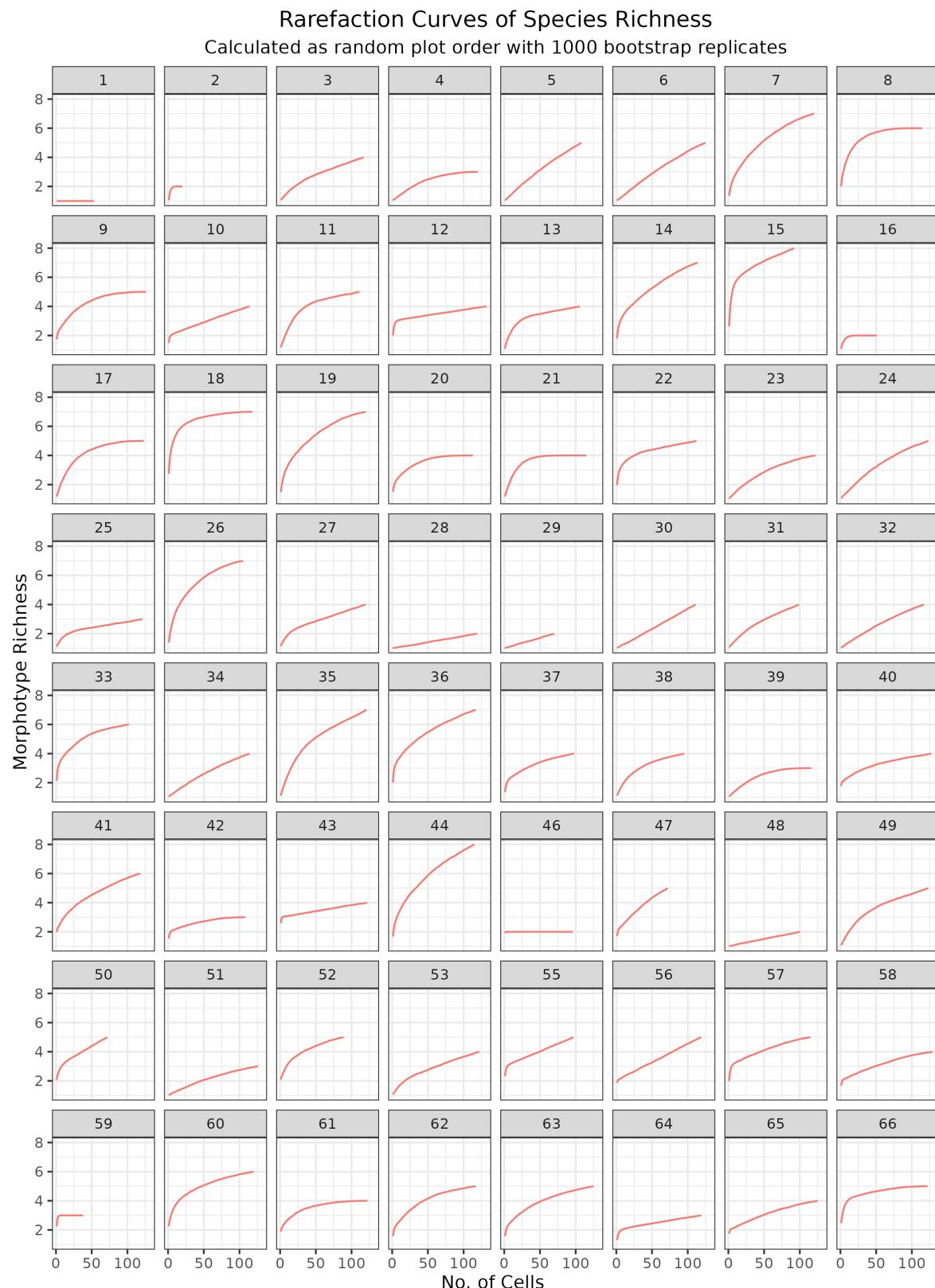
34b: Apertures not annulate – grains appear more or less circular (~ERICACEAE)

Morphotype A: Trifolium, Lupinus, Glycrrhiza, Mitella, Geum

Morphotype B: Lupinus, Lathyrus, Potentilla, Androsace, Bistorta, Vicia

Morphotype C: Jeffersonia, Micranthes, Prunus, Delphinium, Androsace, Penstemon, Orthocarpus, Scutellaria, Aquilegia, Castilleja, Draba)

Morphotype D: Salix, Boechera



Used to assess completeness of subsampling. VEGAN package 'specaccum' function used

Species Richness Abundance Estimating via Hill Numbers ($q = 0$)

Confidence Interval of 99% with 1000 Bootstrap replicates

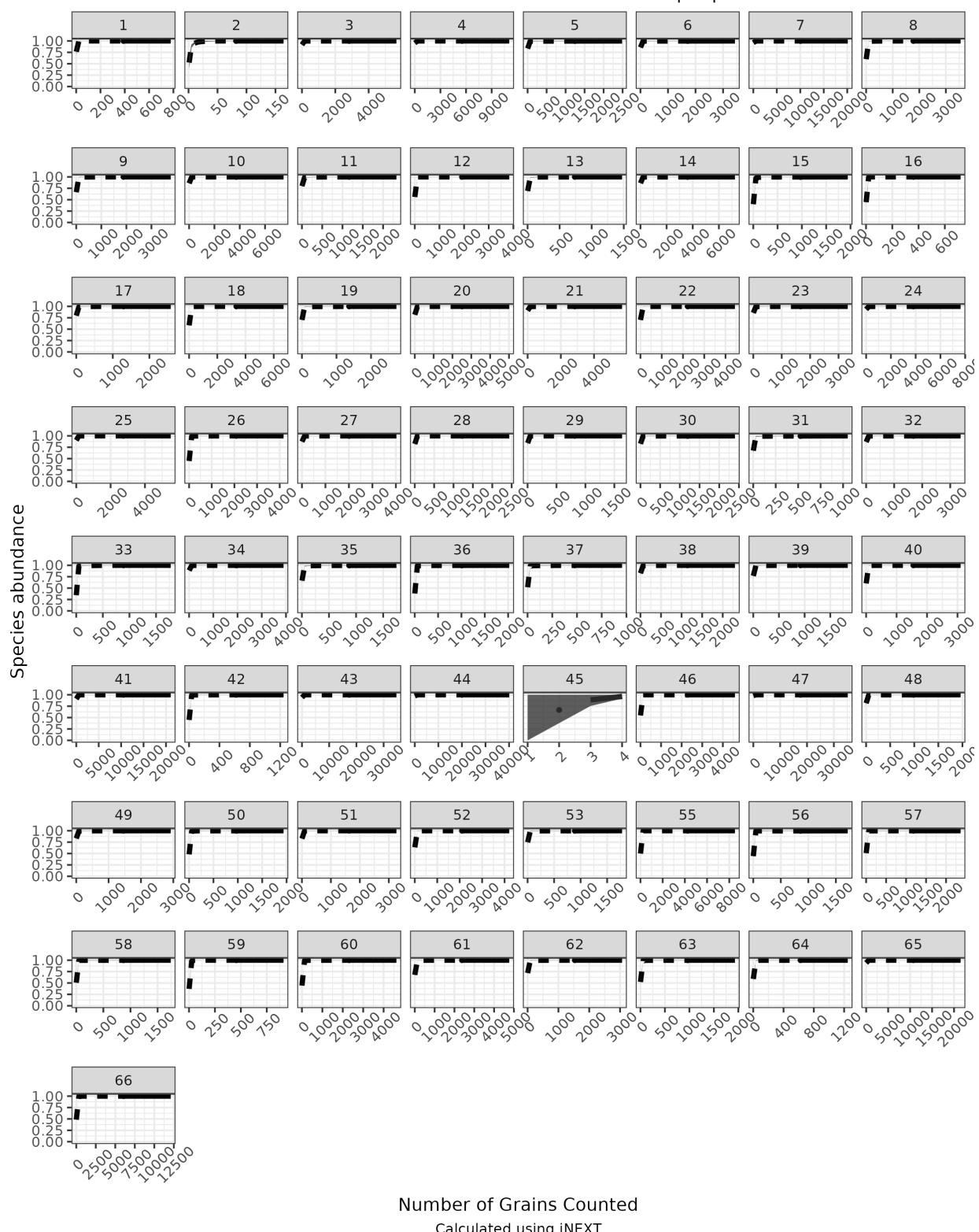


Table 1: All species present in the Reference Sequence Databases
(Kraken and BLAST)

Order	Family	Taxon
Alismatales	Potamogetonaceae	<i>Potamogeton wrightii</i>
Apiales	Apiaceae	<i>Osmorhiza aristata</i>
Asparagales	Amaryllidaceae	<i>Allium stamineum</i>
	Asparagaceae	<i>Streptopus amplexifolius</i>
Asterales	Asteraceae	<i>Anaphalis margaritacea</i> <i>Antennaria carpatica</i> <i>Antennaria dioica</i> <i>Artemisia sibirica</i> <i>Brickellia dentata</i> <i>Chrysanthemus greenei</i> <i>Cirsium pannonicum</i> <i>Cirsium parryi</i> <i>Cirsium vulgare</i> <i>Crepis pygmaea</i> <i>Ericameria parryi</i> <i>Erigeron ecuadoriensis</i> <i>Erigeron grandiflorus</i> <i>Erigeron rosulatus</i> <i>Erigeron uniflorus</i> <i>Helianthella quinquenervis</i> <i>Heterotheca villosa</i> <i>Hieracium avilae</i> <i>Hieracium jubatum</i> <i>Hymenoxys hoopesii</i> <i>Leucanthemum graminifolium</i> <i>Microseris lindleyi</i> <i>Omalotheca supina</i> <i>Packera quercetorum</i> <i>Pseudognaphalium attenuatum</i> <i>Pseudognaphalium frigidum</i> <i>Pseudognaphalium lacteum</i> <i>Pseudognaphalium oxyphyllum</i> <i>Rudbeckia hirta</i> <i>Scabrethia scabra</i> <i>Senecio adenophyllus</i> <i>Senecio algens</i> <i>Senecio apolobambensis</i> <i>Senecio candollei</i> <i>Senecio chionogeton</i> <i>Senecio formosus</i> <i>Senecio funcii</i> <i>Senecio gilliesii</i> <i>Senecio humillimus</i> <i>Senecio nutans</i> <i>Senecio puchei</i> <i>Senecio rufescens</i> <i>Senecio spinosus</i> <i>Senecio tephrosioides</i>

(Continued on Next Page)

Table 1: All species present in the Reference Sequence Databases
(Kraken and BLAST) (*continued*)

Order	Family	Taxon
Boraginales	Campanulaceae	<i>Solidago chilensis</i> <i>Stilpnolepis intricata</i> <i>Symphytum foliaceum</i> <i>Taraxacum cucullatum</i> <i>Taraxacum officinale</i>
		<i>Tonestus lyallii</i>
		<i>Townsendia formosa</i>
		<i>Campanula argaea</i>
		<i>Campanula rotundifolia</i>
	Hydrophyllaceae	<i>Cynoglossum amplifolium</i> <i>Cynoglossum anchusoides</i> <i>Cynoglossum pringlei</i> <i>Mertensia ciliata</i> <i>Mertensia fusiformis</i>
		<i>Hydrophyllum canadense</i>
		<i>Hydrophyllum capitatum</i>
		<i>Hydrophyllum fendleri</i>
		<i>Nemophila menziesii</i>
Caryophyllales	Caryophyllaceae	<i>Arenaria globiflora</i> <i>Arenaria serpyllifolia</i> <i>Cerastium arvense</i> <i>Cerastium lanceolatum</i> <i>Minuartia recurva</i> <i>Odontostemma leucasterium</i> <i>Pseudostellaria heterophylla</i> <i>Sagina procumbens</i> <i>Schizotechium monospermum</i> <i>Shivparvatia glanduligera</i>
		<i>Stellaria graminea</i>
		<i>Stellaria holostea</i>
		<i>Stellaria obtusa</i>
		<i>Rumex induratus</i>
		<i>Rumex spinosus</i>
		<i>Parnassia faberi</i>
		<i>Parnassia palustris</i>
		<i>Paxistima canbyi</i>
		<i>Gaultheria prostrata</i>
Celastrales	Ericaceae	<i>Moneses uniflora</i> <i>Orthilia secunda</i> <i>Vaccinium vitis-idaea</i> <i>Collomia grandiflora</i> <i>Ipomopsis aggregata</i>
		<i>Phlox douglasii</i>
		<i>Primulaceae</i>
		<i>Androsace studiosorum</i>
		<i>Androsace vitaliana</i>
	Fabaceae	<i>Astragalus pelecinus</i>
		<i>Lupinus argenteus</i>
		<i>Lupinus sericeus</i>

(Continued on Next Page)

Table 1: All species present in the Reference Sequence Databases
(Kraken and BLAST) (*continued*)

Order	Family	Taxon
Gentianales	Gentianaceae	<i>Vicia americana</i> <i>Frasera speciosa</i> <i>Gentiana cruciata</i>
Hyphomicrobiales	Xanthobacteraceae	<i>Azorhizobium caulinodans</i>
Lamiales	Lamiaceae	<i>Agastache pallidiflora</i>
Liliales	Colchicaceae	<i>Prosartes smithii</i>
	Liliaceae	<i>Erythronium dens-canis</i>
	Melanthiaceae	<i>Anticlea elegans</i> <i>Veratrum viride</i>
Malpighiales	Hypericaceae	<i>Hypericum perforatum</i>
	Salicaceae	<i>Populus alba</i>
	Violaceae	<i>Viola odorata</i>
Myrtales	Onagraceae	<i>Chamaenerion angustifolium</i> <i>Epilobium canum</i> <i>Epilobium parviflorum</i>
Ranunculales	Berberidaceae	<i>Berberis sibirica</i>
	Papaveraceae	<i>Corydalis aitchisonii</i>
	Ranunculaceae	<i>Actaea heracleifolia</i> <i>Anemone anemonoides</i> <i>Anemone obtusiloba</i> <i>Aquilegia ecalcarata</i> <i>Caltha palustris</i> <i>Delphinium barbeyi</i> <i>Delphinium gracile</i> <i>Delphinium nuttallianum</i> <i>Pulsatilla chinensis</i> <i>Thalictrum thalictroides</i> <i>Thalictrum tuberosum</i> <i>Trollius europaeus</i>
Rosales	Elaeagnaceae	<i>Shepherdia argentea</i>
	Rosaceae	<i>Crataegus bipinnatifida</i> <i>Dasiphora fruticosa</i> <i>Geum ternatum</i> <i>Hedlundia austriaca</i> <i>Holodiscus argenteus</i> <i>Karpatiosorbus devoniensis</i> <i>Micromeles japonica</i> <i>Potentilla anserina</i> <i>Potentilla pulcherrima</i> <i>Potentilla tetrandra</i> <i>Rubus chingii</i>
Sapindales	Sapindaceae	<i>Acer campestre</i>
Saxifragales	Crassulaceae	<i>Rhodiola rosea</i> <i>Sedum nudum</i>
	Grossulariaceae	<i>Ribes rubrum</i>
	Saxifragaceae	<i>Lithophragma parviflorum</i> <i>Saxifraga biflora</i> <i>Saxifraga fortunei</i>

