

<sup>1</sup> Plant Metagenomic Barcoding using Angiosperms353 of Corbiculae  
<sup>2</sup> from Queen Bumble Bees

<sup>3</sup> Reed Clark Benkendorf<sup>1,2\*</sup>, Jane E. Ogilive<sup>3</sup>, Emily J. Woodworth<sup>1,2</sup>,  
Sophie Taddeo<sup>1,2</sup>, Paul J. CaraDonna<sup>1,2,3</sup>, Jeremie B. Fant<sup>1,2</sup>

<sup>1</sup>Chicago Botanic Garden, 1000 Lake Cook Road, Glencoe, Illinois 60022, USA

<sup>2</sup>Plant Biology and Conservation, Northwestern University, Evanston, Illinois 60208, USA

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

<sup>4</sup> **Abstract**

- <sup>5</sup> 1) DNA Barcoding has been successful for the rapid analysis of complex ecological assemblages.
- <sup>6</sup> Barcoding in the plant kingdom barcoding has been more difficult than in others, and hence may
- <sup>7</sup> begin to lag behind other kingdoms.
- <sup>8</sup> 2) Here we test the utilization of Angiosperms 353 probes to barcode plant species found in corbiculae
- <sup>9</sup> pollen loads collected from Queen Bumble Bees.
- <sup>10</sup> 3) To verify the accuracy of this barcoding system we compared the barcoding data to museum plant
- <sup>11</sup> specimens, observations of plant flowering and bee flower visitation in the field, and species dis-
- <sup>12</sup> tribution modelling to identify likely candidate plant species found within the corbiculae pollen
- <sup>13</sup> loads.
- <sup>14</sup> 4) By utilizing Species distribution modelling we allow users to create a regionally appropriate sequence
- <sup>15</sup> databases which may use increase the alignment algorithms minimizing need for large computational
- <sup>16</sup> power, and run time.
- <sup>17</sup> 5) We show that the Angiosperms 353 probes, developed for phylogenomics, and which are currently
- <sup>18</sup> being used in the largest ever plant systematic endeavor, offer significant promise to metagenomic
- <sup>19</sup> approaches.
- <sup>20</sup> 6) The DNA barcoding of bumble bee corbiculae pollen loads was most accurate when combined with
- <sup>21</sup> knowledge of what plant species were flowering in the plant community when they were collected.
- <sup>22</sup> Thus, supplementing DNA barcoding data with ecological context is most accurate and powerful.

---

\*Correspondence: reedbenkendorf2021@u.northwestern.edu

## <sup>23</sup> 1 | INTRODUCTION

<sup>24</sup> The inability to reliably identify plants to terminal taxon can limit our understanding of ecosystem function  
<sup>25</sup> and interactions (Bortolus (2008)). This is especially true for genera where identification down to species  
<sup>26</sup> can be useful as specific bioindicators; defining ecological and behavioral properties (e.g. different species  
<sup>27</sup> of Sagebrush- *Artemisia* L., Willows - *Salix* L., and Sedges - *Carex* L.) (Gage & Cooper (2013), Baner-  
<sup>28</sup> jee *et al.* (2022)). In these instances the lack of species level data can hinder our understanding of the  
<sup>29</sup> breadth of habitat which some species occupy, and their interactions with other species. This can be further  
<sup>30</sup> complicated by the fact that the identification of organisms to species is also often mired by lack of diagnos-  
<sup>31</sup> tic characters(e.g. flowers, fruits, roots or combinations thereof), an increasing lack of taxonomic expertise  
<sup>32</sup> (Hebert *et al.* (2003), Banerjee *et al.* (2022)) and the presence of cryptic species (Janzen *et al.* (2017),  
<sup>33</sup> Oliver *et al.* (2009)). Taxonomic verification can also be limited by the fact that revisiting field sites to  
<sup>34</sup> identify material using morphological or chemical approaches, can be resource intensive and often ineffective.  
<sup>35</sup> The current methods to ameliorate this situation include: (i) ignoring these ecologically relevant levels of  
<sup>36</sup> detail, (ii) revisiting plots as diagnostic material becomes temporally available, (iii) seeking the assistance  
<sup>37</sup> from taxonomic specialists, or (iv) the use of barcoding using molecular techniques (Hebert *et al.* (2003),  
<sup>38</sup> Banerjee *et al.* (2022)).

