

¹ Metagenomic Barcoding of Pollen Loads Offers Insights on the
² Foraging Patterns of Queen Bumble Bees

³ Reed Clark Benkendorf^{*1,2}, Jane E. Ogilvie ³, Emily J. Woodworth ^{1,2}, Sophie Taddeo ^{1,2}, Paul J. Cara

⁴ **Abstract**

⁵ .
⁶ 1) DNA Barcoding has been remarkably successful in nearly all kingdoms of life and has allowed for the
⁷ rapid analysis of ecological assemblies. Successful DNA barcoding in the plant kingdom has been more
⁸ difficult than other kingdoms. Due to this understanding plants in ecological contexts and understandings
⁹ of their synecology in some instances may begin to lag behind other kingdoms.

¹⁰ 2) Here we utilize hyb-seq, museum studies, and species distribution modelling, to detect the plant species
¹¹ present in pollen loads collected from Queen Bumble Bees.

¹² 3) We show that hyb-seq using the Angiosperms 353 probes offers significant promise to metagenomic
¹³ approaches in real world scenarios.

¹⁴ 4) We conclude that these probes offer promise for the identification of plant tissue in both single sample,
¹⁵ and metasample contexts.

¹⁶ **1 | INTRODUCTION**

¹⁷ The inability to reliably identify plants to the level of species often leaves our understanding of ecosystem
¹⁸ function and interactions wanting. Current methods to ameliorate this situation include: ignoring these
¹⁹ ecologically relevant levels of detail, revisiting plots as diagnostic material becomes temporally available,
²⁰ assistance from taxonomic specialists, or the use of barcoding or other molecular techniques. These approaches
²¹ are untenable in light of the benefits offered by: species in several morphologically difficult genera which
²² serve as bioindicators, preferred partners in ecological interactions, as well as an increasing lack of taxonomic
²³ experts (Hebert *et al.* (2003)). Many genera, especially with the formalized advent of integrative taxonomy,
²⁴ have species which are well defined based upon ecological and behavioral rather than morphological properties,

*Correspondence: reedbenkendorf2021@u.northwestern.edu

25 the identification of these taxa in degraded areas or without their mutualistic partners is fraught with difficulty.
26 Hindering an understanding of the breadth of habitat which some species occupy, and the interactions they
27 have with other species.

28 The identification of many plant species to terminal taxon is an essential component of nearly all land
29 management programs, where many species in the same genus (e.g. Sagebrush - *Artemisia* L., Willows -
30 *Salix* L., and Sedges - *Carex* L.) serve as bioindicators (respectively for ‘rangelands’, streams, and wetlands),
31 as well as in academic research (Gage & Cooper (2013), AIM). This endeavour is often mired by lack of
32 diagnostic characters (e.g. flowers, fruits, roots or combinations thereof), and increasingly the description of
33 cryptic species (Janzen *et al.* (2017), Oliver *et al.* (2009)). Solutions to this problem are wanting, certain
34 programmes have relied increasingly upon revisiting field sites to identify material using morphological or
35 chemical approaches, whereas academic research has often used high copy number plastid genes as barcodes
36 (Rosentreter *et al.* 2021, MORE MORE). However, both approaches have significant downsides, the former
37 resource intensive at the landscape scale - and often does not work, while the latter seldom works due to a
38 lack of variability in the currently available barcodes (Liu *et al.* (2014)).

39 Recently barcoding, and metabarcoding, have shown considerable promise in all Kingdoms of life. For example
40 With plants the identification of members of certain clades has been quite successful, whereas with
41 others results have been elusive (Liu *et al.* (2014), Group *et al.* (2011)), while most applications laying along
42 this spectrum (Li *et al.* (2015), Kress & Erickson (2007), Group *et al.* (2009)). Particular challenges with the
43 utilization of high-copy number sequences are associated with their rates of divergence, gene tree conflict,
44 and hybridization (Coissac *et al.* (2016), Fazekas *et al.* (2009)) Herein we have resolved major components
45 of the problems of identifying plant material without diagnostic morphological character states using the
46 Angiosperms353 (A353) Hyb-Seq probes (Johnson *et al.* (2019)), and custom species sequence databases
47 derived via species distribution modelling, and temporal filtering.

48 Our foundation for increasing the quality of metabarcoding results in plants is reducing the number of possible
49 plant species candidates by generating user selected sequence databases relevant to the spatial extent of the
50 study region. While there are numerous possible approaches for this process, we achieve the selection of
51 possible plant candidate species using digital collections gleaned from herbaria, survey work, and citizen
52 science (e.g. iNaturalist), from a domain exceeding the study area. To these candidate species, modelling
53 approaches - such as logistic regression, may be used to identify distances under which taxa warrant further
54 exploration. To these candidate species, we generate species distribution models (SDM’s), which indicate
55 the probability of suitable habitat in a domain, and base the inclusion of these taxa, or representative
56 congeners, upon these results. This approach has the additional benefit of greatly reducing the size of a

57 sequence database, which allows for the usage of genomic size data on personal computers. Moreover, as
58 most next-generation sequence data is deposited as raw-sequence reads, from a processing perspective, it is
59 essential to reduce the candidate species via an approach as such.

60 Currently the largest plant systematic endeavor ever undertaken, the Kew Plant and Fungal Tree of Life
61 (PAFTOL), is approaching completion (Baker *et al.* (2021a)). This dataset will contain Hyb-Seq data from
62 at least one species representing each genus in the plant kingdom using the popular A353 probes (Baker *et*
63 *al.* (2021b)), resulting in over 14,000 represented species. These publicly available data serve to provide a
64 taxonomically comprehensive backbone for plant metabarcoding. Data from the 10kP project, which seeks to
65 develop reference genomes from a phylogenetically diverse suite of plants will contribute many more records
66 upon its intended completion, now slated to be by 2030, similar projects which seek to sequence high amounts
67 of genomes in regions e.g. the ‘Darwin Tree of Life’ are being undertaken which will contribute data for
68 applicable to enormous spatial domains (Cheng *et al.* (2018), Life Project Consortium *et al.* (2022), Lewin *et*
69 *al.* (2022)).

70 Considerable amounts of species interactions are expressed along time (CaraDonna *et al.* (2021)). For the
71 tropics the flowering periods of many plant species display high seasonality, and given the elevated rates of
72 species richness relative to the temperate, this axis may provide an essential filter for identifying material in
73 many metagenomic samples (Janzen (1967), Newstrom *et al.* (1994)). While many pollination interactions
74 are formed and dissolved along the temporal axis in the temperate regions (CaraDonna *et al.* (2017)), the
75 overall shorter extent of the active growing season in these systems results in the presence of few to any
76 natural breaks in these systems which subjugates the utility of these to perform as filters of post-processing
77 results, rather than distinct species assemblage for database generation. Nonetheless, we work through a
78 process which seems applicable to the tropics to utilize the temporal dimension for classifying sequencing
79 results.

80 We apply these metagenomic and informatics approaches to determine whether the foraging record of Queen
81 Bumble Bee’s is consistent across direct observations and the pollen record, an incongruency in several
82 floral visitation networks (Barker & Arceo-Gomez (2021), Zhao *et al.* (2019), Alarcón (2010)). The two
83 foraging phases of the Queen Bumble Bee life cycle is essential to 1) increase their weight before diapause, 2)
84 increase their ovary weights while establishing their recently found nests, both of these time periods represent
85 potential demographic bottlenecks in bumble bee populations (Sarro *et al.* (2022)). Bumblebees are one of
86 the only groups of insects with unequivocal quantitative evidence for numerous populations declines, while
87 simultaneously serving as the most effective pollinators in temperate montane ecosystems (Cameron & Sadd
88 (2020), Goulson *et al.* (2008), Williams (1982), Colla *et al.* (2012), Bergman *et al.* (1996), Bingham & Orthner

89 (1998)). These montane ecosystems represent some of the most ecologically resilient and resistant systems in
90 the temperate and offer unparalleled potential as refugial areas for multiple dimensions of biodiversity under
91 climate change.