(Continued on Next Page)

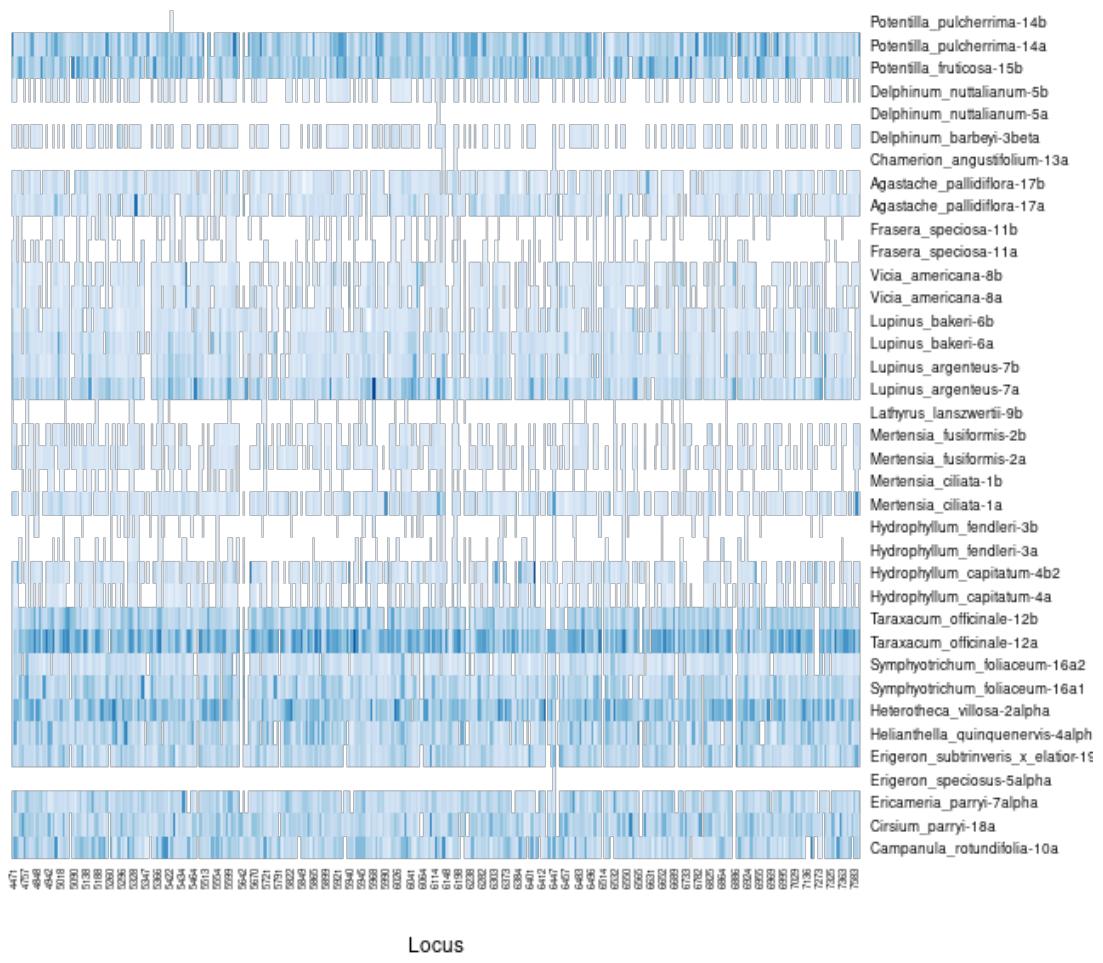
⁸¹⁷ Appendix 10 - All Species in the Sequence Databases (con't)

Table 1: All species present in the Reference Sequence Databases
(Kraken and BLAST) (*continued*)

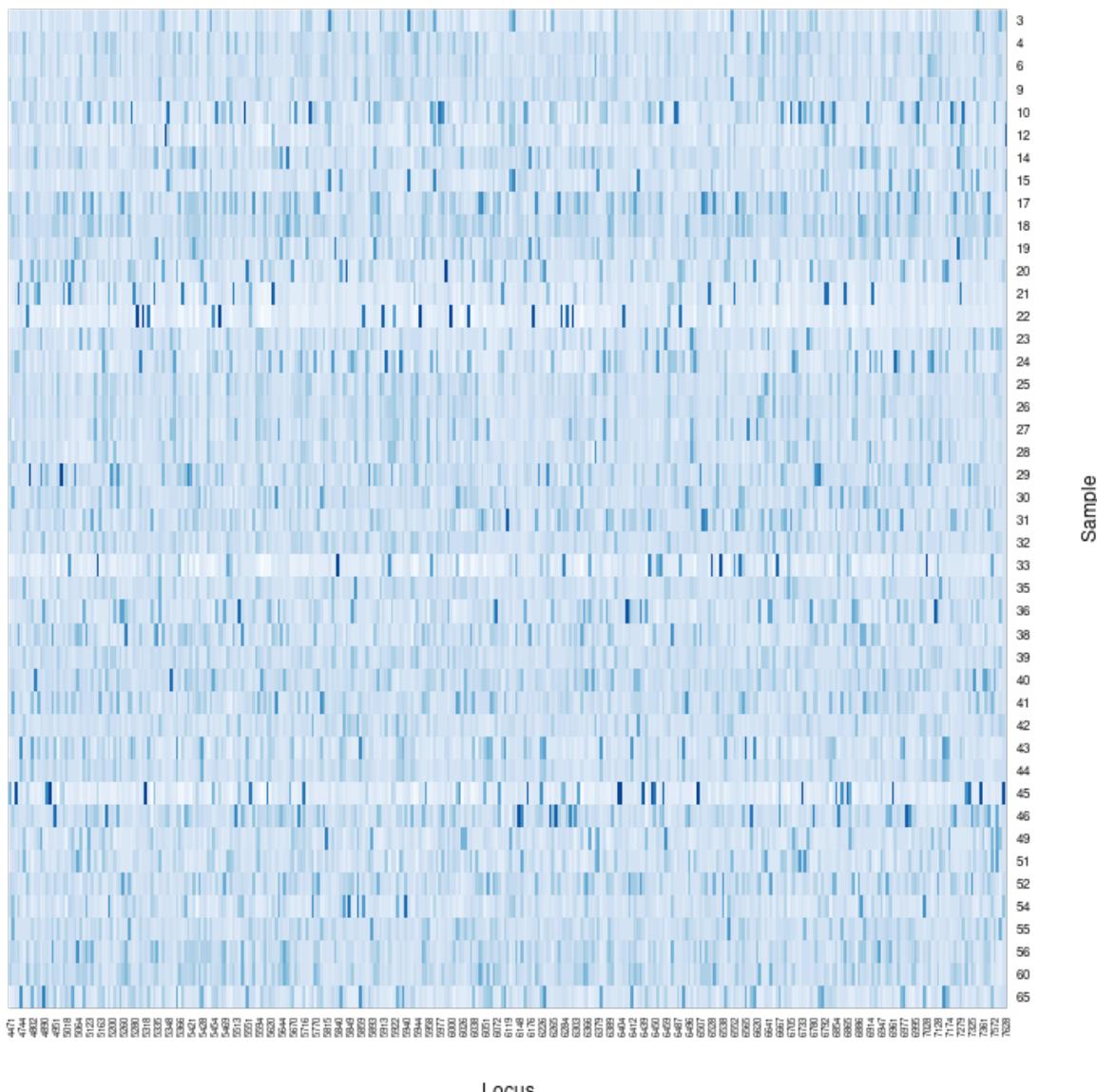
Order	Family	Taxon
		Saxifraga maderensis
		Saxifraga oppositifolia
		Saxifraga portosanctana
		Saxifraga x geum

x geum* \end{longtable}

Loci & Nucleotides Returned per Reference Sample



Percent matched reads



Comparision of Foraging Patterns from Three Sequence Alignment Algorithms



824 Appendix 13 - Models used for Species Distribution Model Ensembles

825 The two machine learning models utilize Ensemble learning.

826 **Ensemble learning** utilizes many sets of trees, each tree being composed of many binary decisions, to
827 create a single model. Each independent variable (- or *feature*) may become a node on the tree - i.e. a
828 location on the tree where a binary decision will move towards a predicted outcome. Each of the decision
829 tree models which ensemble learning utilizes is a weak model, each of which may suffer due to high variance
830 or bias, but which produce better outcomes than would be expected via chance. When ensembled these
831 models generate a strong model, a model which should have more appropriately balanced variance and bias
832 and predicts outcomes which are more strongly correlated with the expected values than the individual weak
833 models.

834 **Random Forest (RF)** the training data are continually bootstrap re-sampled, in combination with random
835 subsets of features, to create nodes which attempt to optimally predict a known outcome. A large number
836 of trees are then aggregated, via the most common predictions, to generate a final classification prediction
837 tree. Each individual prediction tree is generated independently of the others.

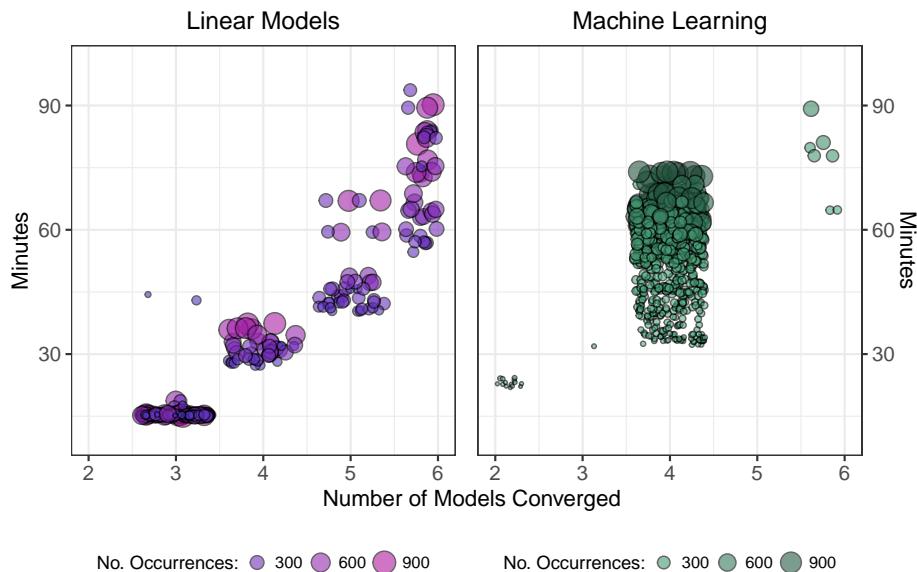
838 **Boosted Regression Tree (BRT)** (or Gradient Boosted tree) An initial tree is grown, and all other trees
839 are derived sequentially from it, as each new tree is grown the errors in responses from the last tree are
840 weighed more heavily so that the model focuses on selecting dependent variables which refine predictions.
841 All response data and predictor variables are kept available to all trees.

842 **Bias** predictions from an algorithm are systematically in error due to being prejudiced for or against certain
843 results, due to assumptions during learning.

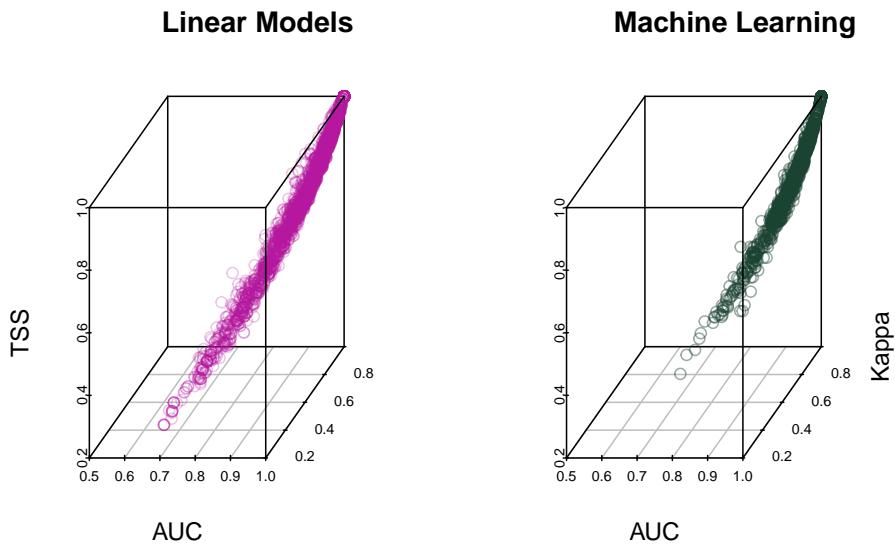
844 **Variance** errors in models due to an over-reliance and sensitivity of training to outliers in training data.

845 In general, Random Forest models have high bias and low variance, where boosted regressions trees have lower
846 bias and higher variance. Theoretically, the weaknesses and strengths of bootstrap aggregation (bagging) as
847 implemented by Random Forests are supplemented by the boosting.

Time Spent Fitting and Projecting Models onto Gridded Surfaces



Collectively it took 215 hours for all of the GLM and GAM to run, and for the converged models to be ensembled, and predicted onto a raster surface; it took 419 hours for the same process to be carried out for the Random Forest and Boosted Regression Tree models.



Results for each converged individual model which were then ensembled, using weights from the True Skill Statistic (TSS).