<sup>39</sup> Recently barcoding (the identification of a sample from a single organism *e.g.* a piece of leaf), and metabar-  
<sup>40</sup> coding (the identification of a sample containing a mix of organisms *e.g.* soil), have shown considerable  
<sup>41</sup> promise in all Kingdoms of Life (Ruppert *et al.* (2019)). With plants the identification of members of cer-  
<sup>42</sup> tain clades using barcoding has been quite successful (Kress (2017)), whereas with many other clades results  
<sup>43</sup> have been elusive (Liu *et al.* (2014), Group *et al.* (2011), Coissac *et al.* (2012)), however metabarcoding  
<sup>44</sup> incurs additional challenges to those which exist for the currently implemented barcodes (Li *et al.* (2015),  
<sup>45</sup> Kress & Erickson (2007), Group *et al.* (2009), Coissac *et al.* (2012)). Particular challenges with the high  
<sup>46</sup> copy number barcodes (*e.g.* ITS2, *rbcL*, *matK*, *trnH-psbA*) include the utilization their rates of divergence,  
<sup>47</sup> gene tree conflict, and hybridization (Coissac *et al.* (2016), Fazekas *et al.* (2009)). Currently, most plant  
<sup>48</sup> metabarcoding endeavors only allow the identification of material to the level of family or genus.

<sup>49</sup> [Table 1 about here.]

<sup>50</sup> Currently the largest plant systematic endeavor ever undertaken,by the Royal Botanic Gardens Kew, the  
<sup>51</sup> Plant and Fungal Tree of Life (PAFTOL) is approaching completion (Baker *et al.* (2021a)). This data  
<sup>52</sup> set will contain hybridization capture (Hyb-Seq) data from at least one species in each genus of the plant

53 kingdom, 14,000 represented species, using the popular Angiosperms353 (A353) probes, which includes 353  
54 single-copy orthologous loci, (Baker *et al.* (2021a), Johnson *et al.* (2019)). These publicly available data  
55 serve to provide a taxonomically comprehensive backbone for plant metabarcoding. The A353 probes are  
56 currently being used in many other plant phylogenetic studies increasing the sampling depth of many clades  
57 (Baker *et al.* (2021b)). Data from the 10kP project, which seeks to develop reference genomes from a  
58 phylogenetically diverse suite of plants, will contribute many more species by 2030 (Cheng *et al.* (2018)).  
59 Similar projects such as the ‘Darwin Tree of Life’ which will sequence all described taxa in Britain and  
60 Ireland, seek to sequence high numbers of genomes in geographic regions will contribute data sets applicable  
61 to enormous spatial domains (Life Project Consortium *et al.* (2022), Lewin *et al.* (2022)). These data will  
62 promote the ability to apply metabarcoding to resolve a diverse array of questions relevant to theoretical  
63 and applied ecology (Kress (2017), Hollingsworth *et al.* (2016)). However, the application of metabarcoding  
64 still faces challenges relating to the enormity of the genomic data sets and the computational power required  
65 to process sequence data.

66 Herein we have resolved major components of the problems of accurately and effectively identifying plant  
67 material without diagnostic morphological character states using the A353 Hyb-Seq probes (Johnson *et*  
68 *al.* (2019)), and custom species sequence databases derived via species distribution modelling, and temporal  
69 filtering. To increase the accuracy and efficiency of metabarcoding results in plants, we are proposing reducing  
70 the number of possible candidate species by generating a user specific databases relevant to the region of study  
71 and the ecological characteristics of interest (Bell *et al.* (2022)). To achieve this goal, we first create a regional  
72 list of candidate species using digital collections gleaned from herbaria, survey work, and citizen science  
73 (e.g. iNaturalist), from a region exceeding the study area. For these candidate species, a modelling approach,  
74 such as logistic regression, may be used to identify taxa which warrant further exploration e.g. determine  
75 their possibility of presence in metabarcoding samples. We then use species distribution models to create  
76 potential distribution maps for the candidate species to limit the impact of spatial and taxonomic biases in  
77 the species list and account for spatial variations in niche availability throughout the study area. Species  
78 distribution models examine the ecological conditions associated with the known occurrence of a species to  
79 identify suitable habitats in the study area. This approach has the benefit of greatly reducing the size of  
80 a sequence database, which allows for the use of genomic data on personal computers. This approach can  
81 significantly reduce processing time, increasing the a projects efficiency, particularly as most next-generation  
82 sequence data is deposited as raw-sequence reads.