92 2 | METHODS

93 Study System & Field Work

94 Observations and sample collection was conducted at The Rocky Mountain Biological Laboratory (RMBL;
95 38°57.5" N, 106°59.3" W (WGS 84), 2900 m.a.s.l.), Gunnison County, Colorado, USA (*APPENDIX 1* for
96 site information). Pollinator observations of *Bombus* Latreille spp. (Apidae Latreille) were conducted from
97 June - August of 2015. Observations of *Bombus* foraging took place for one hour at each field site in
98 three 100m transects, where all flowers were also counted and placed into abundance bins. Corbiculae loads
99 were, non-lethally, collected once from all Queen individuals encountered. The six study sites are in areas
100 characterised by high-montane/subalpine Parkland vegetation communities.

101 2.1 | Spatial Analyses

102 **2.1.1 Candidate Species** To develop an ecologically relevant list of vascular plant species, with expected
103 biotic pollination, which may be present at the study sites all records adjacent to the field site were downloaded
104 from the Botanical Information and Ecology Network ‘BIEN’ (Maitner (2022)), and these taxa had Species
105 Distribution Models (SDMs) generated to predict their suitability. The predicted plant species served as a
106 reference for which species to include in the genomic sequence databases.

107 In order to minimise the number of species for which SDM's were to be generated, BIEN was queried at
108 a distance of up to 100km from our field site and all plant species records were downloaded. In order to
109 emulate the perceived stochasticity of collections, this dataset was bootstrap re-sampled 250 times, with
110 90% of samples selected, to create a testing dataset. The median of the logistic regression assessing the
111 probability of occurrence of a species record as a function of distance from the study area was used as a
112 threshold distance, under which, to include species as candidates for distribution modelling.

113 **2.1.2 Distribution Modelling** Species had all records from BIEN within a 50km border of the Omernik
114 level 3 ecoregion which the site is located in (*No. 21 “Southern Rockies”*), downloaded (n = 23,919) (Omernik
115 (1987)). These records were copied into two, initially identical, sets, one for generating machine learning
116 models (Random Forest, and Boosted Regression Tree's), and the other for Generalised Linear (GLM) and

117 Generalized Additive Models (GAM). The set for generating GLM and GAM records was thinned to reduce
118 spatial autocorrelation in the dataset, as measured by Morans Index (Moran (1950), Bivand & Wong (2018)).
119 To both datasets an additional 4029 plots collected from a random stratification of 19% of the land cover in
120 the area of analysis were searched to create true absences (BLM CITATION - need appropriate format for
121 journal). To achieve a larger absence dataset 1000 pseudo-absence records were generated for each taxon,
122 each of which was greater than 10km from an occurrence record. For ML models, these pseudo-absences
123 were reduced so that the ratio of presence to absence records were balanced. To achieve this, absence records
124 inside of 10% of the mean sample value of the presence records, for any predictor were removed; the required
125 number of absence records were then randomly sampled.

126 Species abiotic niche predictors were 26 variables at 30m resolution, six related to climate, five soil, four
127 topographic, four related to cloud cover, with the remaining reflecting assorted abiotic parameters (Wilson &
128 Jetz (2016), Wang *et al.* (2016), Hengl *et al.* (2017), Robinson *et al.* (2014)) (APPENDIX 6). For linear
129 regression models these predictors underwent both vifstep (theta = 10, max observations = 12,500) and vifcor
130 (theta = 0.7, max observations = 12,500), and collinear features were removed leaving 16 variables (Naimi *et*
131 *al.* (2014)).

132 Modelling: Random Forest and Boosted Regression Trees, were sub sampled with 30% test and two replicates
133 each before weighted ensemble based on True Skill Statistics (tss) (Naimi & Araujo (2016)). Generalised
134 linear models (GLM) and Generalised additive models (GAM) with 30% sub sampling and three replicates
135 each were also ensembled using the tss (Naimi & Araujo (2016)). The results of these models were extracted
136 to a polygon feature derived from a minimum-spanning tree which encompasses the study area, and species
137 from either ensemble with greater than 50% habitat suitability were considered present for further purposes
138 (Prim (1957)).

139 535 species were modelled using Generalized Linear Models and Generalized Additive Models. 534 species
140 were modelled using Random Forest and Boosted Regression Trees. To evaluate the accuracy of the species
141 distribution models, additional presence records from GBIF ($n = 61,789$), and AIM ($n = 12,730$) were used as
142 test and training sets ($n = 74,519$) for logistic regression (CITE AIM AND GBIF). Additional novel absence
143 records were generated from the AIM dataset to create a dataset where each species has balanced presence
144 and absences. 11 or more paired presence and absence records were required for this testing, resulting in 334
145 species being included in the logistic regression ($Mdn = 110.0$, $\bar{x} = 223.1$, max = 1568 record pairs used)
146 with a 70% test split (Kuhn (2022)).

¹⁴⁷ **2.2 | Molecular Lab Work**

¹⁴⁸ All lab work was carried out at The Daniel F. and Ada L. Rice Plant Conservation Science Center at the
¹⁴⁹ Chicago Botanic Garden, Glencoe, Illinois, U.S.A.

¹⁵⁰ **2.2.1 | Reference Plant Library Generation** Using 5 years of observational data on *Bombus Queen*
¹⁵¹ Bee foraging at these studies sites, we identified the plant taxa most frequently visited by Queens across
¹⁵² all years. We sequenced the 12 most commonly visited taxa twice using samples from one site within the
¹⁵³ Gunnison River Drainage and one individual from another population. In addition, for any of these 12 focal
¹⁵⁴ species which did not have a congener pair in this filtered sample, we included a congener - or a species from
¹⁵⁵ a closely related genus to serve as an outgroup. We also sequenced another 15 abundant taxa commonly
¹⁵⁶ visited by *Bombus* workers, based on the aforementioned data set (*APPENDIX 4*).

¹⁵⁷ **2.2.2 | Plant Genomic DNA Extraction** Plant genomic DNA was isolated from ~ 1 cm² of leaf tissue
¹⁵⁸ from silica-gel dried or herbarium material using a modified cetyltrimethylammonium (CTAB) protocol
¹⁵⁹ (Doyle & Doyle (1987)) that included two chloroform washes. DNA was quantified using a Nanodrop 2000
¹⁶⁰ (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and Qubit fluorometer (Thermo Fisher Scientific).

¹⁶¹ **2.2.3 | Pollen Genomic DNA Extraction** Pollen genomic DNA was extracted from corbiculae using a
¹⁶² CTAB based protocol modified from Lahlamgiah et al. and Guertler et al. (2014, 2014). A SDS extraction
¹⁶³ buffer (350µL , 100mM Tris-HCl, 50 mM EDTA, 50 mM NaCl, 10% SDS v/v., pH 7.5) was added followed
¹⁶⁴ by vortexing to allow dissolution of corbiculae. Pollen grains were then macerated with Kontes Pellet Pestles,
¹⁶⁵ and the tip of these washed with 130 µL of the SDS extraction buffer, samples were then incubated for 1 hour
¹⁶⁶ at 30°C. This was followed by the addition of 10% CTAB solution (450ul, of 20 mM Tris-Cl pH. 8.0, 1.4 M
¹⁶⁷ NaCl, 10 mM EDTA pH 7.5, 10% CTAB, 5% PVP, ~85% Deionized water) and RNase (10 uL of 10 mg/mL)
¹⁶⁸ and samples were incubated for 40 minutes at 37°C, on heat block (Multi-Blok, Thermo Fisher Scientific,
¹⁶⁹ Waltham Massachusetts) set to 40°C. After 20 minutes incubation, Proteinase K (15 µL of 20mg/ml) and
¹⁷⁰ DTT (12.5 µL of 1M in water) were added, and the samples were further incubated at 60°C for 1 hour.
¹⁷¹ Samples were then incubated overnight at 40°C. 500 µL of Phenol-Chloroform-Isoamyl alcohol (25:24:1) were
¹⁷² added, vortexed, and centrifuged at 10,000 rpm for 10 minutes and the aqueous phase was pipetted to a 1.5
¹⁷³ ml centrifuge tube.