Table 1: Subset of Possible Combinations for re-classifying Sequences by Incorporating Ecological Factors

Spatial	Temporal	Congener	Confamilial	Congeners	Confamilials	Condition	Return	Rank
1	1	1	1	0	0	A.1	Input	Species
1	1	1	1	1	0	A.2	Input	Species
1	1	1	1	0	1	A.3	Input	Species
1	1	1	1	1	1	A.4	Input	Species
1	1	1	0	0	0	A.5	Input	Species
1	1	1	0	1	0	A.6	Input	Species
1	1	0	1	0	0	A.7	Input	Species
1	1	0	1	0	1	A.8	Input	Species
1	1	0	0	0	0	A.9	Input	Species
1	0	1	1	0	0	B.1	Congener	Species
1	0	1	1	0	1	B.2	Congener	Species
1	0	1	0	0	0	B.3	Congener	Species
1	0	1	1	1	0	C.1	Congener	Genus
1	0	1	1	1	1	C.2	Congener	Genus
1	0	1	0	1	0	C.3	Congener	Genus
1	0	0	1	0	0	D.1	Confamilial	Species
1	0	0	1	0	1	E.1	Confamilial	Family
1	0	0	0	0	0	F.1	Input	Species
0	0	1	1	0	0	G.1	Congener	Species
0	0	1	1	0	1	G.2	Congener	Species
0	0	1	0	0	0	G.3	Congener	Species
0	0	1	1	1	0	H.1	Congener	Genus
0	0	1	1	1	1	H.2	Congener	Genus
0	0	1	0	1	0	H.3	Congener	Genus
0	0	0	1	0	0	I.1	Confamilial	Species
0	0	0	1	0	1	J.1	Confamilial	Family

Note, for both ‘Congener’ and ‘Confamilial’ (*in the singular*) ‘1’ denotes that a species is present; in a sense the genus is monotypic in space and time. For both ‘Congeners’ and ‘Confamilials’ (*in the plural*), ‘1’ denotes that two or more species are present; ‘Confamilial’ again representing a monotypic entity in space and time.

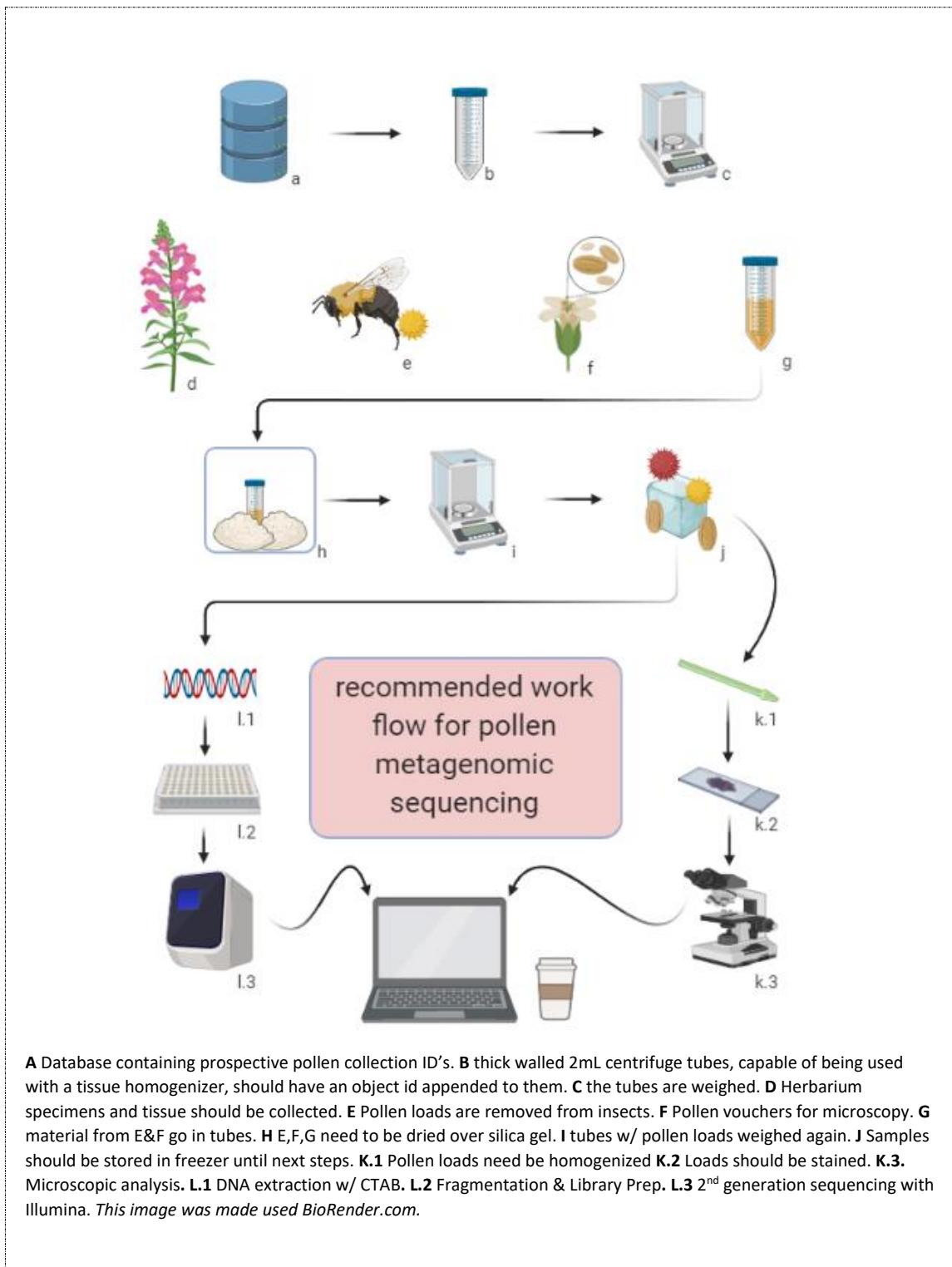
$$\begin{aligned} \text{Spatial} == 1 \& \text{ Temporal} == 1 \sim \mathbf{A} \\ \text{Spatial} == 1 \& \text{ Temporal} == 0 \& \text{ Congener} = 1 \sim \mathbf{B} \end{aligned}$$

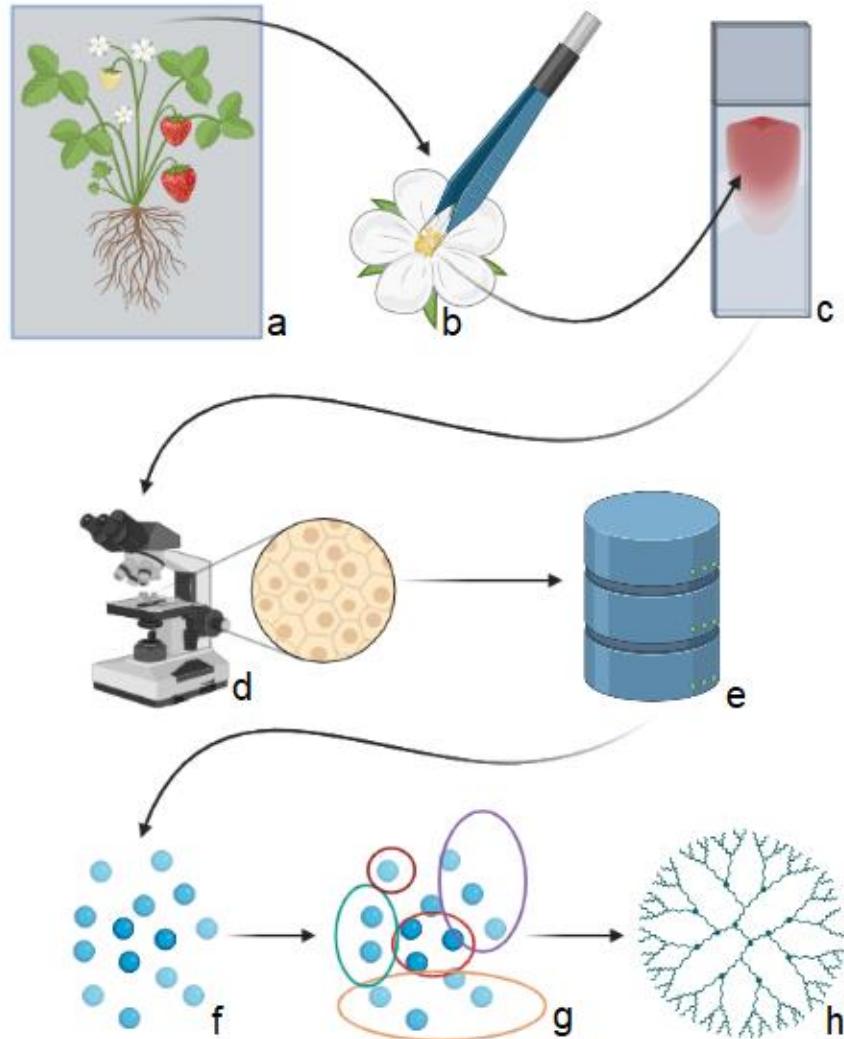
The temporal dimension is now buffered and a form of \mathbf{A} is employed
 $\text{Spatial} == 1 \& \text{ Temporal} +/- \text{ Buffer} == 1 \sim \mathbf{X}$

$$\begin{aligned} \text{Spatial} == 1 \& \text{ Temporal} == 0 \& \text{ Congeners} >= 2 \sim \mathbf{C} \\ \text{Spatial} == 1 \& \text{ Temporal} == 0 \& \text{ Congeners} == 0 \& \text{ Confamilial} == 1 \sim \mathbf{D} \\ \text{Spatial} == 1 \& \text{ Temporal} == 0 \& \text{ Congeners} == 0 \& \text{ Confamilial} >= 2 \sim \mathbf{E} \\ \text{Spatial} == 1 \& \text{ Temporal} == 0 \& \text{ Congener|s} == 0 \& \text{ Confamilial|s} == 0 \sim \mathbf{F} \end{aligned}$$

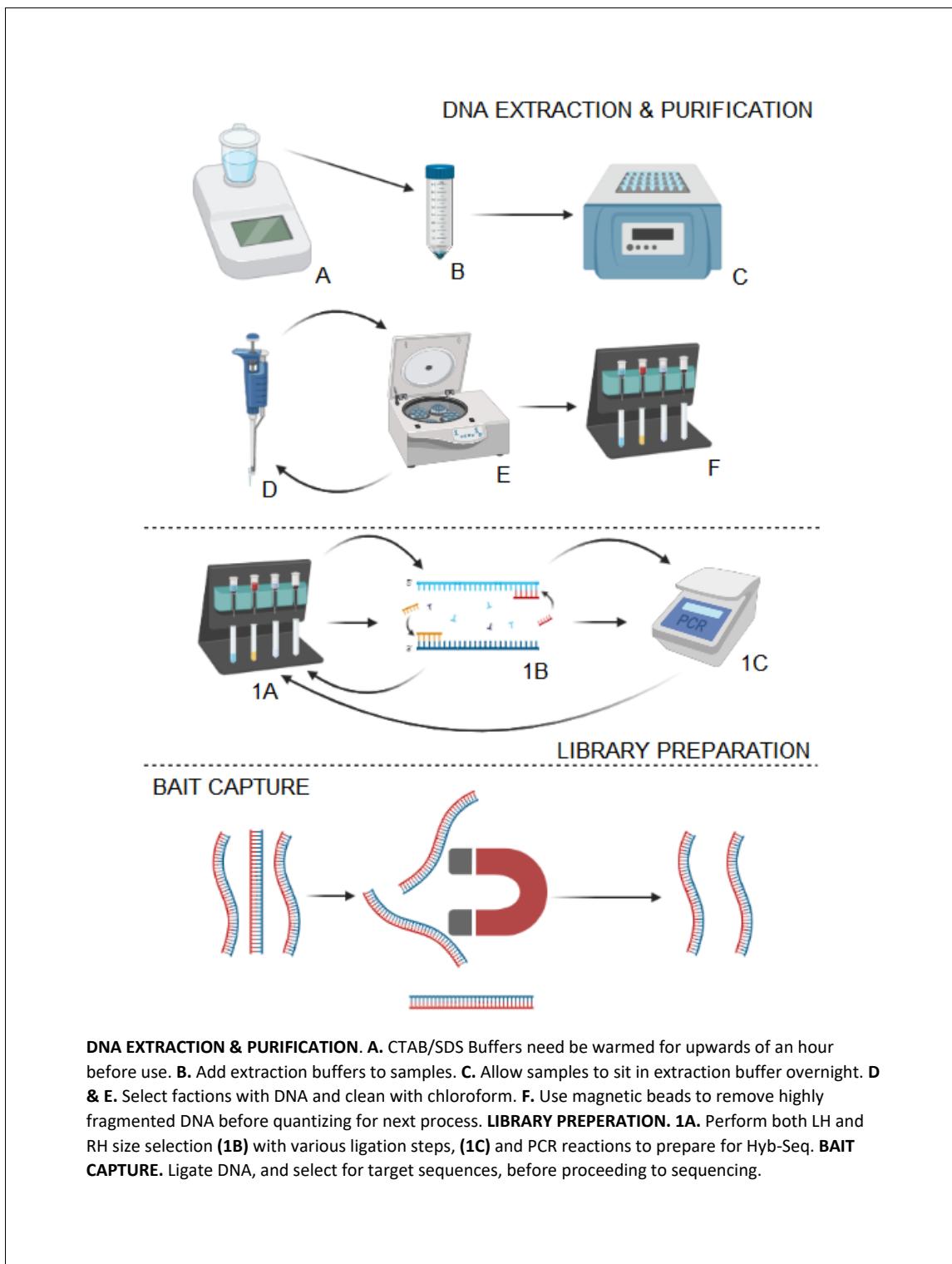
$$\begin{aligned} \text{Spatial} == 0 \& \text{ Temporal} == 0 \& \text{ Congener} == 1 \sim \mathbf{G} \\ \text{Spatial} == 0 \& \text{ Temporal} == 0 \& \text{ Congeners} == 1 \sim \mathbf{H} \\ \text{Spatial} == 0 \& \text{ Temporal} == 0 \& \text{ Confamilial} == 1 \sim \mathbf{I} \\ \text{Spatial} == 0 \& \text{ Temporal} == 0 \& \text{ Confamilials} == 1 \sim \mathbf{J} \end{aligned}$$

While the overall order matters, \mathbf{X} in particular may significantly alter conclusions.





A. An herbarium collection in flower from which pollen may be removed. **B.** The careful removal of pollen from dehiscent anthers using dissection tools. **C.** Place the pollen on a fuchsin jelly cube and melt it with stirring on a hotplate, for ca. 30 seconds. **D.** Microscopic imaging, of specimens and collection of character trait data. **E.** Immediate input and accessioning of files to a database. **F.** Ordination of traits into 2-dimensional space. **G.** Agglomerative clustering of data points into similar groups. **H.** Recovery of bifurcating decisions in development of clusters, or handwritten keys to visually diagnosable groups.