83

[Table 2 about here.]

84 As species interactions vary both in space and time (Thompson (1994), CaraDonna *et al.* (2021)). Contrasts  
85 in the flowering periods of many plant species, can provide an additional filter for increasing the accuracy  
86 of identifying material in certain types of metagenomic samples (Janzen (1967), Newstrom *et al.* (1994)).  
87 In high elevation temperate regions, pollination interactions vary temporally and are characterized by high  
88 turnover in active periods of species (CaraDonna *et al.* (2017)), however the overall shorter extent of the  
89 active growing season in these systems results in the presence of few to any natural breaks, which reduces  
90 the utility of these to operate as filters in the post-processing of sequence matches. Nonetheless, we work  
91 develop a general approach which seems applicable to many areas which utilize the temporal dimension for  
92 classifying sequences in metagenomic samples (but see Davis *et al.* (2022), Figure 1).

93 [Figure 1 about here.]

94 We tested this metagenomic and informatics approach to determine to examine the foraging of queen bum-  
95 blebee, and compare this approach with direct observations and the pollen record, which has shown some  
96 incongruity in several floral visitation networks involving smaller bodied fauna (Barker & Arceo-Gomez  
97 (2021), Zhao *et al.* (2019), Alarcón (2010)). The assessment of the plant species compositions from pollen is  
98 a desirable approach, with several potential applications, despite numerous potential complications (Pornon  
99 *et al.* (2017), Bell *et al.* (2017), Sickel *et al.* (2015), Bell *et al.* (2019), Suchan *et al.* (2019), Johnson *et*  
100 *al.* (2021)). The two foraging phases of the queen bumblebee life cycle are essential to 1) increase their  
101 weight before diapause, and 2) increase their ovary weights while establishing their recently found nests.  
102 Both of these time periods represent potential demographic bottlenecks in bumble bee populations (Sarro  
103 *et al.* (2022)). Bumblebees are one of the only groups of insects with unequivocal quantitative evidence for  
104 numerous population declines, while simultaneously serving as the most effective pollinators in many tem-  
105 perate montane ecosystems (Cameron & Sadd (2020), Goulson *et al.* (2008), Williams (1982), Colla *et al.*  
106 (2012), Bergman *et al.* (1996), Bingham & Orthner (1998), Grixti *et al.* (2009)). These areas often represent  
107 the most diverse areas in the temperate and often offer the sole potential refugia for multiple dimensions of  
108 biodiversity under climate change, whilst simultaneously experiencing the greatest proportional changes in  
109 mean annual temperature (Brito-Morales *et al.* (2018), Pepin *et al.* (2022)). An immediate understanding  
110 of how to manage previously overlooked keystone insect species, such as bumble bees, is essential if the  
111 refugial potential of the temperate mountains are to be incorporated into their current diversity (Loarie *et*  
112 *al.* (2009), Dobrowski & Parks (2016)).

<sup>113</sup> **2 | METHODS**

<sup>114</sup> **2.1 Case Study: Bee-Flower Observations and Pollen Load Collection**