¹⁷⁴ To precipitate the DNA, chilled Isopropyl alcohol & 3 mM Sodium acetate (5:1) equivalent to 2/3 of the
¹⁷⁵ volume of sample were added, with 1 hour of chilling at -20°C, followed by 10 minutes of centrifuging at
¹⁷⁶ 13,000 rpm. The supernatant was pipetted to a new 1.5 ml centrifuge tube, and 70% EtOH (400 µL) were

177 added before chilling at -20°C for 20 minutes followed by centrifugation at 13,000 rpm for 10 minutes. Both
178 tubes were then washed with 75% EtOH (400 µL), inverted, centrifuged at 13,000 rpm for 4 minutes, and the
179 solution discarded, then washed with 95% EtOH (400 µL) , inverted, centrifuged at 13,000 rpm for 4 minutes,
180 and the solution discarded. Pellets were dried at room temperature overnight before resuspension in Nuclease
181 free H₂O. Extractions were assessed using a Nanodrop 2000 (Thermo Fisher Scientific) and Qubit fluorometer
182 (Thermo Fisher Scientific). DNA extracts were then cleaned using 2:1 v./v. Sera-Mag beads (Cytiva, Little
183 Chalfont, UK) to solute following the manufacturer's protocol, eluted in 0.5x TE, and the eluent allowed to
184 reduce by half volume in ambient conditions. DNA was quantified using a Qubit fluorometer.

185 **2.2.4 | Fragmentation, Library Preparation & Target Enrichment** Library preparation was per-
186 formed using the NEBNext Ultra II FS-DNA Library Prep Kit for Illumina (New England BioLabs, Ipswich,
187 Massachusetts, USA) using slightly modified manufacturers recommendation. Fragmentation was performed
188 at ½ volume of reagents and ¼ enzyme mix for 40 minutes at 37°C, with an input of 500 ng cleaned DNA.
189 Adapter Ligation and PCR enrichment were performed with ½ volumes, while cleanup of products was
190 performed with ½ volume of SPRI beads (Beckman Coulter, Indianapolis, Indiana, USA) and recommended
191 volumes of 80% v./v. ethanol washes. The exception was the herbarium specimens which were not fragmented
192 and only end repaired, with similar library preparation of all samples. Products were analysed on 4% agarose
193 gels, and a Qubit fluorometer. Libraries were pooled and enriched with the Angiosperms 353 probe kit V.4
194 (Arbor Biosciences myBaits Target Sequence Capture Kit) by following the manufacturer's protocol and
195 Brewer et al. 2019. Sequencing was performed using an Illumina mi-Seq with 150-bp end reads, (NUSeq Core,
196 Chicago, Illinois).

197 **2.2.5 | Computational Processes and Analyses.**

198 **2.2.5.1 | Reference Library Data Processing** Sequences were processed using Trimmomatic, which
199 removed sequence adapters, clipped the first 3 bp, discarding reads less than 36 bp, and removing reads
200 if their average PHRED score dropped beneath 20 over a window of 5 bp (Bolger & Giorgi (2014), Tange
201 (2021)). Contigs were generated using HybPiper using target files created by M353 (Johnson *et al.* (2016),
202 McLay *et al.* (2021)).

203 **2.2.5.2 | Sequence Identification** A custom Kraken2 database was created by downloading representative
204 species of each genus indicated as being present in the study area by the spatial analyses from the Sequence
205 Read Archive (SRA) NCBI (Wood *et al.* (2019)). These sequences were processed in the same manner as
206 our novel sequences before being placed into the database. The Kraken2 database was built using default

parameters. Kraken2 was run on sequences using default parameters (*APPENDIX 5*). Following Kraken2, Bracken was used to classify sequences to terminal taxa (Lu *et al.* (2017)). Results from both Kraken2 and Bracken, results were reclassified manually to identify terminal taxa. For example, when only a single species of a genus was known in the study area, but our database used a representative of another taxon in the genus, this species was coded as the result. The re-coding of sequences from another representative species for the genus to the sole RMBL representative allowed the identification of XX & % more species.

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

2.2.5.4 | Morphological Pollen identification

To develop a reference library of pollen grains which may be present in corbiculae loads, an image reference collection of fuchsin-jelly stained (Beattie (1971)) slides was assembled from slides previously prepared by the authors (n = 21), and other researchers (n = 38) (Brosi & Briggs (2013)). Using 5 years of observational data on *Bombus* Queen Bee foraging at these studies sites (Ogilvie unpublished), as well as the Vascular Plant Checklist (Frase & Buck (2007)), an additional 62 voucher slides for species were prepared and imaged at 400x (Leica DMLB, Leica MC170 HD Camera, Leica Application Suite V. 4.13.0) from non accessioned herbarium collections to supplement the number of species and clades covered (Appendix 3).

In order to determine which plant taxa were distinguishable via light microscopy, and to develop a dichotomous key to pollen morphotypes, Divisive Hierarchical Clustering techniques were used. Ten readily discernible categorical traits were collected from each specimen in the image collection. These traits were transformed using Gower distances, and clustered using Divisive Hierarchical clustering techniques (Maechler *et al.* (2022)).

Using the cluster dendrogram, elbow plot, and heatmaps (Hennig (2020)), of these results morphological groups of pollen which could not be resolved via microscopy were delineated, and a dichotomous key was prepared (APPENDIX NO.). This key was then used to identify the pollen grains sampled from corbiculae loads to

morphotypes in a consistent manner. To prepare the pollen slides from corbiculae, all corbiculae loads were broken apart and rolled using dissection needlepoints to increase heterogeneity of samples. *Cerca* 0.5mm² of

pollen was placed onto a ~4mm² fuchsin jelly cube (Beattie (1971)) atop a graticulated microscope slide, with 20 transects and 20 rows (400 quadrants) (EMS, Hartfield, PA). The jelly was melted, with stirring, until pollen grains were homogeneously spread across the microscope slide. Slides were sealed with Canada Balsam (Rublev Colours, Willits, CA) followed by sealing with nail polish; all samples are noted in APPENDIX 3.

To identify the pollen present in corbiculae loads, light microscopy at 400x (Zeiss Axioscope A1) was used. In initial sampling in three transects, each pollen grain was identified to morphotype and counted; an additional

238 two transects were scanned for morphotypes unique to that slide, if either transect contained an unique
239 morphotype than all grains in that transect were also identified and counted. Subsequent to the first round
240 of sampling, non-parametric species richness rarefaction curves (Oksanen *et al.* (2022)), and non-parametric
241 species diversity rarefaction curves were used to assess the completeness of sampling (Chao *et al.* (2014),
242 Hsieh *et al.* (2020)). Slides not approaching the asymptote of the rarefaction curve were then re-sampled,
243 and analysed iteratively for up to a total of seven transects *APPENDIX 2*.

244 **2.3 | Temporal Analyses**

245 To estimate the duration of dates in which plant species were flowering weibull estimates of several phenological
246 parameters all spatially modelled taxa were developed (Belitz *et al.* (2020), Pearse *et al.* (2017)). Only
247 BIEN records which occurred in the Omernik Level 4 Ecoregions within 15km of the study area ($n = 5$, or
248 conditionally 6 if enough records not be found in the nearest 5), and which were from herbarium records
249 were included. To remove temporally irrelevant herbarium records, i.e. material collected during times which
250 flowering is impossible at the study area due to snow cover, the SnowUS dataset (Iler *et al.* (2021), Tran *et al.*
251 (2019)) from 2000-2017 was analyzed for the first three days of contiguous snow absence, and the first three
252 days of contiguous snow cover in Fall. Herbarium records after the 3rd quantile for melt, and the 1st quantile
253 for snow cover of these metrics were removed. Species with > 10 records had their weibull distributions
254 generated for the date when 10% of individuals had begun flowering, when 50% were flowering, and when
255 90% of individuals had flowered.

256 **2.4 | Floral Observations**

257 **3 | RESULTS**

258 **3.1 | Spatial Analyses**

259 [Table 1 about here.]

260 [Table 2 about here.]

261 The median (25.009 km) of the logistic regression assessing the probability of occurrence of a species record as
262 a function of distance from the study area was used as a threshold distance to include species for distribution
263 modelling. A 2-sample test for equality of proportions with continuity correction ($X^2 = 13.254$, df
264 = 1, p-value = 0.000136, 95% CI 0.04-1.00) was used to test whether more of the records located in the
265 broad ecological sites present at the field station, between the distance of the median (25.009 km) to the

266 third quantile (ca 43.830 km) of the regression distance, where true presences at the field station. Including
267 these records would have resulted in modelling an additional 222 species distributions of which 30 are true
268 presences these taxa were not modelled.