DNA EXTRACTION & PURIFICATION. A. CTAB/SDS Buffers need to be warmed for upwards of an hour before use. B. Add extraction buffers to samples. C. Allow samples to sit in extraction buffer overnight. D & E. Select fractions with DNA and clean with chloroform. F. Use magnetic beads to remove highly fragmented DNA before quantizing for next process. **LIBRARY PREPARATION.** 1A. Perform both LH and RH size selection (1B) with various ligation steps, (1C) and PCR reactions to prepare for Hyb-Seq. **BAIT CAPTURE.** Ligate DNA, and select for target sequences, before proceeding to sequencing.

CTAB-DNA POLLEN EXTRACTIONS

Adapted from Lalhmangiahi et. al & Guertler et al. by Benkendorf, Fant, & Noble.

SAMPLE PREPARATION AND GRINDING

- a1) Add 380 μ L extraction buffer (100 mM Tris-HCl, 50 mM EDTA, 50 mM NaCl, 10% SDS, pH 7.5). This solution will need to be warm enough for the SDS to be in solution, requires heat and stirring.
- a2) Vortex samples at speed > 2000, until pellet breaks apart, ca. 20-30 seconds.
- a3) Use the Pellet Pestle Motor (Kontes) for ca. 15 seconds to macerate samples.
- a4) Add 100 μ L extraction buffer to wash the tip of the pestle into the centrifuge tube, and burst bubbles.
- a5) Allow to sit at 35°C for 1 hour, use vortex occasionally if sedimentation of pollen occurs.

EXTRACTION AND ISOLATION OF DNA

- b1) Warm CTAB buffer to remove any precipitants if present.
- b2) Add 480 μ L 10% CTAB buffer.
- b3) Add 10 μ L RNase (10mg/mL); invert by hand, incubate for 40 minutes at 37°C, increase heat to 60°C wait 20 minutes before continuing to b4.
- b4) Add 15 μ L proteinase K (20mg/mL) & 12.5 μ L DTT (1 molar in H₂O); invert by hand, incubate for 1 hr. at 60°C.
- b5) Incubate overnight at 40°C (*note: this is a hard stopping point*)
- b6) Add 500 μ L of Phenol-chloroform-isoamyl alcohol vortex samples, centrifuge at 10,000 rpm (10 min.)
- b7) Transfer the uppermost aqueous layer to a new 2 mL centrifuge tube.

DNA PRECIPITATION

- c1) Add slightly chilled Isopropyl alcohol & Sodium Acetate 3mM 5:1, equivalent to ca. 2/3 of the removed layer. Store at -20°C, 1 hour to allow precipitation.
(Note: potential stopping point for a day or more, samples can stay at -20°C for days)
- c2) Centrifuge at 13,000 rpm for 10 minutes.
- c3) Pour supernatant into new 2mL centrifuge tube, add 400 μ L 70% EtOH. Store at -20°C for 20 minutes.
- c4) Spin at 13,000 rpm for 10 minutes, discard supernatant.
for both tubes the following steps apply
- c3) Add 400 μ L of 75% EtOH, invert tube x3, centrifuge at 13,000 rpm for 4 minutes; discard supernatant
- c4) Add 400 μ L of 95% EtOH, invert tube x3, centrifuge at 13,000 rpm for 4 minutes, discard supernatant
- c5) Dry tubes in vacuum centrifuge for 30 minutes on medium heat at 15 mmHG.

RESUSPENSION OF DNA

- d1) Add 40 μ L of dna free H₂O to sample.
- d2) place on heat block at 37°C until pellet resuspends with occasional use of vortexes.

NOTES: a 10% CTAB preparation will not readily stay in solution, maintain it on heatblock until you are ready to use it. After adding it to extraction tubes move them to heat block immediately (*i.e.* in batches of 5-10).

Solutions

Extraction buffer (100 mM Tris-HCl, 50 mM EDTA, 50 mM NaCl, 10% SDS - pH 7.5, ca. 32 mL H₂O)

For 100 samples (50 mL solution)

10 grams SDS (Sodium Dodecyl Sulfate , d = 1.01 g/cm3)

146.1 mg Sodium Chloride (NaCl, mw = 58.4 g/mol)

930.6 mg EDTA (EthyleneDiamineTetraacetic Acid Disodium Salt dihydrate, mw = 372.24 g/mol)

Add 20 mL deH₂O

5 mL Tris-HCl pH 8.0 (1 molar- kept in fridge)

Fill to 50 mL with deH₂O

Auto clave on 'Liquid' setting for 15 minutes.

Dissolution may require heat and stirring (3 & 4 settings respectively, ca. 15 min.)

10% CTAB solution (20 mM Tris-Cl pH 8.0, 1.4 M NaCl, 10 mM EDTA pH 7.5, 10% CTAB, 5% PVP, 40 mL DiH₂O)

For 100 samples (50 mL solution)

add ~30 mL deH₂O,

1 ml Tris-HCl pH 8.0 (1 molar- kept in fridge; 2-Amino-2-(hydroxymethyl)propane-1,3-diol)

4.08 g Sodium Chloride (NaCl, mw = 58.4 g/mol)

4 mL EDTA pH 7.5 (0.125 molar – kept in fridge; 2,2',2'',2'''-(Ethane-1,2-diyl)dinitrilo)tetraacetic acid)

5 g CTAB (hexadecyl(trimethyl)ammonium bromide, mw = 364.45, FYI this is 274 mM)

Auto clave on 'Liquid' setting for 15 minutes.

2.5 g PVP-40 (1-ethenylpyrrolidin-2-one) – add after autoclave

Fill to 50 mL with deH₂O

Dissolution of PVP will require 2-3 hrs, at 65°C with stirring. Before use allow one hour of stirring and heat to resuspend all salts in the solution.

Sodium acetate solution (3mM)

For 100 samples (10 mL solution)

20.4 mg Sodium Acetate trihydrate (mw = 136.08 g/M)

to 50 mL deH₂O

Auto clave on 'Liquid' setting for 15 minutes.

Phenol-chloroform Isoamyl alcohol (25:24:1) Saturated with 10 mM Tris pH 8.0, EDTA

For 100 samples (50 mL solution) (no need to make, is bought)

25 mL Phenol

24 mL Chloroform (Trichloromethane)

1 mL Isoamyl alcohol

Literature cited

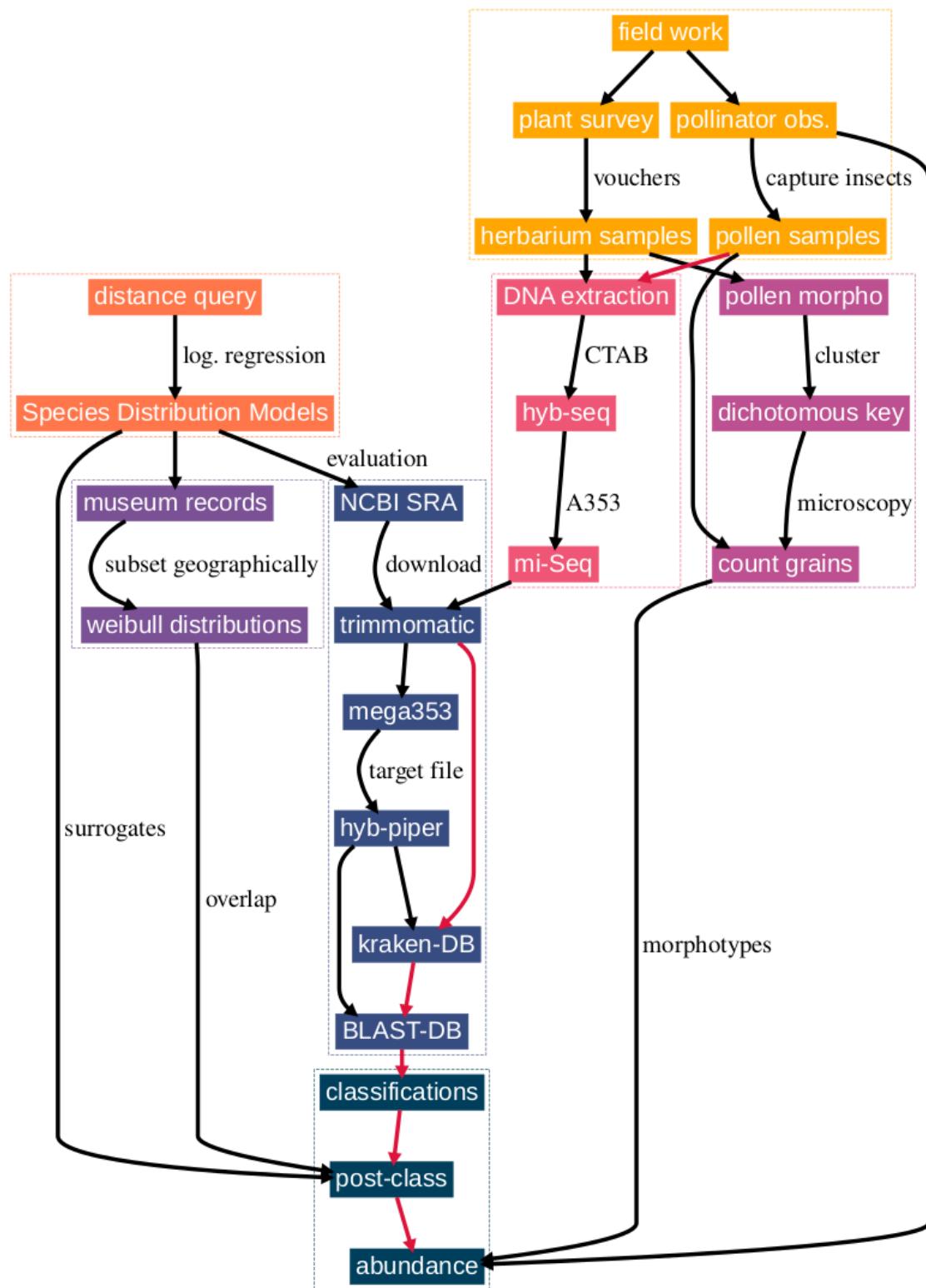
Lalhmangaibi, R., Ghatak, S., Laba, R., Gurusubramian, G., Jumar, N.S. *Protocol for Optimal Quality and Quantity Pollen DNA Isolation from Honey Samples*. 2014. Journal of Biomolecular Techniques 25:92-95

Guertler, P., Eicheldinger, A., Muschler, P., Goerlich, O., Bursch, U. *Automated DNA extraction from pollen in honey* 2014. Food Chemistry 149:302-306

Table 1: Queen Bee Pollen Loads examined

Subgenus	Species	Author	Tongue Length	Microscope Slides	Metagenome Samples
Subterraneobombus	<i>Vogt</i>	<i>B. appositus</i>	Cresson 1879	Long	11 NA
Pyrobonbus	<i>Dalla Torre</i>	<i>B. bifarius</i>	Cresson 1879	Short	11 NA
Thoracobombus	<i>Dalla Torre</i>	<i>B. californicus</i>	Smith 1854	Long	8 NA
Pyrobonbus	<i>Dalla Torre</i>	<i>B. flavifrons</i>	Cresson 1864	Medium	13 NA
Pyrobonbus	<i>Dalla Torre</i>	<i>B. mixtus</i>	Cresson 1879	Short	3 NA
Bonbius	<i>Robertson</i>	<i>B. nevadensis</i>	Cresson 1874	Long	5 NA
Cullumanobombus	<i>Vogt</i>	<i>B. rufocinctus</i>	Cresson 1864	Short	13 NA
Pyrobonbus	<i>Dalla Torre</i>	<i>B. sylvicola</i>	Kirby 1837	Short	1 NA

^a All subgenera follow the system of Williams et al. 2008, and placements were found from the NMH website.
^b Tongue Lengths collected from Pyke et al. 2012



862 **References**

- 863 Ackerfield, J. (2015). *Flora of colorado*. BRIT Press Fort Worth.
- 864 Agrawal, A.A., Ackerly, D.D., Adler, F., Arnold, A.E., Caceres, C., Doak, D.F., Post, E., Hudson, P.J.,
865 Maron, J., Mooney, K.A. & others. (2007). Filling key gaps in population and community ecology.
866 *Frontiers in Ecology and the Environment*, **5**, 145–152.
- 867 Alarcón, R. (2010). Congruence between visitation and pollen-transport networks in a california plant–
868 pollinator community. *Oikos*, **119**, 35–44. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0706.2009.17694.x>
- 869 Aldous, A.E. (1919). *Eradicating tall larkspur on cattle ranges in the national forest*. US Department of
870 Agriculture.
- 871 Allouche, O., Tsoar, A. & Kadmon, R. (2006). Assessing the accuracy of species distribution models:
872 Prevalence, kappa and the true skill statistic (TSS). *Journal of applied ecology*, **43**, 1223–1232.
- 873 Allred, K.W. & Ivey, R. (2012). Flora neomexicana III: An illustrated identification manual. *Lulu. com*.
- 874 Araujo, M.B. & New, M. (2007). Ensemble forecasting of species distributions. *Trends in ecology & evolution*,
875 **22**, 42–47.
- 876 Ashman, T.-L. & Arceo-Gómez, G. (2013). Toward a predictive understanding of the fitness costs of het-
877 erospecific pollen receipt and its importance in co-flowering communities. *American Journal of Botany*,
878 **100**, 1061–1070.
- 879 Augspurger, C.K. & Zaya, D.N. (2020). Concordance of long-term shifts with climate warming varies among
880 phenological events and herbaceous species. *Ecological Monographs*, **90**, e01421.
- 881 Baker, W.J., Bailey, P., Barber, V., Barker, A., Bellot, S., Bishop, D., Botigué, L.R., Brewer, G., Carruthers,
882 T., Clarkson, J.J., Cook, J., Cowan, R.S., Dodsworth, S., Epitawalage, N., Francoso, E., Gallego, B.,
883 Johnson, M.G., Kim, J.T., Leempoel, K., Maurin, O., McGinnies, C., Pokorny, L., Roy, S., Stone, M.,
884 Toledo, E., Wickett, N.J., Zuntini, A.R., Eiserhardt, W.L., Kersey, P.J., Leitch, I.J. & Forest, F. (2021a).
885 A Comprehensive Phylogenomic Platform for Exploring the Angiosperm Tree of Life. *Systematic Biology*,
886 **71**, 301–319. Retrieved from <https://doi.org/10.1093/sysbio/syab035>
- 887 Baker, W., Dodsworth, S., Forest, F., Graham, S., Johnson, M., McDonnell, A., Pokorny, L., Tate, J., Wicke,
888 S. & Wickett, N. (2021b). Exploring Angiosperms353: An open, community toolkit for collaborative
889 phylogenomic research on flowering plants. *American Journal of Botany*, **108**.
- 890 Barbet-Massin, M., Jiguet, F., Albert, C.H. & Wilfried. (2012). Selecting pseudo-absences for species
891 distribution models: How, where and how many? *Methods in ecology and evolution*, **3**, 327–338.
- 892 Barker, D.A. & Arceo-Gómez, G. (2021). Pollen transport networks reveal highly diverse and temporally
893