<sup>115</sup> Bee and flower observations and bee corbiculae pollen collection was conducted around the Rocky Mountain  
<sup>116</sup> Biological Laboratory (RMBL; 38°57.5" N, 106°59.3" W (WGS 84), 2900 m.a.s.l.), Colorado, USA (Appendix  
<sup>117</sup> 1 for site information). The area is characterized by subalpine meadow vegetation communities. Pollinator  
<sup>118</sup> observations of *Bombus* Latreille spp. (Apidae Latreille) were conducted from May 29th – July 23rd of  
<sup>119</sup> 2015 in six study sites as a part of a larger study (described in Ogilvie and CaraDonna 2022). Observations  
<sup>120</sup> of *Bombus* foraging took place for one hour at each field site, with equal time spent searching for bee in  
<sup>121</sup> the major vegetation types (dry, and wet meadows, and aspen forest). Corbiculae pollens loads were, non-  
<sup>122</sup> lethally, collected from queens encountered by capturing them in an insect net and transferring them into a  
<sup>123</sup> restraining device (“bee squeezer”, Kearns *et al.* (2001)). We then collected a single pollen load (i.e., from  
<sup>124</sup> one leg) from the bee and then released it. At weekly intervals at each site, we also recorded the abundances  
<sup>125</sup> of flowers visited by bumble bees within belt transects spread over the three vegetation types (0.5 x 40 m  
<sup>126</sup> transects in each vegetation type, 60 m<sup>2</sup> total area per site).

<sup>127</sup> **2.3 | Pollen Morphological identification**

<sup>128</sup> **2.3.1 | Pollen Reference Library**

<sup>129</sup> To develop a reference library of pollen grains which may be present in corbiculae loads, an image reference  
<sup>130</sup> collection of fuchsin-jelly stained (Beattie (1971)) slides was assembled from slides previously prepared by the  
<sup>131</sup> authors (n = 21), and other researchers (n = 38) (Brosi & Briggs (2013)). Using five years of observational  
<sup>132</sup> data on *Bombus* Queen Bee foraging at these studies sites (Ogilvie unpublished), as well as the Vascular  
<sup>133</sup> Plant Checklist (Frase & Buck (2007)), an additional 62 voucher slides for species were prepared and imaged  
<sup>134</sup> at 400x (Leica DMLB, Leica MC170 HD Camera, Leica Application Suite V. 4.13.0) from non-accessioned  
<sup>135</sup> herbarium collections to supplement the number of species and clades covered (Appendix 3).

<sup>136</sup> We used Divisive Hierarchical Clustering techniques to determine which plant taxa were distinguishable via  
<sup>137</sup> light microscopy, and to develop a dichotomous key to pollen morphotypes. Ten readily discernible categorical  
<sup>138</sup> traits were collected from each specimen in the image collection. These traits were transformed using Gower  
<sup>139</sup> distances, and clustered using Divisive Hierarchical clustering techniques (Maechler *et al.* (2022)). Using  
<sup>140</sup> the cluster dendrogram, elbow plot, and heatmaps (Hennig (2020)), of these results morphological groups  
<sup>141</sup> of pollen which could not be resolved via microscopy were delineated, and a dichotomous key was prepared

<sup>142</sup> (Appendix 6). This key was then used to identify the pollen grains sampled from corbiculae loads to  
<sup>143</sup> morphotypes in a consistent manner.

<sup>144</sup> **2.3.2 | Pollen Corbiculae Loads**

<sup>145</sup> To prepare the pollen slides from corbiculae, all corbiculae loads were broken apart and rolled using dissection  
<sup>146</sup> needlepoints to increase heterogeneity of samples. *Cerca* 0.5mm<sup>2</sup> of pollen was placed onto a ~4mm<sup>2</sup> fuchsin  
<sup>147</sup> jelly cube (Beattie (1971)) atop a graticulated microscope slide, with 20 transects and 20 rows (400 quadrants)  
<sup>148</sup> (EMS, Hartfield, PA). The jelly was melted, with stirring, until pollen grains were homogeneously spread  
<sup>149</sup> across the microscope slide. Slides were sealed with Canada Balsam (Rublev Colours, Willits, CA) followed  
<sup>150</sup> by sealing with clear nail polish to prevent oxidation; all samples are noted in Appendix 4. To identify the  
<sup>151</sup> pollen present in corbiculae loads, light microscopy at 400x (Zeiss Axioscope A1) was used. In initial sampling  
<sup>152</sup> in three transects, each pollen grain was identified to morphotype and counted; an additional two transects  
<sup>153</sup> were scanned for morphotypes unique to that slide, if either transect contained a unique morphotype than  
<sup>154</sup> all grains in that transect were also identified and counted. Subsequent to the first round of sampling, non-  
<sup>155</sup> parametric species richness rarefaction curves (Oksanen *et al.* (2022)), and non-parametric species diversity  
<sup>156</sup> rarefaction curves were used to assess the completeness of sampling (Chao *et al.* (2014), Hsieh *et al.* (2020)).  
<sup>157</sup> Slides not approaching the asymptote of the rarefaction curve were then re-sampled, and analysed iteratively  
<sup>158</sup> for up to a total of seven transects (Appendix 7 & 8).