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

271 At the field site, of the 554 vascular plants with biotic pollination syndromes, the 493 ML ensembles accurately
272 predicted the presence of 362 (65.3%), incorrectly predicted the presence of 64 (11.6%), incorrectly predicted
273 34 true presences (6.1%) as being absent, and correctly predicted the true absence of 33 (6.0%). The balanced
274 accuracy of the ensembled models is 0.627 (Sensitivity = 0.340, Specificity 0.914). Of the 554 vascular plants
275 with biotic pollination syndromes, the 475 LM ensembles accurately predicted the presence of 286 (51.6%),
276 incorrectly predicted the presence of 41 (14.3%), incorrectly predicted 93 true presences (16.8%) as being
277 absent, and correctly predicted the true absence of 55 (9.9%). The balanced accuracy of the ensembled
278 models is 0.664 (Sensitivity = 0.573, Specificity 0.754). Of the 554 vascular plants with biotic pollination
279 syndromes in the flora 13 (2.3%) were in the Orchid family and 41 (7.4%) are non-natives, both of which are
280 restricted from the database, and can only reduce the number of true predicted presences by roughly 10%.

281 At the six study plots, of the 117 plant species identified to the species level across the spatial extents of all
282 plots and duration of queen bee activity, the ML ensembles predicted the presence of 105 (89.7%) of them,
283 and LM ensembles 102 (87.2%). Of the missing species two (1.7%) are Orchids, six (5.1%) are non-native,
284 and one (0.85%) is of contested taxonomic standing, all of which (7.65%) are restricted from the initial query
285 database.

286 3.2 | Microscopic Pollen identification

287 Using the fuchsin jelly preparation and light microscopic analyses of grains and scoring of 10 character
288 states resulted in the establishment of 28 morphotypes which grains could be classified into. APPENDIX
289 7. 60 samples were counted and based on rarefaction **had over % of expected morphotypes found**
290 (morphotype richness, $\bar{x} = 4.5$, Mdn = 4, min = 1, max = 9), all samples had expected morphotype diversity
291 reach the asymptote APPENDIX 8. The number of counted pollen grains in each sample range from (MIN -
292 16,293, $\bar{x} = 2788.685$, Mdn = 1453).

293 **3.3 | Metabarcoding Pollen identification**

294 54 corbiculae loads had DNA extracted and underwent various steps towards hyb-seq, in the end a total of 44
295 corbiculae samples were sequenced, 7,752,353 reads were recovered from sequencing. The number of reads
296 per sequence varied widely (range = 76 - 508,795, $\bar{x} = 176,189.8$, Mdn = 138,395). Of the possible 353 loci,
297 the number which were recovered from each sample, and informative to BLAST were range = 24 - 353, $\bar{x} =$
298 305.5, Mdn = 331. The number of reads per loci from across all samples had a range of 178 - 506,653, $\bar{x} =$
299 20,688, Mdn = 12,616.

300 **APPENDIX X.**

301 After trimming 7,865,680 sequences remained. 10,682,538 reads were matched using Kraken, of the reads
302 classified by Kraken 10,160,768 reads were matched using Bracken, of the reads classified by Kraken 7,302,876
303 reads were matched using BLAST.

304 Based upon subjective review of the three classifiers, BLAST was chosen as the classification method which
305 yielded the most probable results, and it's values were used for all subsequent analyses.

306 **3.4 | Temporal Analyses**

307 The first date of modeled snow melt in the Gothic area ($n = 17$, $\bar{x} = 137.9$, Mdn = 135, 3rd quantile = 151),
308 and the first date of a consistent winter snow base ($n = 17$, $\bar{x} = 299.9$, Mdn = 300, 1st quantile = 291) from
309 2000-2017, were used as delimiters for the inclusions of herbarium records in modelling. Of the **500** species
310 predicted likely present in the area via logistic regression, 332 species (64.4%) with more than 10 records
311 in the focal level 4 ecoregions ($\bar{x} = 35.01657$, Mdn = 35, max = 96) had weibull estimates calculated, an
312 additional 56 species (11.2%) with enough contributing records from the 'Sedimentary Mid-Elevation Forests',
313 a large ecoregion in general just beneath the elevation bands occupied by the five ecoregions around the study
314 area had weibull estimates also calculated ($\bar{x} = 13.86885$, Mdn = 13, max = 24).

315 Only 58 of these 388 species ($n = 34.56897$, Mdn = 31) were able to be compared to plot based observational
316 data from the long term (1974–2012) dataset. Of these species relatively high accord was observed between
317 the long-term ground truthed dataset, and the modelled species. There was very strong evidence that the
318 weibull estimates were positively associated with the observed onset ($r^2 = 0.72$, $p < 0.0001$, $\tau = 0.61$) and
319 peak ($r^2 = 0.70$, $p < 0.0001$, $\tau = 0.65$) of flowering, and that the number of herbarium samples had a
320 moderate effect on the estimates ($p = 0.004$ and $p = 0.034$ respectively). There was very strong evidence
321 that the weibull estimates had a positive association with the observed cessation of flowering ($r^2 = 0.4339$, p
322 < 0.0001 , $\tau = 0.489$), however their was no evidence that sample size had an effect ($p = 0.349$). There was

323 moderate evidence that the weibull estimates, with an effect of sample size, had a weak positive association
324 with the observed duration of flowering ($p = 0.0401$, $r^2 = 0.07$, $\tau = 0.17$).

325 [Figure 1 about here.]

326 3.5 | Floral Observations

327 The six sites were surveyed for a total of 52 hours from May 27-July 27. A total of 723 queen-pollen foraging
328 interactions were observed (range per bee species by week range = 1 - 18, $\bar{x} = 3.46$, Mdn = 2), with a range
329 of total observed interactions per bee species across this time period (min = 1, $\bar{x} = 59.08$, Mdn = 19, max =
330 184). Plants varied widely in the number of interactions which they partook in with each species of bee (range
331 per plant species by week min = 1 - 20, $\bar{x} = 3.51$, Mdn = 2), with a range of total observed interactions per
332 plant species over this time period (min = 1, $\bar{x} = 20.26$, Mdn = 4, max = 141). The number of plant species
333 which bees were observed interacting with varied more narrowly (range = 1 - 18, $\bar{x} = 8$, Mdn = 6).

334 A total of 66 corbiculae loads were collected from Bees, 64 of them from Queens.

335 4 | DISCUSSION

336 Although we were able to use an actually fine scale flora to determine the species present at the field site,
337 we suspect a similar approach may be accomplished via quick species richness inventories at sites, and then
338 utilizing a bootstrap approach akin to ours, to the taxa returned from databases to derive these estimates.

339 ** Mention in here that the general effect of sample size on predictions means that larger samples sizes are
340 required for this to work!!!! ** Although our temporal results were lackluster, we note that our study area
341 has an incredibly brief growing period. and we suspect these temporal results would be useful in sub-tropical
342 and tropical ecosystems. FURTHER, the sites used for ground truthing the temporal flowering periods
343 were not randomly selected across the study area, and cannot be used to make inference to the population
344 across the entirety of the study area as we did here. Regardless both show good agreement on flower onset,
345 peak flowering, and moderate agreement with flowering cessation. The disagreement in flowering cessation
346 is perhaps due to more microclimates which retain water, rather than microclimates which allow the early
347 accumulation of heat.

348 Fewer modelling runs for SDM's likely to be effective for determining inclusion, elastic inclusion criteria. The
349 actual dataset which was used for training and testing all of the models incorporated into SDM's represented
350 only roughly one quarter of the records available for such purposes. We consciously chose to do this in order

351 to showcase the possibility of this approach working in less data rich areas.