- stable plant–pollinator interactions in an Appalachian floral community. *AoB PLANTS*, **13**. Retrieved from <https://doi.org/10.1093/aobpla/plab062>
- Bascompte, J. (2009). Mutualistic networks. *Frontiers in Ecology and the Environment*, **7**, 429–436.
- Bascompte, J., Jordano, P. & Olesen, J.M. (2006). Asymmetric coevolutionary networks facilitate biodiversity maintenance. *Science*, **312**, 431–433.
- Basey, A.C., Fant, J.B. & Kramer, A.T. (2015). Producing native plant materials for restoration: 10 rules to collect and maintain genetic diversity. *Native Plants Journal*, **16**, 37–53.
- Beattie, A. (1971). A technique for the study of insect-borne pollen. *The Pan-Pacific Entomologist*, **47**, 82.
- Beck, J.B., Markley, M.L., Zielke, M.G., Thomas, J.R., Hale, H.J., Williams, L.D. & Johnson, M.G. (2021). Are palmer's elm-leaf goldenrod and the smooth elm-leaf goldenrod real? The Angiosperms353 kit provides within-species signal in solidago ulmifolia sl. *Systematic Botany*, **46**, 1107–1113.
- Belitz, M.W., Larsen, E.A., Ries, L. & Guralnick, R.P. (2020). The accuracy of phenology estimators for use with sparsely sampled presence-only observations. *Methods in Ecology and Evolution*, **11**, 1273–1285.
- Bell, K.L., Burgess, K.S., Botsch, J.C., Dobbs, E.K., Read, T.D. & Brosi, B.J. (2019). Quantitative and qualitative assessment of pollen DNA metabarcoding using constructed species mixtures. *Molecular Ecology*, **28**, 431–455.
- Bell, K.L., Fowler, J., Burgess, K.S., Dobbs, E.K., Gruenewald, D., Lawley, B., Morozumi, C. & Brosi, B.J. (2017a). Applying pollen DNA metabarcoding to the study of plant–pollinator interactions. *Applications in plant sciences*, **5**, 1600124.
- Bell, K.L., Fowler, J., Burgess, K.S., Dobbs, E.K., Gruenewald, D., Lawley, B., Morozumi, C. & Brosi, B.J. (2017b). Applying pollen DNA metabarcoding to the study of plant–pollinator interactions. *Applications in plant sciences*, **5**, 1600124.
- Bell, K.L., Petit III, R.A., Cutler, A., Dobbs, E.K., Macpherson, J.M., Read, T.D., Burgess, K.S. & Brosi, B.J. (2021). Comparing whole-genome shotgun sequencing and DNA metabarcoding approaches for species identification and quantification of pollen species mixtures. *Ecology and Evolution*, **11**, 16082–16098.
- Bell, K.L., Turo, K.J., Lowe, A., Nota, K., Keller, A., Encinas-Viso, F., Parducci, L., Richardson, R.T., Leggett, R.M., Brosi, B.J. & others. (2022). Plants, pollinators and their interactions under global ecological change: The role of pollen DNA metabarcoding. *Molecular ecology*.
- Belsky, A.J., Matzke, A. & Uselman, S. (1999). Survey of livestock influences on stream and riparian ecosystems in the western united states. *Journal of Soil and water Conservation*, **54**, 419–431.
- Bergman, P., Molau, U. & Holmgren, B. (1996). Micrometeorological impacts on insect activity and plant reproductive success in an alpine environment, swedish lapland. *Arctic and alpine research*, **28**, 196–202.

- 927 Betancourt, J.L., Schwartz, M.D., Breshears, D.D., Cayan, D.R., Dettinger, M.D., Inouye, D.W., Post, E.
928 & Reed, B.C. (2005). Implementing a US national phenology network.
- 929 Bingham, R.A. & Orthner, A.R. (1998). Efficient pollination of alpine plants. *Nature*, **391**, 238–239.
- 930 Bolger, A. & Giorgi, F. (2014). Trimmomatic: A flexible read trimming tool for illumina NGS data. *Bioin-*
931 *formatics*, **30**, 2114–2120.
- 932 Bontsutsnaja, A., Karise, R., Mand, M. & Smagghe, G. (2021). Bumble bee foraged pollen analyses in spring
933 time in southern estonia shows abundant food sources. *Insects*, **12**, 922.
- 934 Breinholt, J.W., Carey, S.B., Tiley, G.P., Davis, E.C., Endara, L., McDaniel, S.F., Neves, L.G., Sessa, E.B.,
935 Konrat, M. von, Chantanaorrapint, S. & others. (2021). A target enrichment probe set for resolving the
936 flagellate land plant tree of life. *Applications in Plant Sciences*, **9**, e11406.
- 937 Brewen, C.J., Berrill, J.-P., Ritchie, M.W., Boston, K., Dagley, C.M., Jones, B., Coppoletta, M. & Burnett,
938 C.L. (2021). 76-year decline and recovery of aspen mediated by contrasting fire regimes: Long-unburned,
939 infrequent and frequent mixed-severity wildfire. *Plos one*, **16**, e0232995.
- 940 Brito-Morales, I., Molinos, J.G., Schoeman, D.S., Burrows, M.T., Poloczanska, E.S., Brown, C.J., Ferrier,
941 S., Harwood, T.D., Klein, C.J., McDonald-Madden, E. & others. (2018). Climate velocity can inform
942 conservation in a warming world. *Trends in ecology & evolution*, **33**, 441–457.
- 943 Brosi, B.J. (2016). Pollinator specialization: From the individual to the community. *New Phytologist*, **210**,
944 1190–1194.
- 945 Brosi, B.J. & Briggs, H.M. (2013). Single pollinator species losses reduce floral fidelity and plant reproductive
946 function. *Proceedings of the National Academy of Sciences*, **110**, 13044–13048.
- 947 Calabrese, J.M., Certain, G., Kraan, C. & Dormann, C.F. (2014). Stacking species distribution models and
948 adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography*, **23**, 99–112.
- 949 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T.L. (2009).
950 BLAST+: Architecture and applications. *BMC bioinformatics*, **10**, 1–9.
- 951 Cameron, S.A. & Sadd, B.M. (2020). Global trends in bumble bee health. *Annual review of entomology*, **65**,
952 209–232.
- 953 CaraDonna, P.J., Iler, A.M. & Inouye, D.W. (2014). Shifts in flowering phenology reshape a subalpine plant
954 community. *Proceedings of the National Academy of Sciences*, **111**, 4916–4921.
- 955 CaraDonna, P.J., Petry, W.K., Brennan, R.M., Cunningham, J.L., Bronstein, J.L., Waser, N.M. & Sanders,
956 N.J. (2017). Interaction rewiring and the rapid turnover of plant–pollinator networks. *Ecology letters*,
957 **20**, 385–394.
- 958 CaraDonna, P.J. & Waser, N.M. (2020). Temporal flexibility in the structure of plant–pollinator interaction
959 networks. *Oikos*, **129**, 1369–1380.

- 960 Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M.,
961 Tilman, D., Wardle, D.A. & others. (2012). Biodiversity loss and its impact on humanity. *Nature*, **486**,
962 59–67.
- 963 Chao, A., Gotelli, N.J., Hsieh, T.C., Sande, E.L., Ma, K.H., Colwell, R.K. & Ellison, A.M. (2014). Rarefac-
964 tion and extrapolation with hill numbers: A framework for sampling and estimation in species diversity
965 studies. *Ecological Monographs*, **84**, 45–67.
- 966 Cheng, S., Melkonian, M., Smith, S.A., Brockington, S., Archibald, J.M., Delaux, P.-M., Li, F.-W., Melko-
967 nian, B., Mavrodiev, E.V., Sun, W., Fu, Y., Yang, H., Soltis, D.E., Graham, S.W., Soltis, P.S., Liu,
968 X., Xu, X. & Wong, G.K.-S. (2018). 10KP: A phylodiverse genome sequencing plan. *GigaScience*, **7**.
969 Retrieved from <https://doi.org/10.1093/gigascience/giy013>
- 970 Coissac, E., Hollingsworth, P.M., Lavergne, S. & Taberlet, P. (2016). From barcodes to genomes: Extending
971 the concept of DNA barcoding.
- 972 Coissac, E., Riaz, T. & Puillandre, N. (2012). Bioinformatic challenges for DNA metabarcoding of plants
973 and animals. *Molecular ecology*, **21**, 1834–1847.
- 974 Colla, S.R., Gadallah, F., Richardson, L., Wagner, D. & Gall, L. (2012). Assessing declines of north american
975 bumble bees (*bombus* spp.) Using museum specimens. *Biodiversity and Conservation*, **21**, 3585–3595.
- 976 Collen, B., Ram, M., Zamin, T. & McRae, L. (2008). The tropical biodiversity data gap: Addressing
977 disparity in global monitoring. *Tropical Conservation Science*, **1**, 75–88.
- 978 Cooke, R.U. & Reeves, R.W. (1976). *Arroyos and environmental change in the american south-west*. Claren-
979 don Press.
- 980 Crisci, J.V., Katinas, L., Apodaca, M.J. & Hoch, P.C. (2020). The end of botany. *Trends in Plant Science*,
981 **25**, 1173–1176.
- 982 Cronquist, A., Holmgren, A.H., Holmgren, N.H., Reveal, J.L., Holmgren, P.K., Barneby, R & others.
983 (1977+). *Intermountain flora. Vascular plants of the intermountain west, USA volume six. The mono-*
984 *cotyledons*. Columbia University.
- 985 Dahl, T.E. (1990). *Wetlands losses in the united states, 1780's to 1980's*. US Department of the Interior,
986 Fish; Wildlife Service.
- 987 Dobrowski, S.Z. & Parks, S.A. (2016). Climate change velocity underestimates climate change exposure in
988 mountainous regions. *Nature Communications*, **7**, 1–8.
- 989 Doyle, J.J. & Doyle, J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue.
990 *Phytochemical Bulletin*, **19**, 11–15.
- 991 Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.-P. & Guisan, A. (2011). Predicting spatial
992 patterns of plant species richness: A comparison of direct macroecological and species stacking modelling

- 993 approaches. *Diversity and Distributions*, **17**, 1122–1131.
- 994 Duffy, J.E., Cardinale, B.J., France, K.E., McIntyre, P.B., Thébaud, E. & Loreau, M. (2007). The functional
995 role of biodiversity in ecosystems: Incorporating trophic complexity. *Ecology letters*, **10**, 522–538.
- 996 Elith*, J., H. Graham*, C., P. Anderson, R., Dudik, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann,
997 F., R. Leathwick, J., Lehmann, A. & others. (2006). Novel methods improve prediction of species'
998 distributions from occurrence data. *Ecography*, **29**, 129–151.
- 999 Fazekas, A.J., Kesakurti, P.R., Burgess, K.S., Percy, D.M., Graham, S.W., Barrett, S.C., Newmaster,
1000 S.G., Hajibabaei, M. & Husband, B.C. (2009). Are plant species inherently harder to discriminate than
1001 animal species using DNA barcoding markers? *Molecular Ecology Resources*, **9**, 130–139.
- 1002 Fenn, M.E., Baron, J.S., Allen, E.B., Rueth, H.M., Nydick, K.R., Geiser, L., Bowman, W.D., Sickman, J.O.,
1003 Meixner, T., Johnson, D.W. & others. (2003). Ecological effects of nitrogen deposition in the western
1004 united states. *BioScience*, **53**, 404–420.
- 1005 Flora of North America Editorial Committee, eds. (1993+). *Flora of north america north of mexico [online]*.
1006 Oxford University Press on Demand.
- 1007 Frase, Barbara A. & Buck, P. (2007). Vascular Plants of the Gothic Area. Retrieved from https://www.digitalrmbi.org/wp-content/uploads/2016/05/vascularplantlist_20071.pdf
- 1008 Freitas, T.M., Montag, L.F., De Marco Jr, P. & Hortal, J. (2020). How reliable are species identifica-
1009 tions in biodiversity big data? Evaluating the records of a neotropical fish family in online repositories.
1010 *Systematics and biodiversity*, **18**, 181–191.
- 1011 Futuyma, D.J. & Agrawal, A.A. (2009). Macroevolution and the biological diversity of plants and herbivores.
1012 *Proceedings of the National Academy of Sciences*, **106**, 18054–18061.
- 1013 Galloni, M., Podda, L., Vivarelli, D., Quaranta, M. & Cristofolini, G. (2008). Visitor diversity and pollinator
1014 specialization in mediterranean legumes. *Flora-Morphology, Distribution, Functional Ecology of Plants*,
1015 **203**, 94–102.
- 1016 Genissel, A., Aupinel, P., Bressac, C., Tasei, J.-N. & Chevrier, C. (2002). Influence of pollen origin on
1017 performance of *bombus terrestris* micro-colonies. *Entomologia Experimentalis et Applicata*, **104**, 329–
1018 336.
- 1019 Goulson, D. (2010). *Bumblebees: Behaviour, ecology, and conservation*. Oxford University Press on Demand.
- 1020 Goulson, D., Hanley, M.E., Darvill, B., Ellis, J. & Knight, M.E. (2005). Causes of rarity in bumblebees.
1021 *Biological conservation*, **122**, 1–8.
- 1022 Goulson, D., Lye, G.C. & Darvill, B. (2008a). Diet breadth, coexistence and rarity in bumblebees. *Biodi-
1023 versity and Conservation*, **17**, 3269–3288.
- 1024 Goulson, D., Lye, G. & Darvill, B. (2008b). The decline and conservation of bumblebees. *Annual review of*