<sup>159</sup> **2.4 | Molecular Barcoding**

<sup>160</sup> **2.4.1 | Species reference list**

<sup>161</sup> **2.4.1.1 Spatial Analyses** We first generated a short list of potential candidate species. We downloaded  
<sup>162</sup> from the Botanical Information and Ecology Network ‘BIEN’ (Maitner (2022)) all records adjacent to the  
<sup>163</sup> field sites to develop an ecologically relevant list of vascular plant species, with expected biotic pollination,  
<sup>164</sup> which may be present at the study area. To reduce the list of species to include in the genomic sequence  
<sup>165</sup> databases, we then generated Species Distribution Models (SDMs) for these taxa to predict their distribution  
<sup>166</sup> throughout the study area.

<sup>167</sup> To minimize the number of species for which SDM’s were to be generated, BIEN was queried at a distance  
<sup>168</sup> of up to 100km from our study area and all plant species records were downloaded. To account for the  
<sup>169</sup> stochasticity of botanical collecting and offset the number of records associated with the research station,

170 this data set was bootstrap re-sampled 250 times, with 90% of samples selected, to create a testing data  
171 set. The median of the logistic regression assessing the probability of occurrence of a species record as a  
172 function of distance from the study area was used as a threshold distance, under which, to include species  
173 as candidates for distribution modelling.

174 **2.4.1.2 Distribution Modelling** To determine which clades to include in the reference sequence database  
175 we used Species Distribution Modelling. We used all occurrence records from BIEN ( $n = 23,919$ ) within a  
176 50km border of the ecoregion, Omernik level 3, which includes the study area (*No. 21 “Southern Rockies”*)  
177 to construct the species distribution model (Omernik (1987)). These records were copied into two, initially  
178 identical, sets, one for generating machine learning models (ML; Random Forest, and Boosted Regression  
179 Tree’s), and the other for Generalised Linear (GLM) and Generalized Additive Models (GAM) (Barbet-  
180 Massin *et al.* (2012)). Ensembled predictions have been shown to outperform their constituent models,  
181 on average, and to reduce the ecological signal to the analytical noise of individual runs (Araujo & New  
182 (2007)). No single method of producing SDMs has been shown to universally outperform others when faced  
183 with a large and diverse number of applications, in our case a great number of species with different biology  
184 and ecology (Elith\* *et al.* (2006), Qiao *et al.* (2015)). In the spirit of these findings, multiple families of  
185 models, which can be generated together as they have similar requirements regarding the number and ratios  
186 of Presence to Absence records were ensembled together (Barbet-Massin *et al.* (2012)).

187 We then generated 4,029 absence points, locations where the focal taxon is anticipated missing, through a  
188 random stratification of 19% of the land cover in the area and included them in (Land Management (2019)).  
189 To achieve a larger absence data set, we generated 1,000 pseudo-absence records for each taxon by randomly  
190 selecting coordinates located at least 10km away from an occurrence record. For ML models, these pseudo-  
191 absences were reduced so that the ratio of presence to absence records were balanced (Barbet-Massin *et al.*  
192 (2012)). To achieve this, we removed absence records inside of 10% of the mean sample value of any predictor  
193 variable the presence records; the required number of absence records were then randomly sampled.

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

201 (2014)).