352 Bayesian framework

353 Future Directions:

354 While at the time of writing this there are limited A353 sequence data, the Plant and Fungal Trees of Life
355 (PAFTOL) project, which is sequencing at least a species of each genera in the plant Kingdom will produce
356 sequence data from over 14,000 species. Given the extant publicly available genomic data, we conservatively
357 estimate that upon completion of PAFTOL there will be no fewer than 15,500 species (4.4% of all ca. 350,000
358 plant species) for which sequence data of a majority of these loci exist (Govaerts *et al.* (2021)). Accordingly,
359 projects in the near future may increase the number of metagenomics samples while decreasing the need to
360 create their own plant sequence reference libraries. As a result of PAFTOL the first ever comprehensive
361 phylogenetic hypotheses of all plant genera will be presented. In tandem with an increased number of
362 digitised and geo-referenced herbarium specimens, and monitoring programs in natural areas, we believe that
363 geo-informatics, and phylogenetic inference will increase the ability of researchers applying this technique
364 to identifying sequence reads. While our approach emphasises the use of this metagenomic technique for
365 the purpose of identifying pollen, I argue the template and resources we provide here make this approach a
366 suitable candidate for many plant metagenomic tasks. While we did not have the resources to explore the
367 possibility of characterising infraspecific characteristics, preliminary results from others (Wenzell *et al.* (2021),
368 Loke *et al.* in prep) indicate a possibility for these probes to also collect data at the level of populations and
369 individuals. **

370 In regards to better understanding the foraging preferences of *Bombus* feeding in subalpine ecosystems.

371 JANE AND PAUL SET UP FOR NEAR FUTURE RESULTS?

372 **AUTHOR CONTRIBUTIONS:** R.C.B conducted botanical collections, conducted all molecular lab work,
373 lead all analyses, and writing. J.E.O conceived, designed, and conducted all ecological fieldwork, assisted
374 with analyses, and writing. E.J.W. prepared, imaged, and collected trait data on pollen reference slides, and
375 assisted with analysis of trait data and writing a dichotomous key. S.T. assisted with spatial analyses and
376 writing. P.J.C assisted with ecological analyses and writing. J.B.F. conceived, and designed all lab work,
377 analyses, assisted with writing, and secured funding for molecular work.

378 **ACKNOWLEDGMENTS:** Nyree Zerega for assistance obtaining herbaria loans and accessioning our
379 collections at CHIC. Pat Herendeen for assistance with virtually all aspects of preparing pollen vouchers and
380 the identification process. Hilary Noble, Zoe Diaz-Martinez, Angela McDonnell, & Elena Loke for assistance
381 with genomic library preparation. We thank the curators at the following herbaria for supplying tissue: Ben

382 Legler at Stillinger (ID), Charles (Rick) Williams at Ray J. Davis (IDS), (B)Ernie Nelson at Rocky Mountain
383 (RM). We thank the original collectors of these specimens: D. Knoke, L. Brummer, J. Boyd, C. Davidson, I.
384 Gilman, M. Kirkpatrick, S. McCauley, J. Smith, K. Taylor, & C. Williams. David Giblin for sharing relevant
385 sections of an advanced draft of FNA V. 15. The Bureau of Land Management is thanked as many plant
386 specimens were collected by R.C.B as a partner or contractor to the agency, Sarah Burnett and Lauren Price
387 are thanked for sharing AIM data. Sanda and NEB are gratefully acknowledged for technical support and
388 generously sharing samples.

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

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

391 **DATA AVAILABILITY STATEMENT** The queries required to download all data used in this project
392 are located in.... All novel sequencing data are located at NCBI...

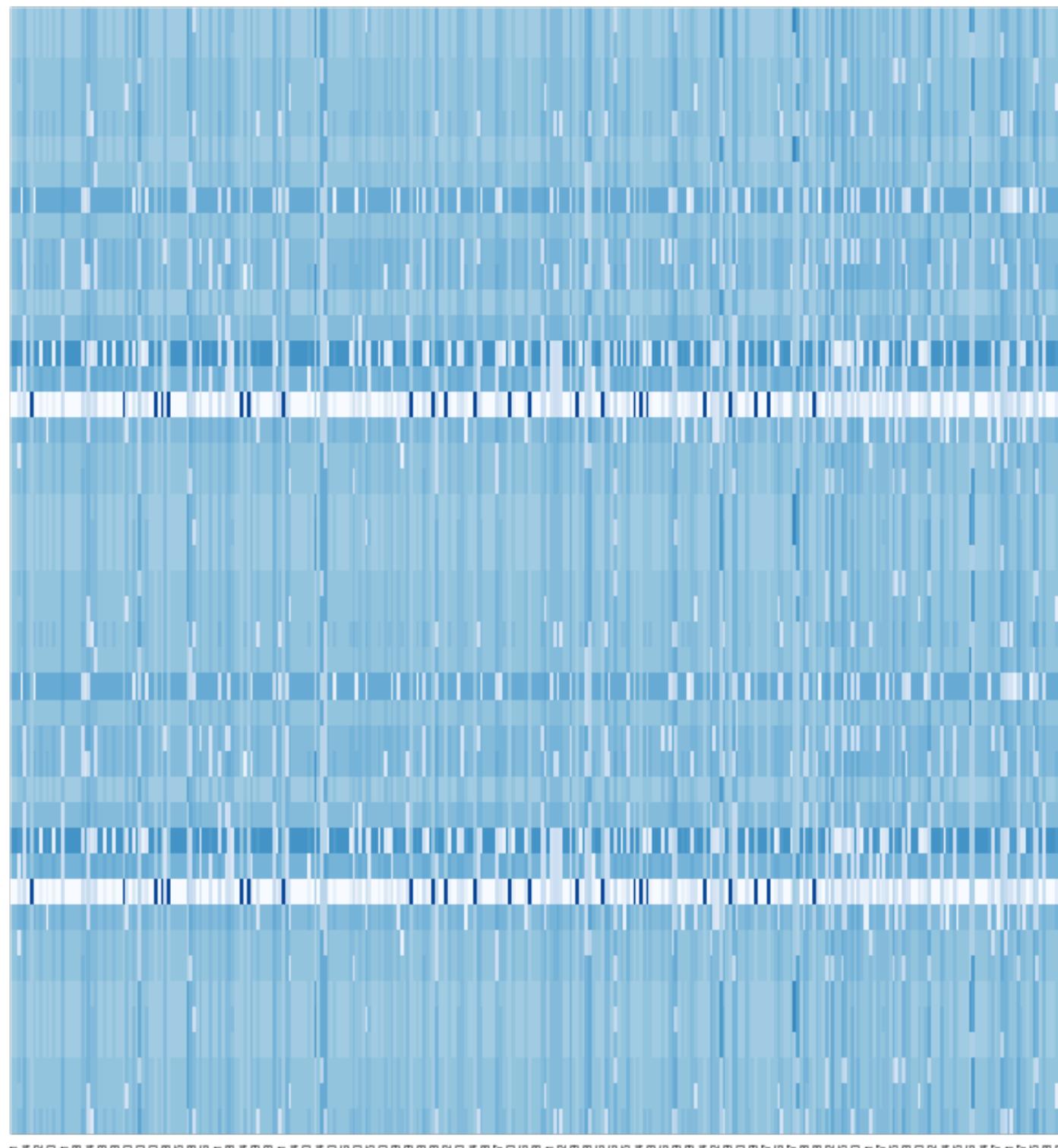
393 **ORCID**

394 Paul CaraDonna <https://orcid.org/0000-0003-3517-9090>
395 Jeremie Fant <https://orcid.org/0000-0001-9276-1111>
396 Jane Ogilvie <https://orcid.org/0000-0001-8546-0417>
397 Sophie Taddeo <https://orcid.org/0000-0002-7789-1417>

398 **References**

399 **Supporting**

Percent matched reads per locus by sample



- 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 (corporate)
- 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 aperture
- 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 an 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