- 1026 *entomology*, **53**, 191–208.
- 1027 Grixti, J.C., Wong, L.T., Cameron, S.A. & Favret, C. (2009). Decline of bumble bees (*bombus*) in the north
1028 american midwest. *Biological conservation*, **142**, 75–84.
- 1029 Group, C.P.W., Hollingsworth, P.M., Forrest, L.L., Spouge, J.L., Hajibabaei, M., Ratnasingham, S., Bank,
1030 M. van der, Chase, M.W., Cowan, R.S., Erickson, D.L. & others. (2009). A DNA barcode for land
1031 plants. *Proceedings of the National Academy of Sciences*, **106**, 12794–12797.
- 1032 Group, C.P.B., Li, D.-Z., Gao, L.-M., Li, H.-T., Wang, H., Ge, X.-J., Liu, J.-Q., Chen, Z.-D., Zhou, S.-L.,
1033 Chen, S.-L. & others. (2011). Comparative analysis of a large dataset indicates that internal transcribed
1034 spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National
1035 Academy of Sciences*, **108**, 19641–19646.
- 1036 Guisan, A. & Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. *Ecological
1037 modelling*, **135**, 147–186.
- 1038 Hanley, M.E., Franco, M., Pichon, S., Darvill, B. & Goulson, D. (2008). Breeding system, pollinator choice
1039 and variation in pollen quality in british herbaceous plants. *Functional Ecology*, 592–598.
- 1040 Havens, K., Vitt, P., Schwarz, J., Orr, B. & Crimmins, T. (2007). Chicago botanic garden's conservation
1041 and outreach efforts on climate change. *BGjournal*, **4**, 13–16.
- 1042 Hebert, P.D., Cywinski, A., Ball, S.L. & DeWaard, J.R. (2003). Biological identifications through DNA
1043 barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**, 313–321.
- 1044 Hembry, D.H. & Weber, M.G. (2020). Ecological interactions and macroevolution: A new field with old
1045 roots. *Annual Review of Ecology, Evolution, and Systematics*, **51**, 215–243.
- 1046 Hengl, T., Mendes de Jesus, J., Heuvelink, G.B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A.,
1047 Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B. & others. (2017). SoilGrids250m:
1048 Global gridded soil information based on machine learning. *PLoS one*, **12**, e0169748.
- 1049 Hennig, C. (2020). *Fpc: Flexible procedures for clustering*. Retrieved from <https://CRAN.R-project.org/>
1050 package=fpc
- 1051 Hinchliff, C.E., Smith, S.A., Allman, J.F., Burleigh, J.G., Chaudhary, R., Coghill, L.M., Crandall, K.A.,
1052 Deng, J., Drew, B.T., Gazis, R. & others. (2015). Synthesis of phylogeny and taxonomy into a compre-
1053 hensive tree of life. *Proceedings of the National Academy of Sciences*, **112**, 12764–12769.
- 1054 Hitchcock, C.L. & Cronquist, A. (2018). *Flora of the pacific northwest: An illustrated manual*. University
1055 of Washington Press.
- 1056 Hollingsworth, P.M., Li, D.-Z., Bank, M. van der & Twyford, A.D. (2016). Telling plant species apart with
1057 DNA: From barcodes to genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*,
1058 **371**, 20150338.

- 1059 Hsieh, T.C., Ma, K.H. & Chao, A. (2020). *iNEXT: Interpolation and extrapolation for species diversity*.
1060 Retrieved from http://chao.stat.nthu.edu.tw/wordpress/software_download/
- 1061 Iler, A.M., Humphrey, P.T., Ogilvie, J.E. & CaraDonna, P.J. (2021). Conceptual and practical issues limit
1062 the utility of statistical estimators of phenological events. *Ecosphere*, **12**, e03828.
- 1063 Janzen, D.H. (1967). Synchronization of sexual reproduction of trees within the dry season in central america.
1064 *Evolution*, **21**, 620–637.
- 1065 *Jepson flora project*. (2020).
- 1066 Johnson, M.D., Fokar, M., Cox, R.D. & Barnes, M.A. (2021). Airborne environmental DNA metabarcoding
1067 detects more diversity, with less sampling effort, than a traditional plant community survey. *BMC Ecology*
1068 and *Evolution*, **21**, 1–15.
- 1069 Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J. & Wickett,
1070 N.J. (2016). HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput
1071 sequencing reads using target enrichment. *Applications in plant sciences*, **4**, 1600016.
- 1072 Johnson, M.G., Pokorny, L., Dodsworth, S., Botigue, L.R., Cowan, R.S., Devault, A., Eiserhardt, W.L.,
1073 Epitawalage, N., Forest, F., Kim, J.T. & others. (2019). A universal probe set for targeted sequencing
1074 of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic biology*,
1075 **68**, 594–606.
- 1076 Jones, M.R. & Good, J.M. (2016). Targeted capture in evolutionary and ecological genomics. *Molecular
1077 ecology*, **25**, 185–202.
- 1078 Joppa, L.N., Roberts, D.L., Myers, N. & Pimm, S.L. (2011). Biodiversity hotspots house most undiscovered
1079 plant species. *Proceedings of the National Academy of Sciences*, **108**, 13171–13176.
- 1080 Joshi, J., Schmid, B., Caldeira, M., Dimitrakopoulos, P., Good, J., Harris, R., Hector, A., Huss-Danell, K.,
1081 Jumpponen, A., Minns, A. & others. (2001). Local adaptation enhances performance of common plant
1082 species. *Ecology Letters*, **4**, 536–544.
- 1083 Katoh, K. & Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: Improvements
1084 in performance and usability. *Molecular biology and evolution*, **30**, 772–780.
- 1085 Keane, R.E. (2002). Cascading effects of fire exclusion in rocky mountain ecosystems: A literature review.
- 1086 Kearns, C.A., Thomson, J.D. & others. (2001). *Natural history of bumblebees*. University Press of Colorado.
- 1087 Kramer, A.T. & Havens, K. (2015). Report in brief: Assessing botanical capacity to address grand challenges
1088 in the united states. *Natural Areas Journal*, **35**, 83–89.
- 1089 Kress, W.J. (2017). Plant DNA barcodes: Applications today and in the future. *Journal of systematics and
1090 evolution*, **55**, 291–307.
- 1091 Kress, W.J. & Erickson, D.L. (2007). A two-locus global DNA barcode for land plants: The coding rbcL

- 1092 gene complements the non-coding trnH-psbA spacer region. *PLoS one*, **2**, e508.
- 1093 Kuhn, M. (2022). *Caret: Classification and regression training*. Retrieved from <https://CRAN.R-project.org/package=caret>
- 1094
- 1095 Lamb, P.D., Hunter, E., Pinnegar, J.K., Creer, S., Davies, R.G. & Taylor, M.I. (2019). How quantitative is
1096 metabarcoding: A meta-analytical approach. *Molecular ecology*, **28**, 420–430.
- 1097 Land Management, B. of. (2019). U.s. Department of interior bureau of land management, BLM - assess-
1098 ment, inventory, and monitoring (AIM) terrestrial indicators raw dataset. Retrieved from <https://gbp->
1099 blm-egis.hub.arcgis.com/pages/aim
- 1100 Lang, D., Tang, M., Hu, J. & Zhou, X. (2019). Genome-skimming provides accurate quantification for pollen
1101 mixtures. *Molecular Ecology Resources*, **19**, 1433–1446.
- 1102 Lesica, P., Lavin, M. & Stickney, P.F. (2012). *Manual of montana vascular plants*. Brit Press.
- 1103 Lewin, H.A., Richards, S., Aiden, E.L., Allende, M.L., Archibald, J.M., Bálint, M., Barker, K.B., Baumgart-
1104 ner, B., Belov, K., Bertorelle, G., Blaxter, M.L., Cai, J., Caperello, N.D., Carlson, K., Castilla-Rubio,
1105 J.C., Chaw, S.-M., Chen, L., Childers, A.K., Coddington, J.A., Conde, D.A., Corominas, M., Crandall,
1106 K.A., Crawford, A.J., DiPalma, F., Durbin, R., Ebenezer, T.E., Edwards, S.V., Fedrigo, O., Flicek, P.,
1107 Formenti, G., Gibbs, R.A., Gilbert, M.T.P., Goldstein, M.M., Graves, J.M., Greely, H.T., Grigoriev,
1108 I.V., Hackett, K.J., Hall, N., Haussler, D., Helgen, K.M., Hogg, C.J., Isobe, S., Jakobsen, K.S., Janke,
1109 A., Jarvis, E.D., Johnson, W.E., Jones, S.J.M., Karlsson, E.K., Kersey, P.J., Kim, J.-H., Kress, W.J.,
1110 Kuraku, S., Lawniczak, M.K.N., Leebens-Mack, J.H., Li, X., Lindblad-Toh, K., Liu, X., Lopez, J.V.,
1111 Marques-Bonet, T., Mazard, S., Mazet, J.A.K., Mazzoni, C.J., Myers, E.W., O'Neill, R.J., Paez, S.,
1112 Park, H., Robinson, G.E., Roquet, C., Ryder, O.A., Sabir, J.S.M., Shaffer, H.B., Shank, T.M., Sherkow,
1113 J.S., Soltis, P.S., Tang, B., Tedersoo, L., Uliano-Silva, M., Wang, K., Wei, X., Wetzer, R., Wilson,
1114 J.L., Xu, X., Yang, H., Yoder, A.D. & Zhang, G. (2022). The earth BioGenome project 2020: Start-
1115 ing the clock. *Proceedings of the National Academy of Sciences*, **119**, e2115635118. Retrieved from
1116 <https://www.pnas.org/doi/abs/10.1073/pnas.2115635118>
- 1117 Liang, H., Zhao, Y.-H., Rafferty, N.E., Ren, Z.-X., Zhong, L., Li, H.-D., Li, D.-Z. & Wang, H. (2021).
1118 Evolutionary and ecological factors structure a plant–bumblebee network in a biodiversity hotspot, the
1119 himalaya–hengduan mountains. *Functional Ecology*, **35**, 2523–2535.
- 1120 Life Project Consortium, D.T. of, Blaxter, M., Mieszkowska, N., Palma, F.D., Holland, P., Durbin, R.,
1121 Richards, T., Berriman, M., Kersey, P., Hollingsworth, P., Wilson, W., Twyford, A., Gaya, E., Lawniczak,
1122 M., Lewis, O., Broad, G., Howe, K., Hart, M., Flicek, P. & Barnes, I. (2022). Sequence locally, think glob-
1123 ally: The darwin tree of life project. *Proceedings of the National Academy of Sciences*, **119**, e2115642118.
1124 Retrieved from <https://www.pnas.org/doi/abs/10.1073/pnas.2115642118>

- 1125 Li, X., Jiang, L., Meng, F., Wang, S., Niu, H., Iler, A.M., Duan, J., Zhang, Z., Luo, C., Cui, S. & others.
1126 (2016). Responses of sequential and hierarchical phenological events to warming and cooling in alpine
1127 meadows. *Nature Communications*, **7**, 1–8.
- 1128 Liu, O.R. & Gaines, S.D. (2022). Environmental context dependency in species interactions. *Proceedings of*
1129 *the National Academy of Sciences*, **119**, e2118539119.
- 1130 Liu, J., Shi, L., Han, J., Li, G., Lu, H., Hou, J., Zhou, X., Meng, F. & Downie, S.R. (2014). Identification
1131 of species in the angiosperm family apiaceae using DNA barcodes. *Molecular ecology resources*, **14**,
1132 1231–1238.
- 1133 Li, X., Yang, Y., Henry, R.J., Rossetto, M., Wang, Y. & Chen, S. (2015). Plant DNA barcoding: From gene
1134 to genome. *Biological Reviews*, **90**, 157–166.
- 1135 Loarie, S.R., Duffy, P.B., Hamilton, H., Asner, G.P., Field, C.B. & Ackerly, D.D. (2009). The velocity of
1136 climate change. *Nature*, **462**, 1052–1055.
- 1137 Lu, J., Breitwieser, F.P., Thielen, P. & Salzberg, S.L. (2017). Bracken: Estimating species abundance in
1138 metagenomics data. *PeerJ Computer Science*, **3**, e104.
- 1139 Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. (2022). *Cluster: Cluster analysis basics*
1140 *and extensions*. Retrieved from <https://CRAN.R-project.org/package=cluster>
- 1141 Mainali, K. & Slud, E. (2022). *CooccurrenceAffinity: Affinity in cooccurrence data*.
- 1142 Mainali, K.P., Slud, E., Singer, M.C. & Fagan, W.F. (2022). A better index for analysis of co-occurrence
1143 and similarity. *Science Advances*, **8**, eabj9204.
- 1144 Maitner, B. (2022). *BIEN: Tools for accessing the botanical information and ecology network database*.
1145 Retrieved from <https://CRAN.R-project.org/package=BIEN>
- 1146 Manzano, S. (2021). Flippant attitudes towards plant identification jeopardize early career botanists. *Trends*
1147 *in Plant Science*, **26**, 987–988.
- 1148 McLay, T.G., Birch, J.L., Gunn, B.F., Ning, W., Tate, J.A., Nauheimer, L., Joyce, E.M., Simpson, L.,
1149 Schmidt-Lebuhn, A.N., William J & others. (2021). New targets acquired: Improving locus recovery
1150 from the Angiosperms353 probe set. *Applications in plant sciences*, **9**.
- 1151 Meyer, C., Kreft, H., Guralnick, R. & Jetz, W. (2015). Global priorities for an effective information basis of
1152 biodiversity distributions. *Nature communications*, **6**, 1–8.
- 1153 Mohlenbrock, R.H. (2002). *Vascular flora of illinois*. SIU Press.
- 1154 Moore, A.J. & Bohs, L. (2003). An ITS phylogeny of balsamorhiza and wyethia (asteraceae: heliantheae).
1155 *American Journal of Botany*, **90**, 1653–1660.
- 1156 Naiman, R.J., Johnston, C.A. & Kelley, J.C. (1988). Alteration of north american streams by beaver.
1157 *BioScience*, **38**, 753–762.