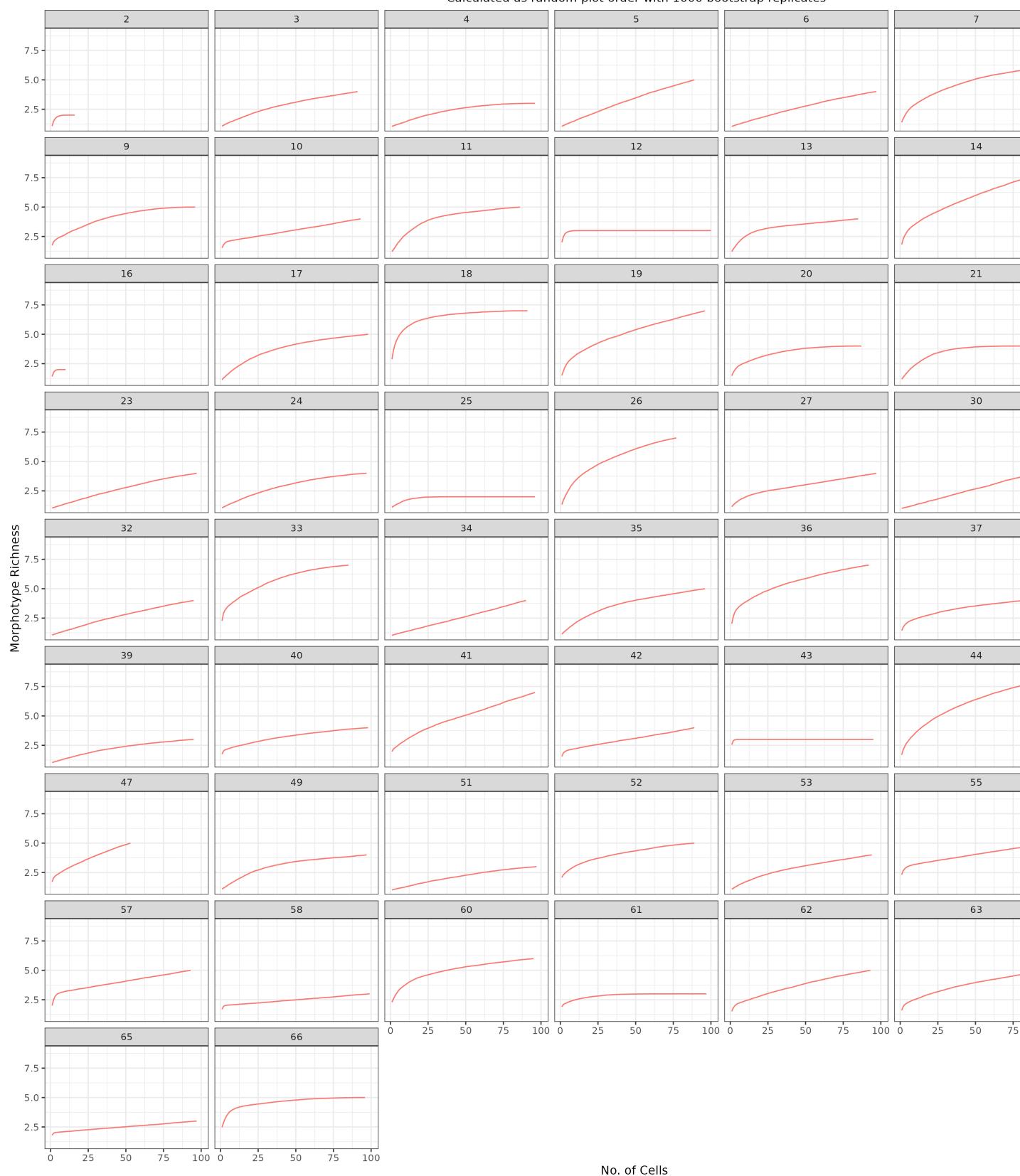
Morphotype B: Lupinus, Lathyrus, Potentilla, Androsace, Bistorta, Vicia

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

Morphotype D: Salix, Boechera

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

Rarefaction Curves of Species Richness
Calculated as random plot order with 1000 bootstrap replicates



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

Species richness
Number of species
Number of individuals

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 Candolle	Asteraceae	CHIC tba	S	Colorado, Gunnison	18.VII.2020	tba	3.6
<i>Erigeron subirinervis</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>Helianthus 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 serra</i> Hook.	Asteraceae	CHIC tba	P	Idaho, Idaho	26.VII.2020	tba	1035.0
<i>Sympphytidium foliacetum</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>Mertensia ciliata</i> (James ex Torr.) G. Don	Boraginaceae	ID 175485	S	Idaho, Valley	18.VI.2018	tba	979.3
<i>Mertensia 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.VII.1997	tba	44.8
<i>Mertensia fusiformis</i> Greene	Boraginaceae	RMH 720600	P	Colorado, Gunnison	9.VII.1997	tba	38.9
<i>Campionula rotundifolia</i> L.	Campanulaceae	CHIC tba	S	Colorado, Gunnison	18.VII.2020	tba	3.6
<i>Lathyrus lauszweitii</i> Kellogg var. <i>leucanthus</i> (Ryd.) Dorn	Fabaceae	CHIC tba	S	Colorado, Gunnison	18.VII.2020	tba	3.6
<i>Lathyrus lauszweitii</i> Kellogg var. <i>leucanthus</i> (Ryd.) Dorn	Fabaceae	CHIC tba	S	Colorado, Gunnison	18.VII.2020	tba	3.6
<i>Lupinus argenteus</i> Pursh	Fabaceae	CHIC tba	S	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	Washington, Yakima	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.VII.2011	tba	65.3
<i>Hydrophyllum fendleri</i> (Gray) Heller	Hydrophyllaceae	ID 161100	P	Washington, Yakima	9.VII.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>Agastache pallidiflora</i> (Heller) Rydberg	Lamiaceae	CHIC tba	S	Arizona, Coconino	17.VII.2020	tba	617.7
<i>Chamerion angustifolium</i> (L.) Holub	Onagraceae	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 166462	P	Idaho, Gem	15.VI.2011	tba	9825.5
<i>Delphinium nuttallianum</i> Pritz.	Ranunculaceae	ID 179376	P	Idaho, Gooding	29.IV.2017	tba	793.7
<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

- 407 Alarcón, R. (2010). Congruence between visitation and pollen-transport networks in a California plant–
408 pollinator community. *Oikos*, **119**, 35–44. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0706.2009.17694.x>
- 410 Baker, W.J., Bailey, P., Barber, V., Barker, A., Bellot, S., Bishop, D., Botigué, L.R., Brewer, G., Carruthers,
411 T., Clarkson, J.J., Cook, J., Cowan, R.S., Dodsworth, S., Epitawalage, N., Françoso, E., Gallego, B.,
412 Johnson, M.G., Kim, J.T., Leempoel, K., Maurin, O., McGinnies, C., Pokorny, L., Roy, S., Stone, M.,
413 Toledo, E., Wickett, N.J., Zuntini, A.R., Eiserhardt, W.L., Kersey, P.J., Leitch, I.J. & Forest, F. (2021a).
414 A Comprehensive Phylogenomic Platform for Exploring the Angiosperm Tree of Life. *Systematic Biology*,
415 **71**, 301–319. Retrieved from <https://doi.org/10.1093/sysbio/syab035>
- 416 Baker, W.J., Dodsworth, S., Forest, F., Graham, S.W., Johnson, M.G., McDonnell, A., Pokorny, L., Tate,
417 J.A., Wicke, S. & Wickett, N.J. (2021b). Exploring Angiosperms353: An open, community toolkit for
418 collaborative phylogenomic research on flowering plants. *American Journal of Botany*, **108**, 1059–1065.
419 Retrieved from <https://bsapubs.onlinelibrary.wiley.com/doi/abs/10.1002/ajb2.1703>
- 420 Barker, D.A. & Arceo-Gomez, G. (2021). Pollen transport networks reveal highly diverse and temporally
421 stable plant–pollinator interactions in an Appalachian floral community. *AoB PLANTS*, **13**. Retrieved
422 from <https://doi.org/10.1093/aobpla/plab062>
- 423 Beattie, A. (1971). A technique for the study of insect-borne pollen. *The Pan-Pacific Entomologist*, **47**, 82.
- 424 Belitz, M.W., Larsen, E.A., Ries, L. & Guralnick, R.P. (2020). The accuracy of phenology estimators for use
425 with sparsely sampled presence-only observations. *Methods in Ecology and Evolution*, **11**, 1273–1285.
- 426 Bergman, P., Molau, U. & Holmgren, B. (1996). Micrometeorological impacts on insect activity and plant
427 reproductive success in an alpine environment, Swedish Lapland. *Arctic and Alpine Research*, **28**, 196–202.
- 428 Bingham, R.A. & Orthner, A.R. (1998). Efficient pollination of alpine plants. *Nature*, **391**, 238–239.
- 429 Bivand, R. & Wong, D.W.S. (2018). Comparing implementations of global and local indicators of spatial
430 association. *TEST*, **27**, 716–748.
- 431 Bolger, A. & Giorgi, F. (2014). Trimmomatic: A flexible read trimming tool for Illumina NGS data.
432 *Bioinformatics*, **30**, 2114–2120.
- 433 Brosi, B.J. & Briggs, H.M. (2013). Single pollinator species losses reduce floral fidelity and plant reproductive
434 function. *Proceedings of the National Academy of Sciences*, **110**, 13044–13048.
- 435 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T.L. (2009).
436 BLAST+: Architecture and applications. *BMC Bioinformatics*, **10**, 1–9.
- 437 Cameron, S.A. & Sadd, B.M. (2020). Global trends in bumble bee health. *Annual Review of Entomology*, **65**,
438 209–232.
- 439 CaraDonna, P.J., Burkle, L.A., Schwarz, B., Resasco, J., Knight, T.M., Benadi, G., Blüthgen, N., Dormann,