- 1158 Naimi, B. & Araujo, M.B. (2016). Sdm: A reproducible and extensible r platform for species distribution
1159 modelling. *Ecography*, **39**, 368–375.
- 1160 Naimi, B., Hamm, N. a.s., Groen, T.A., Skidmore, A.K. & Toxopeus, A.G. (2014). Where is positional
1161 uncertainty a problem for species distribution modelling. *Ecography*, **37**, 191–203.
- 1162 Nazaire, M. & Hufford, L. (2014). Phylogenetic systematics of the genus mertensia (boraginaceae). *Systematic
1163 Botany*, **39**, 268–303.
- 1164 Nevado, B., Atchison, G.W., Hughes, C.E. & Filatov, D.A. (2016). Widespread adaptive evolution during
1165 repeated evolutionary radiations in new world lupins. *Nature communications*, **7**, 1–9.
- 1166 Newstrom, L.E., Frankie, G.W. & Baker, H.G. (1994). A new classification for plant phenology based on
1167 flowering patterns in lowland tropical rain forest trees at la selva, costa rica. *Biotropica*, **26**, 141–159.
- 1168 Oberhauser, K.S., Nail, K.R. & Altizer, S. (2015). *Monarchs in a changing world: Biology and conservation
1169 of an iconic butterfly*. Cornell University Press.
- 1170 Occownload Gbif.Org. (2021). Occurrence download. Retrieved from <https://www.gbif.org/occurrence/>
1171 download/0206948-200613084148143
- 1172 Ogilvie, J.E. & CaraDonna, P.J. (2022). The shifting importance of abiotic and biotic factors across the life
1173 cycles of wild pollinators. *Journal of Animal Ecology*.
- 1174 Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Solymos,
1175 P., Stevens, M.H.H., Szoeecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D.,
1176 Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H.B.A., FitzJohn, R., Friendly,
1177 M., Furneaux, B., Hannigan, G., Hill, M.O., Lahti, L., McGlinn, D., Ouellette, M.-H., Ribeiro Cunha,
1178 E., Smith, T., Stier, A., Ter Braak, C.J.F. & Weedon, J. (2022). *Vegan: Community ecology package*.
1179 Retrieved from <https://CRAN.R-project.org/package=vegan>
- 1180 Olden, J.D., Poff, N.L., Douglas, M.R., Douglas, M.E. & Fausch, K.D. (2004). Ecological and evolutionary
1181 consequences of biotic homogenization. *Trends in ecology & evolution*, **19**, 18–24.
- 1182 Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H., Rasmussen, C. & Jordano, P. (2011). Missing and
1183 forbidden links in mutualistic networks. *Proceedings of the Royal Society B: Biological Sciences*, **278**,
1184 725–732.
- 1185 Omernik, J.M. (1987). Ecoregions of the conterminous united states. *Annals of the Association of American
1186 geographers*, **77**, 118–125.
- 1187 Ottenlips, M.V., Mansfield, D.H., Buerki, S., Feist, M.A.E., Downie, S.R., Dodsworth, S., Forest, F., Plun-
1188 kett, G.M. & Smith, J.F. (2021). Resolving species boundaries in a recent radiation with the An-
1189 giosperms353 probe set: The lomatium packardiae/l. anomalam clade of the l. triternatum (apiaceae)
1190 complex. *American journal of botany*, **108**, 1217–1233.

- 1191 Park, D.S., Lyra, G.M., Ellison, A.M., Maruyama, R.K.B., Torquato, D. dos R., Asprino, R.C., Cook, B.I. &
1192 Davis, C.C. (2022). Herbarium records provide reliable phenology estimates in the understudied tropics.
1193 *Journal of Ecology*.
- 1194 Pearse, W.D., Davis, C.C., Inouye, D.W., Primack, R.B. & Davies, T.J. (2017). A statistical estimator
1195 for determining the limits of contemporary and historic phenology. *Nature Ecology & Evolution*, **1**,
1196 1876–1882.
- 1197 Pecl, G.T., Araújo, M.B., Bell, J.D., Blanchard, J., Bonebrake, T.C., Chen, I.-C., Clark, T.D., Colwell,
1198 R.K., Danielsen, F., Evengård, B. & others. (2017). Biodiversity redistribution under climate change:
1199 Impacts on ecosystems and human well-being. *Science*, **355**, eaai9214.
- 1200 Peel, N., Dicks, L.V., Clark, M.D., Heavens, D., Percival-Alwyn, L., Cooper, C., Davies, R.G., Leggett,
1201 R.M. & Yu, D.W. (2019). Semi-quantitative characterisation of mixed pollen samples using MinION
1202 sequencing and reverse metagenomics (RevMet). *Methods in Ecology and Evolution*, **10**, 1690–1701.
- 1203 Pepin, N., Arnone, E., Gobiet, A., Haslinger, K., Kotlarski, S., Notarnicola, C., Palazzi, E., Seibert, P.,
1204 Serafin, S., Schöner, W. & others. (2022). Climate changes and their elevational patterns in the mountains
1205 of the world. *Reviews of geophysics*, **60**, e2020RG000730.
- 1206 Pereira, H.M. & Cooper, H.D. (2006). Towards the global monitoring of biodiversity change. *Trends in
1207 Ecology & Evolution*, **21**, 123–129.
- 1208 Pereira, H.M., Leadley, P.W., Proença, V., Alkemade, R., Scharlemann, J.P., Fernandez-Manjarrés, J.F.,
1209 Araújo, M.B., Balvanera, P., Biggs, R., Cheung, W.W. & others. (2010). Scenarios for global biodiversity
1210 in the 21st century. *Science*, **330**, 1496–1501.
- 1211 Pereira, H.M., Navarro, L.M. & Martins, I.S. (2012). Global biodiversity change: The bad, the good, and
1212 the unknown. *Annual Review of Environment and Resources*, **37**, 25–50.
- 1213 Pineda, E. & Lobo, J.M. (2009). Assessing the accuracy of species distribution models to predict amphibian
1214 species richness patterns. *Journal of Animal Ecology*, **78**, 182–190.
- 1215 Pinto-Ledezma, J.N. & Cavender-Bares, J. (2021). Predicting species distributions and community compo-
1216 sition using satellite remote sensing predictors. *Scientific Reports*, **11**, 1–12.
- 1217 Porron, A., Andalo, C., Burrus, M. & Escaravage, N. (2017). DNA metabarcoding data unveils invisible
1218 pollination networks. *Scientific Reports*, **7**, 1–11.
- 1219 Prather, L.A., Alvarez-Fuentes, O., Mayfield, M.H. & Ferguson, C.J. (2004a). Implications of the decline in
1220 plant collecting for systematic and floristic research. *Systematic Botany*, **29**, 216–220.
- 1221 Prather, L.A., Alvarez-Fuentes, O., Mayfield, M.H. & Ferguson, C.J. (2004b). The decline of plant collecting
1222 in the united states: A threat to the infrastructure of biodiversity studies. *Systematic Botany*, **29**, 15–28.
- 1223 Prim, R.C. (1957). Shortest connection networks and some generalisations. *Bell System Technical Journal*,

- 1224 **36**, 1389–1401.
- 1225 Pusalkar, P.K. & Singh, D.K. (2015). Taxonomic rearrangement of arenaria (caryophyllaceae) in indian
1226 western himalaya. *Journal of Japanese Botany*, **90**, 77–91.
- 1227 Qiao, H., Soberon, J. & Peterson, A.T. (2015). No silver bullets in correlative ecological niche modelling:
1228 Insights from testing among many potential algorithms for niche estimation. *Methods in Ecology and*
1229 *Evolution*, **6**, 1126–1136.
- 1230 Rabeler, R.K. & Wagner, W.L. (2016). New combinations in odontostemma (caryophyllaceae). *PhytoKeys*,
1231 77.
- 1232 Ralphs, M.H. & Ueckert, D.N. (1988). Herbicide control of locoweeds: A review. *Weed Technology*, **2**,
1233 460–465.
- 1234 Ralphs, M., Woolsey, L. & Bowns, J. (2003). Mechanism by which ammonium fertilizers kill tall larkspur.
1235 *Rangeland Ecology & Management/Journal of Range Management Archives*, **56**, 524–528.
- 1236 Robinson, N., Regetz, J. & Guralnick, R.P. (2014). EarthEnv-DEM90: A nearly-global, void-free, multi-
1237 scale smoothed, 90m digital elevation model from fused ASTER and SRTM data. *ISPRS Journal of*
1238 *Photogrammetry and Remote Sensing*, **87**, 57–67.
- 1239 Ruete, A. (2015). Displaying bias in sampling effort of data accessed from biodiversity databases using
1240 ignorance maps. *Biodiversity Data Journal*.
- 1241 Ruppert, K.M., Kline, R.J. & Rahman, M.S. (2019). Past, present, and future perspectives of environmental
1242 DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global
1243 eDNA. *Global Ecology and Conservation*, **17**, e00547.
- 1244 Sadeghian, S., Zarre, S., Rabeler, R.K. & Heubl, G. (2015). Molecular phylogenetic analysis of arenaria
1245 (caryophyllaceae: Tribe arenarieae) and its allies inferred from nuclear DNA internal transcribed spacer
1246 and plastid DNA rps16 sequences. *Botanical Journal of the Linnean Society*, **178**, 648–669.
- 1247 Sarro, E., Tripodi, A. & Woodard, S.H. (2022). Bumble bee (*bombus vosnesenskii*) queen nest searching
1248 occurs independent of ovary developmental status. *Integrative Organismal Biology*, **4**, obac007.
- 1249 Schmitt, S., Pouteau, R., Justeau, D., De Boissieu, F. & Birnbaum, P. (2017). Ssdm: An r package to
1250 predict distribution of species richness and composition based on stacked species distribution models.
1251 *Methods in Ecology and Evolution*, **8**, 1795–1803.
- 1252 Sennikov, A.N. & Kurtto, A. (2017). A phylogenetic checklist of sorbus sl (rosaceae) in europe.
- 1253 Sickel, W., Ankenbrand, M.J., Grimmer, G., Holzschuh, A., Hartel, S., Lanzen, J., Steffan-Dewenter, I. &
1254 Keller, A. (2015). Increased efficiency in identifying mixed pollen samples by meta-barcoding with a
1255 dual-indexing approach. *BMC ecology*, **15**, 1–9.
- 1256 Slimp, M., Williams, L.D., Hale, H. & Johnson, M.G. (2021). On the potential of Angiosperms353 for

- 1257 population genomic studies. *Applications in Plant Sciences*, **9**.
- 1258 Smith, S.A. & Brown, J.W. (2018). Constructing a broadly inclusive seed plant phylogeny. *American journal*
1259 *of botany*, **105**, 302–314.
- 1260 Sork, V.L. (2018). Genomic studies of local adaptation in natural plant populations. *Journal of Heredity*,
1261 **109**, 3–15.
- 1262 Stevens, C.J., David, T.I. & Storkey, J. (2018). Atmospheric nitrogen deposition in terrestrial ecosystems: Its
1263 impact on plant communities and consequences across trophic levels. *Functional Ecology*, **32**, 1757–1769.
- 1264 Stroud, S., Fennell, M., Mitchley, J., Lydon, S., Peacock, J. & Bacon, K.L. (2022). The botanical education
1265 extinction and the fall of plant awareness. *Ecology and Evolution*, **12**, e9019.
- 1266 Suchan, T., Talavera, G., Saez, L., Ronikier, M. & Vila, R. (2019). Pollen metabarcoding as a tool for
1267 tracking long-distance insect migrations. *Molecular Ecology Resources*, **19**, 149–162.
- 1268 Tange, O. (2021). GNU parallel 20220322 (savannah). Retrieved from <https://doi.org/10.5281/zenodo.6377950>
- 1269
- 1270 Tasei, J.-N. & Aupinel, P. (2008). Nutritive value of 15 single pollens and pollen mixes tested on larvae
1271 produced by bumblebee workers (*bombus terrestris*, hymenoptera: apidae). *Apidologie*, **39**, 397–409.
- 1272 The biodiversity of species and their rates of extinction, distribution, and protection. (2014). *science*, **344**,
1273 1246752.
- 1274 Thompson, J.N. (1994). *The coevolutionary process*. University of Chicago press.
- 1275 Thuiller, W., Lavorel, S., Sykes, M.T. & Araújo, M.B. (2006). Using niche-based modelling to assess the
1276 impact of climate change on tree functional diversity in europe. *Diversity and Distributions*, **12**, 49–60.
- 1277 Tran, H., Nguyen, P., Ombadi, M., Hsu, K., Sorooshian, S. & Qing, X. (2019). A cloud-free MODIS snow
1278 cover dataset for the contiguous united states from 2000 to 2017. *Scientific data*, **6**, 1–13.
- 1279 Tylianakis, J.M., Laliberté, E., Nielsen, A. & Bascompte, J. (2010). Conservation of species interaction
1280 networks. *Biological conservation*, **143**, 2270–2279.
- 1281 Valiente-Banuet, A., Aizen, M.A., Alcantara, J.M., Arroyo, J., Cocucci, A., Galetti, M., Garcia, M.B.,
1282 Garcia, D., Gomez, J.M., Jordano, P. & others. (2015). Beyond species loss: The extinction of ecological
1283 interactions in a changing world. *Functional Ecology*, **29**, 299–307.
- 1284 Wang, T., Hamann, A., Spittlehouse, D. & Carroll, C. (2016). Locally downscaled and spatially customizable
1285 climate data for historical and future periods for north america. *PLoS one*, **11**, e0156720.
- 1286 Wang, S., Loreau, M., De Mazancourt, C., Isbell, F., Beierkuhnlein, C., Connolly, J., Deutschman, D.H.,
1287 Doležal, J., Eisenhauer, N., Hector, A. & others. (2021). Biotic homogenization destabilizes ecosystem
1288 functioning by decreasing spatial asynchrony. *Ecology*, **102**, e03332.
- 1289 Weber, W. (1998). New names and combinations in asteraceae: Heliantheae-ecliptinae. *Phytologia*, **85**,

- 1290 19–21.
- 1291 Wenzell, K.E., McDonnell, A.J., Wickett, N.J., Fant, J.B. & Skogen, K.A. (2021). Incomplete reproductive
1292 isolation and low genetic differentiation despite floral divergence across varying geographic scales in
1293 *castilleja*. *American Journal of Botany*, **108**, 1270–1288.
- 1294 Wilkinson, D.P., Golding, N., Guillera-Arroita, G., Tingley, R. & McCarthy, M.A. (2021). Defining and
1295 evaluating predictions of joint species distribution models. *Methods in Ecology and Evolution*, **12**, 394–
1296 404.
- 1297 Williams, P.H. (1982). The distribution and decline of british bumble bees (*bombus latr.*). *Journal of*
1298 *Apicultural Research*, **21**, 236–245. Retrieved from <https://doi.org/10.1080/00218839.1982.11100549>
- 1299 Wilson, A.M. & Jetz, W. (2016). Remotely sensed high-resolution global cloud dynamics for predicting
1300 ecosystem and biodiversity distributions. *PLoS biology*, **14**, e1002415.
- 1301 Wood, D.E., Lu, J. & Langmead, B. (2019). Improved metagenomic analysis with kraken 2. *Genome biology*,
1302 **20**, 1–13.
- 1303 Xie, Y., Wang, X. & Silander Jr, J.A. (2015). Deciduous forest responses to temperature, precipitation, and
1304 drought imply complex climate change impacts. *Proceedings of the National Academy of Sciences*, **112**,
1305 13585–13590.
- 1306 Xie, Y., Wang, X., Wilson, A.M. & Silander Jr, J.A. (2018). Predicting autumn phenology: How deciduous
1307 tree species respond to weather stressors. *Agricultural and Forest Meteorology*, **250**, 127–137.
- 1308 Yoccoz, N.G., Nichols, J.D. & Boulinier, T. (2001). Monitoring of biological diversity in space and time.
1309 *Trends in ecology & evolution*, **16**, 446–453.
- 1310 Zhao, Y.-H., Lázaro, A., Ren, Z.-X., Zhou, W., Li, H.-D., Tao, Z.-B., Xu, K., Wu, Z.-K., Wolfe, L.M., Li,
1311 D.-Z. & Wang, H. (2019). The topological differences between visitation and pollen transport networks:
1312 A comparison in species rich communities of the himalaya–hengduan mountains. *Oikos*, **128**, 551–562.
1313 Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1111/oik.05262>