- 440 C.F., Fang, Q., Fründ, J. & others. (2021). Seeing through the static: The temporal dimension of
441 plant–animal mutualistic interactions. *Ecology Letters*, **24**, 149–161.
- 442 CaraDonna, P.J., Petry, W.K., Brennan, R.M., Cunningham, J.L., Bronstein, J.L., Waser, N.M. & Sanders,
443 N.J. (2017). Interaction rewiring and the rapid turnover of plant–pollinator networks. *Ecology letters*, **20**,
444 385–394.
- 445 Chao, A., Gotelli, N.J., Hsieh, T.C., Sande, E.L., Ma, K.H., Colwell, R.K. & Ellison, A.M. (2014). Rarefaction
446 and extrapolation with hill numbers: A framework for sampling and estimation in species diversity studies.
447 *Ecological Monographs*, **84**, 45–67.
- 448 Cheng, S., Melkonian, M., Smith, S.A., Brockington, S., Archibald, J.M., Delaux, P.-M., Li, F.-W., Melkonian,
449 B., Mavrodiev, E.V., Sun, W., Fu, Y., Yang, H., Soltis, D.E., Graham, S.W., Soltis, P.S., Liu, X., Xu, X.
450 & Wong, G.K.-S. (2018). 10KP: A phylogenetic genome sequencing plan. *GigaScience*, **7**. Retrieved from
451 <https://doi.org/10.1093/gigascience/giy013>
- 452 Coissac, E., Hollingsworth, P.M., Lavergne, S. & Taberlet, P. (2016). From barcodes to genomes: Extending
453 the concept of DNA barcoding.
- 454 Colla, S.R., Gadallah, F., Richardson, L., Wagner, D. & Gall, L. (2012). Assessing declines of north american
455 bumble bees (*bombus* spp.) Using museum specimens. *Biodiversity and Conservation*, **21**, 3585–3595.
- 456 Doyle, J.J. & Doyle, J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue.
457 *Phytochemical Bulletin*, **19**, 11–15.
- 458 Fazekas, A.J., Kesanakurti, P.R., Burgess, K.S., Percy, D.M., Graham, S.W., Barrett, S.C., Newmaster, S.G.,
459 Hajibabaei, M. & Husband, B.C. (2009). Are plant species inherently harder to discriminate than animal
460 species using DNA barcoding markers? *Molecular Ecology Resources*, **9**, 130–139.
- 461 Frase, Barbara A. & Buck, P. (2007). Vascular Plants of the Gothic Area. Retrieved from https://www.digitallibrarystatic.org/wp-content/uploads/2016/05/vascularplantlist_20071.pdf
- 462 Gage, E. & Cooper, D.J. (2013). Historical range of variation assessment for wetland and riparian ecosystems,
463 u.s. Forest service rocky mountain region
- 464 Goulson, D., Lye, G. & Darvill, B. (2008). The decline and conservation of bumblebees. *Annual review of
465 entomology*, **53**, 191–208.
- 466 Govaerts, R., Nic Lughadha, E., Black, N., Turner, R. & Paton, A. (2021). The world checklist of vascular
467 plants, a continuously updated resource for exploring global plant diversity. *Scientific Data*, **8**, 1–10.
- 468 Group, C.P.W., Hollingsworth, P.M., Forrest, L.L., Spouge, J.L., Hajibabaei, M., Ratnasingham, S., Bank,
469 M. van der, Chase, M.W., Cowan, R.S., Erickson, D.L. & others. (2009). A DNA barcode for land plants.
470 *Proceedings of the National Academy of Sciences*, **106**, 12794–12797.
- 471 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.,

- 473 Chen, S.-L. & others. (2011). Comparative analysis of a large dataset indicates that internal transcribed
474 spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National*
475 *Academy of Sciences*, **108**, 19641–19646.
- 476 Hebert, P.D., Cywinska, A., Ball, S.L. & DeWaard, J.R. (2003). Biological identifications through DNA
477 barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**, 313–321.
- 478 Hengl, T., Mendes de Jesus, J., Heuvelink, G.B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A.,
479 Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B. & others. (2017). SoilGrids250m:
480 Global gridded soil information based on machine learning. *PLoS one*, **12**, e0169748.
- 481 Hennig, C. (2020). *Fpc: Flexible procedures for clustering*. Retrieved from <https://CRAN.R-project.org/package=fpc>
- 482 Hsieh, T.C., Ma, K.H. & Chao, A. (2020). *iNEXT: Interpolation and extrapolation for species diversity*.
483 Retrieved from http://chao.stat.nthu.edu.tw/wordpress/software_download/
- 484 Iler, A.M., Humphrey, P.T., Ogilvie, J.E. & CaraDonna, P.J. (2021). Conceptual and practical issues limit
485 the utility of statistical estimators of phenological events. *Ecosphere*, **12**, e03828.
- 486 Janzen, D.H. (1967). Synchronization of sexual reproduction of trees within the dry season in central america.
487 *Evolution*, **21**, 620–637.
- 488 Janzen, D.H., Burns, J.M., Cong, Q., Hallwachs, W., Dapkey, T., Manjunath, R., Hajibabaei, M., Hebert, P.D.
489 & Grishin, N.V. (2017). Nuclear genomes distinguish cryptic species suggested by their DNA barcodes
490 and ecology. *Proceedings of the National Academy of Sciences*, **114**, 8313–8318.
- 491 Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J. & Wickett,
492 N.J. (2016). HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput
493 sequencing reads using target enrichment. *Applications in plant sciences*, **4**, 1600016.
- 494 Johnson, M.G., Pokorny, L., Dodsworth, S., Botigue, L.R., Cowan, R.S., Devault, A., Eiserhardt, W.L.,
495 Epitawalage, N., Forest, F., Kim, J.T. & others. (2019). A universal probe set for targeted sequencing of
496 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic biology*, **68**,
497 594–606.
- 498 Kress, W.J. & Erickson, D.L. (2007). A two-locus global DNA barcode for land plants: The coding rbcL
499 gene complements the non-coding trnH-psbA spacer region. *PLoS one*, **2**, e508.
- 500 Kuhn, M. (2022). *Caret: Classification and regression training*. Retrieved from <https://CRAN.R-project.org/package=caret>
- 501 Lewin, H.A., Richards, S., Aiden, E.L., Allende, M.L., Archibald, J.M., Bálint, M., Barker, K.B., Baumgartner,
502 Bélov, K., Bertorelle, G., Blaxter, M.L., Cai, J., Caperello, N.D., Carlson, K., Castilla-Rubio, J.C.,
503 Chaw, S.-M., Chen, L., Childers, A.K., Coddington, J.A., Conde, D.A., Corominas, M., Crandall, K.A.,