1314 List of Figures

1315 1	Simplified Conceptual Diagram of three approaches leading towards classification of sequencing results, and the number of species associated with them in our area. The upper three boxes indicate a common coarse approach, assuming one has a digitized Flora, which is not always the case. The center two boxes indicate the computational approach illustrated here. The final box indicates the use of the expert field data in the case study. The stem of the final applies to both Computational and Expert Survey results, and should be thought of as using time ala chromatography, in the post-classification process.	73
1322 2	Comparision of Accuracy between the initial output data from BLAST, and these same data subjected to the post-classification process which removes surrogate, and temporally restricted species	74
1325 3	Modelled dates of when major flowering events occurred compared between long term and modelled data	75
1327 4	Modelled dates of when major flowering events occurred compared between 2015 and modelled data	76
1329 5	Number of the ten most commonly visited plants which are also in the top ten most common sequences	77
1331 6	Phylogenetic tree of all biotically pollinated plant genera in the study area. The innermost ring indicates every genus which Queen Bee's were observed to visit. The intermediate ring indicates that at least a single morphological pollen voucher slide was prepared for a member of the genus. The outermost ring indicates that sequence data were available for at least a member of that genus. Branch colors follow APG 4.	78
1336 7	Relationship between morphological count data and sequence reads	79

Filtering Species by Geography and Ecology
and stratifying over a temporal gradient

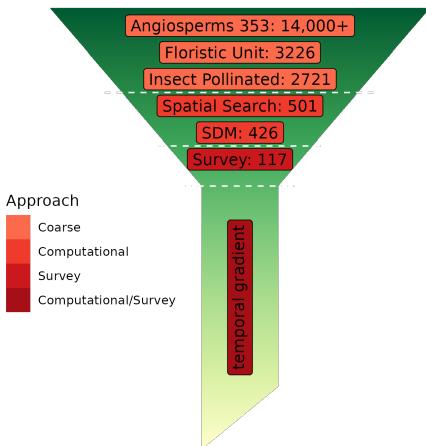


Figure 1: Simplified Conceptual Diagram of three approaches leading towards classification of sequencing results, and the number of species associated with them in our area. The upper three boxes indicate a common coarse approach, assuming one has a digitized Flora, which is not always the case. The center two boxes indicate the computational approach illustrated here. The final box indicates the use of the expert field data in the case study. The stem of the final applies to both Computational and Expert Survey results, and should be thought of as using time ala chromatography, in the post-classification process.

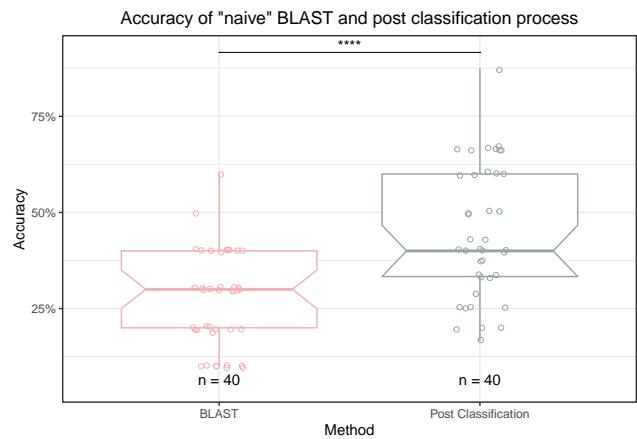


Figure 2: Comparision of Accuracy between the initial output data from BLAST, and these same data subjected to the post-classification process which removes surrogate, and temporally restricted species

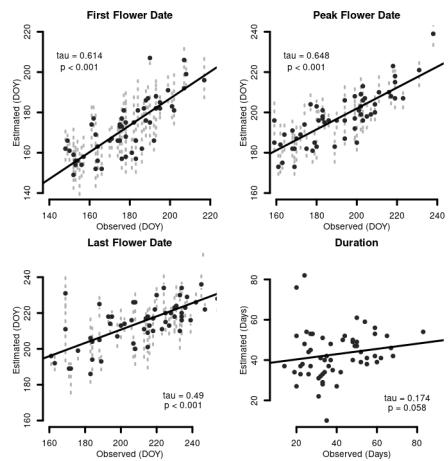


Figure 3: Modelled dates of when major flowering events occurred compared between long term and modelled data

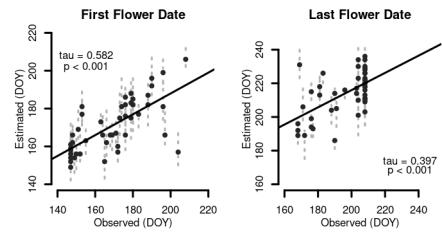


Figure 4: Modelled dates of when major flowering events occurred compared between 2015 and modelled data

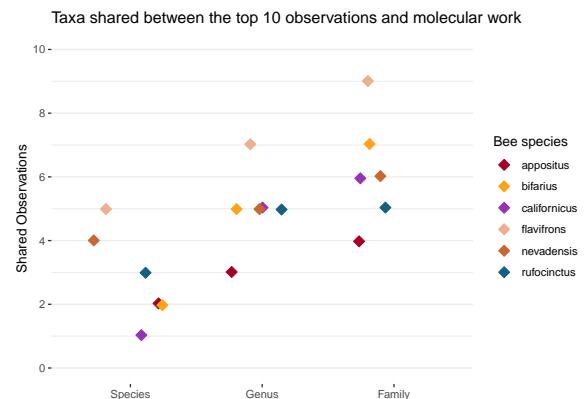
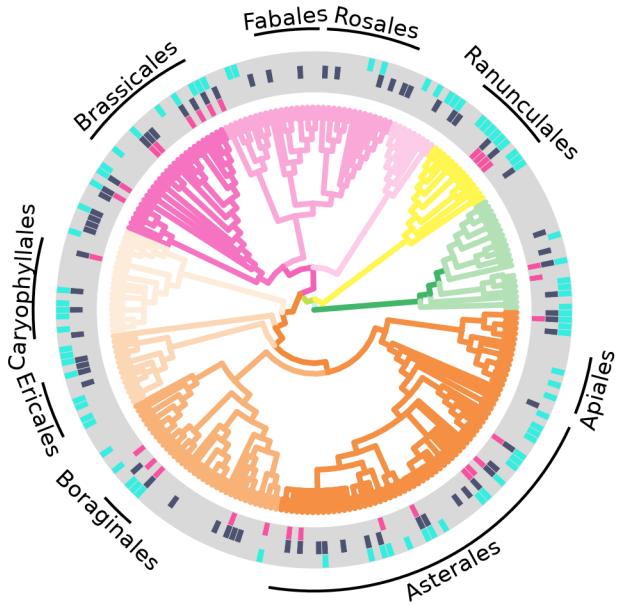


Figure 5: Number of the ten most commonly visited plants which are also in the top ten most common sequences

Biotically pollinated plant genera with morphological or molecular data



Status lacking observed sequenced slide

Figure 6: Phylogenetic tree of all biotically pollinated plant genera in the study area. The innermost ring indicates every genus which Queen Bee's were observed to visit. The intermediate ring indicates that at least a single morphological pollen voucher slide was prepared for a member of the genus. The outermost ring indicates that sequence data were available for at least a member of that genus. Branch colors follow APG 4.

Correlation of Proportion Counted Grains and Sequence Reads

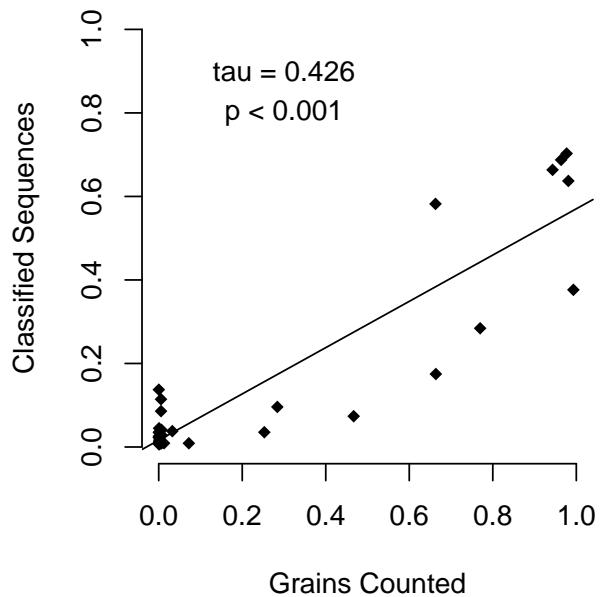


Figure 7: Relationship between morphological count data and sequence reads

₁₃₃₇ **List of Tables**

₁₃₃₈ 2	Applications of Plant Metabarcoding	81
₁₃₃₉ 3	Current Issues Facing Plant Metagenomics	82
₁₃₄₀ 4	Post classification of Sequences via Taxonomy and Ecology, top 15 most abundant reads . . .	83
₁₃₄₁ 5	Plant species detected in five or more corbiculae loads	84
₁₃₄₂ 6	Logistic regression assessing accuracy of SDMs	85
₁₃₄₃ 7	Species Distribution Modeling evaluation contingency table	86
₁₃₄₄ 8	Limitations Faced and Possible Solutions	87
₁₃₄₅ 9	Errors faced with components of the framework	88

Table 2: Applications of Plant Metabarcoding

Application	Example	Citations
Plant-Animal Interactions	Dietary Preferences of herbivores	Soininen et al. 2009; reviewed in Banerjee et al. 2022
Noxious Weed Detection	Presence of species by hydrologic Basin	Coghlan et al. 2021; Xu et al. 2018
Detection of Rare Species	Presence of Rare Aquatic Species; Others	Tsukamoto et al. 2021, reviewed in Banerjee et al. 2022
Forensic Science	Identifying the provenance of materials found at crime scene	Allwood et al. 2020
Pharmaceutical	Identifying adulterants in wholesale products	Bell et al. 2022

Table 3: Current Issues Facing Plant Metagenomics

Issue	Our Approach	Possible Advances
Taxonomic Resolution	A353	Coissac et al. 2016, Kress 2017, Johnson et al. 2023
Reference Library - Phylogenetic	Kew PAFTOL, no phylogenetic biases	Kress 2017, Bell et al. 2022, Johnson et al. 2023
Reference Library - Spatial	Some bias persists towards Europe	Cheng et al. 2018, Darwin Tree of Life 2022, Lewin et al. 2020, Bell et al. 2021
Reference Library Generation	Spatial Modelling; Code within	Bell et al. 2022
Uncertainty with Matches	Temporal Filter System	Bell et al. 2022
Species Surrogates	Temporal Filter System	?
False Positives	Spatial & Temporal Modelling, Jaccard Index, high quality reference loci	Bell et al. 2021

Table 4: Post classification of Sequences via Taxonomy and Ecology, top 15 most abundant reads

Condition	No. Class.	Prcnt. Class.	Total Seqs	Rank
A	143	21.0	32.0	Species
B	205	30.1	10.5	Species
C	5	0.7	0.4	Genus
G	29	4.3	7.8	Species
H	280	41.2	47.9	Genus
None met	18	2.6	1.4	Multiple

Table 5: Plant species detected in five or more corbiculae loads

Family	Genus	Species	No. Samples	Mean Prop. ^a
Asteraceae	Erigeron	sp.	6	0.5
	Senecio	integerrimus		
	Symphyotrichum	eatonii	5	
	Taraxacum	officinale	11	6.9
Boraginaceae	Mertensia	fusiformis	22	23.7
		ciliata	7	39.9
Celastraceae	Parnassia	palustris	5	0.5
Fabaceae	Lupinus	sericeus	23	15.5
Hydrophyllaceae	Hydrophyllum	fendleri	22	15.1
		capitatum	6	32.7
Ranunculaceae	Delphinium	barbeyi	7	45.9
		nuttallianum	21	70.0
Rosaceae	Dasiphora	fruticosa	7	0.6
Salicaceae	Salix	sp.	9	8.3
Violaceae	Viola	praemorsa	6	0.5

^a The mean only calculated across the samples where the species was detected

Table 6: Logistic regression assessing accuracy of SDMs

Metric	Value	Metric	Value
Accuracy (Training)	83.75	F-Score	0.84
Accuracy (Test)	84.00	AUC	0.92
Recall	81.03	Concordance	0.92
True Neg. Rate	86.97	Discordance	0.08
Precision	88.04	Tied	0.00

Table 7: Species Distribution Modeling evaluation contingency table

		Training		Testing	
		Absence	Presence	Absence	Presence
Absence	Absence	25620	3838	11130	1653
	Presence	6614	28248	2758	12024

Table 8: Limitations Faced and Possible Solutions

Method Component	Limitations	Paths Forward
Stage 1 Species Filter	Test Data	Flash Plant Species Surveys on Plot
Species Distribution Modelling	Number of Records; Taxonomically Difficult Groups	Develop and Disseminate Education Materials; Herbaria Collections
Phenological Modelling	Post-Initiation of Climate Change Records	Advocate Herbarium Collections
Database Generation	Adequate Phylogenetic/Spatial Representation	Plant and Fungal Tree of Life; 10kP
Read Re-assignment	Discrete Frequentist Data	Posterior-Probabilities; Floral Abundance, Nectar/Pollen Nutrition
False Positives	Which True Species?	Jaccard Index, Plot Abundance
Semi-Quantitative Inference	Genome Size, Pollen Grain Size?	Spike Samples with Reference Materials; Several C Sizes

Table 9: Errors faced with components of the framework

Component	Example	Correction
Spatial	"Scabrethia Scabra" (false positive)	<i>Wyethia arizonica</i>
Molecular (Reference Library)	"Asteraceae" (wrong order)	<i>Frasera</i>
Molecular Reference (Loci)	"Trollium" (false positives, same family as target)	<i>Lupinus</i> sp.
DNA difficult to extract	"Agastache" (Hemiparasites)	<i>Pedicularis</i>
Reclassification (Genus)	"Epilobium"	<i>Chamaenerion latifolium?</i>
Reclassification (Family)	"Paxistima myrsinoides"	<i>Parnassia palustris</i>