- 506 Crawford, A.J., DiPalma, F., Durbin, R., Ebenezer, T.E., Edwards, S.V., Fedrigo, O., Flieck, P., Formenti,
507 G., Gibbs, R.A., Gilbert, M.T.P., Goldstein, M.M., Graves, J.M., Greely, H.T., Grigoriev, I.V., Hackett,
508 K.J., Hall, N., Haussler, D., Helgen, K.M., Hogg, C.J., Isobe, S., Jakobsen, K.S., Janke, A., Jarvis, E.D.,
509 Johnson, W.E., Jones, S.J.M., Karlsson, E.K., Kersey, P.J., Kim, J.-H., Kress, W.J., Kuraku, S., Lawniczak,
510 M.K.N., Leebens-Mack, J.H., Li, X., Lindblad-Toh, K., Liu, X., Lopez, J.V., Marques-Bonet, T., Mazard,
511 S., Mazet, J.A.K., Mazzoni, C.J., Myers, E.W., O'Neill, R.J., Paez, S., Park, H., Robinson, G.E., Roquet,
512 C., Ryder, O.A., Sabir, J.S.M., Shaffer, H.B., Shank, T.M., Sherkow, J.S., Soltis, P.S., Tang, B., Tedersoo,
513 L., Uliano-Silva, M., Wang, K., Wei, X., Wetzer, R., Wilson, J.L., Xu, X., Yang, H., Yoder, A.D. & Zhang,
514 G. (2022). The earth BioGenome project 2020: Starting the clock. *Proceedings of the National Academy
of Sciences*, **119**, e2115635118. Retrieved from <https://www.pnas.org/doi/abs/10.1073/pnas.2115635118>
- 515 Life Project Consortium, D.T. of, Blaxter, M., Mieszkowska, N., Palma, F.D., Holland, P., Durbin, R.,
516 Richards, T., Berriman, M., Kersey, P., Hollingsworth, P., Wilson, W., Twyford, A., Gaya, E., Lawniczak,
517 M., Lewis, O., Broad, G., Howe, K., Hart, M., Flieck, P. & Barnes, I. (2022). Sequence locally,
518 think globally: The darwin tree of life project. *Proceedings of the National Academy of Sciences*, **119**,
519 e2115642118. Retrieved from <https://www.pnas.org/doi/abs/10.1073/pnas.2115642118>
- 520 Liu, J., Shi, L., Han, J., Li, G., Lu, H., Hou, J., Zhou, X., Meng, F. & Downie, S.R. (2014). Identification
521 of species in the angiosperm family apiaceae using DNA barcodes. *Molecular ecology resources*, **14**,
522 1231–1238.
- 523 Li, X., Yang, Y., Henry, R.J., Rossetto, M., Wang, Y. & Chen, S. (2015). Plant DNA barcoding: From gene
524 to genome. *Biological Reviews*, **90**, 157–166.
- 525 Lu, J., Breitwieser, F.P., Thielen, P. & Salzberg, S.L. (2017). Bracken: Estimating species abundance in
526 metagenomics data. *PeerJ Computer Science*, **3**, e104.
- 527 Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. (2022). *Cluster: Cluster analysis basics
and extensions*. Retrieved from <https://CRAN.R-project.org/package=cluster>
- 528 Maitner, B. (2022). BIEN: Tools for accessing the botanical information and ecology network database.
529 Retrieved from <https://CRAN.R-project.org/package=BIEN>
- 530 McLay, T.G., Birch, J.L., Gunn, B.F., Ning, W., Tate, J.A., Nauheimer, L., Joyce, E.M., Simpson, L.,
531 Schmidt-Lebuhn, A.N., Baker, W.J. & others. (2021). New targets acquired: Improving locus recovery
532 from the Angiosperms353 probe set. *Applications in plant sciences*, **9**.
- 533 Moran, P.A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, **37**, 17–23.
- 534 Naimi, B. & Araujo, M.B. (2016). Sdm: A reproducible and extensible r platform for species distribution
535 modelling. *Ecography*, **39**, 368–375.
- 536 Naimi, B., Hamm, N. a.s., Groen, T.A., Skidmore, A.K. & Toxopeus, A.G. (2014). Where is positional
537

- 539 uncertainty a problem for species distribution modelling. *Ecography*, **37**, 191–203.
- 540 Newstrom, L.E., Frankie, G.W. & Baker, H.G. (1994). A new classification for plant phenology based on
541 flowering patterns in lowland tropical rain forest trees at la selva, costa rica. *Biotropica*, **26**, 141–159.
- 542 Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Solymos,
543 P., Stevens, M.H.H., Szoeecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho,
544 G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H.B.A., FitzJohn, R., Friendly, M., Furneaux,
545 B., Hannigan, G., Hill, M.O., Lahti, L., McGlinn, D., Ouellette, M.-H., Ribeiro Cunha, E., Smith, T.,
546 Stier, A., Ter Braak, C.J.F. & Weedon, J. (2022). *Vegan: Community ecology package*. Retrieved from
547 <https://CRAN.R-project.org/package=vegan>
- 548 Oliver, P.M., Adams, M., Lee, M.S., Hutchinson, M.N. & Doughty, P. (2009). Cryptic diversity in vertebrates:
549 Molecular data double estimates of species diversity in a radiation of australian lizards (*diplodactylus*,
550 *gekkota*). *Proceedings of the Royal Society B: Biological Sciences*, **276**, 2001–2007.
- 551 Omernik, J.M. (1987). Ecoregions of the conterminous united states. *Annals of the Association of American
552 geographers*, **77**, 118–125.
- 553 Pearse, W.D., Davis, C.C., Inouye, D.W., Primack, R.B. & Davies, T.J. (2017). A statistical estimator for
554 determining the limits of contemporary and historic phenology. *Nature Ecology & Evolution*, **1**, 1876–1882.
- 555 Prim, R.C. (1957). Shortest connection networks and some generalisations. *Bell System Technical Journal*,
556 **36**, 1389–1401.
- 557 Robinson, N., Regetz, J. & Guralnick, R.P. (2014). EarthEnv-DEM90: A nearly-global, void-free, multi-
558 scale smoothed, 90m digital elevation model from fused ASTER and SRTM data. *ISPRS Journal of
559 Photogrammetry and Remote Sensing*, **87**, 57–67.
- 560 Sarro, E., Tripodi, A. & Woodard, S.H. (2022). Bumble bee (*bombus vosnesenskii*) queen nest searching
561 occurs independent of ovary developmental status. *Integrative Organismal Biology*, **4**, obac007.
- 562 Tange, O. (2021). GNU parallel 20220322 (savannah). Retrieved from <https://doi.org/10.5281/zenodo.6377950>
- 563 Tran, H., Nguyen, P., Ombadi, M., Hsu, K., Sorooshian, S. & Qing, X. (2019). A cloud-free MODIS snow
564 cover dataset for the contiguous united states from 2000 to 2017. *Scientific data*, **6**, 1–13.
- 565 Wang, T., Hamann, A., Spittlehouse, D. & Carroll, C. (2016). Locally downscaled and spatially customizable
566 climate data for historical and future periods for north america. *PloS one*, **11**, e0156720.
- 567 Wenzell, K.E., McDonnell, A.J., Wickett, N.J., Fant, J.B. & Skogen, K.A. (2021). Incomplete reproductive
568 isolation and low genetic differentiation despite floral divergence across varying geographic scales in
569 *castilleja*. *American Journal of Botany*, **108**, 1270–1288.
- 570 Williams, P.H. (1982). The distribution and decline of british bumble bees (*bombus latr.*). *Journal of
571 Apicultural Research*, **21**, 236–245. Retrieved from <https://doi.org/10.1080/00218839.1982.11100549>

- 572 Wilson, A.M. & Jetz, W. (2016). Remotely sensed high-resolution global cloud dynamics for predicting
573 ecosystem and biodiversity distributions. *PLoS biology*, **14**, e1002415.
- 574 Wood, D.E., Lu, J. & Langmead, B. (2019). Improved metagenomic analysis with kraken 2. *Genome biology*,
575 **20**, 1–13.
- 576 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,
577 D.-Z. & Wang, H. (2019). The topological differences between visitation and pollen transport networks:
578 A comparison in species rich communities of the himalaya–hengduan mountains. *Oikos*, **128**, 551–562.
579 Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1111/oik.05262>

⁵⁸⁰ **List of Figures**

⁵⁸¹	1	A caption	30
----------------	---	---------------------	----

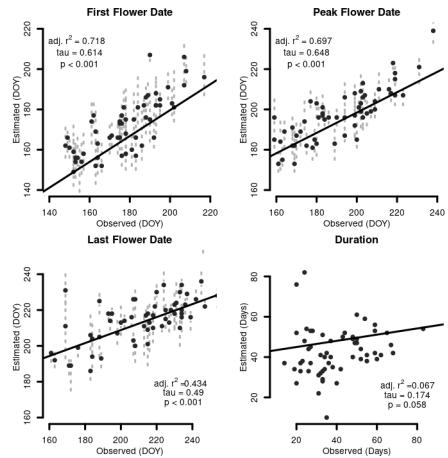


Figure 1: A caption

582 **List of Tables**

583 2	Logistic regression assessing accuracy of SDMs	32
584 3	SDM evaluation contingency table	33

Table 2: 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 3: SDM evaluation contingency table

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