202 Modelling: Random Forest and Boosted Regression Trees, were sub sampled with 30% test and two replicates  
203 each before weighted ensemble based on True Skill Statistics (tss) (Naimi & Araujo (2016)). Generalised  
204 linear models (GLM) and Generalised additive models (GAM) with 30% sub sampling and three replicates  
205 each were also ensembled using the tss (Naimi & Araujo (2016), @). TSS was chosen as the ensemble  
206 criterion as it has been shown to work across a wide range of species occurrences prevalence (Allouche *et*  
207 *al.* (2006)). The results of these models were extracted on a cell-by-cell basis to a polygon feature derived  
208 from a minimum-spanning tree which encompasses the study sites, and species from either ensemble with  
209 greater than 50% mean habitat suitability across all cells were considered present for further purposes (Prim  
210 (1957)).

211 A total of 535 species were modelled using Generalized Linear Models and Generalized Additive Models and  
212 534 species were modelled using Random Forest and Boosted Regression Trees. To evaluate the accuracy of  
213 the species distribution models, additional presence records from GBIF ( $n = 61,789$ ), and AIM ( $n = 12,730$ )  
214 were used as test and training sets ( $n = 74,519$ ) for logistic regression (Ocdownload Gbif.Org (2021), Land  
215 Management (2019)). Additional novel absence records were generated from the AIM data set to create a  
216 data set where each species has balanced presence and absences. Eleven or more paired presence and absence  
217 records were required for this testing, resulting in 334 species being included in the logistic regression (Mdn  
218 = 110.0,  $\bar{x} = 223.1$ , max = 1568 record pairs used) with a 70% test split (Kuhn (2022)).

## 219 2.4.2 | Temporal Analyses

220 For assignment of reads to ecologically probabilistic species subsequent to BLAST, flowering time was used  
221 as a filter. To estimate the duration of dates in which plant species were flowering Weibull estimates of  
222 several phenological parameters all spatially modelled taxa were developed (Belitz *et al.* (2020), Pearse *et*  
223 *al.* (2017)). Only BIEN records which occurred in the Omernik Level 4 Ecoregions within 15km of the  
224 study area ( $n = 5$  Level 4 Ecoregions, or conditionally 6 ecoregions if enough records were not found in  
225 the nearest 5), and which were from herbarium records were included. To remove temporally irrelevant  
226 herbarium records, i.e. material collected during times which flowering is impossible at the study area due  
227 to snow cover, we used the SnowUS data set (Iler *et al.* (2021), Tran *et al.* (2019)) from 2000-2017 were  
228 analyzed for the first three days of contiguous snow absence, and the first three days of contiguous snow cover  
229 in fall. Herbarium records after the 3<sup>rd</sup> quantile for melt, and the 1<sup>st</sup> quantile for snow cover of these metrics  
230 were removed. Species with  $> 10$  records had their Weibull distributions generated for the date when 10%  
231 of individuals had begun flowering, when 50% were flowering, and when 90% of individuals had flowered,

232 we used the initiation and cessation dates, respectively, as effective start and ends of flowering. These  
233 estimates were compared to a long-term observational study of flowering phenology 1974-2012 (CaraDonna  
234 *et al.* (2014)), and the floral abundance data from 2015, using Kendall's tau.

235 **2.5.2 | Barcode references library**

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

238 **2.5.2.1 | Sampling Species for Barcoding** Using five years (2015-2020) of observational data on *Bombus*  
239 queen interactions with flowering plants at these studies sites, we identified the plant taxa most frequently  
240 visited by queens across all years. We sequenced the 12 most visited taxa twice using samples collected from  
241 one site within the Gunnison Basin River Drainage and one individual collected from another more distal  
242 population. In addition we included a congener - or a species from a closely related genus to serve as an  
243 outgroup for all 12 taxa. We also sequenced another 15 taxa commonly visited by *Bombus* workers, based  
244 on the abundances, and immediate access to plant tissue, in the aforementioned data set (Appendix 3).  
245 Plant collections were identified typically using a combination, of dichotomous keys and primary literature  
246 as required (Flora of North America Editorial Committee (1993+), Hitchcock & Cronquist (2018), Ackerfield  
247 (2015), Lesica *et al.* (2012), Cronquist *et al.* (1977+), Allred & Ivey (2012), *Jepson flora project* (2020),  
248 Mohlenbrock (2002)).

249 **2.5.2.2 | Plant Genomic DNA Extraction** Plant genomic DNA was isolated from ~ 1 cm<sup>2</sup> of leaf tissue  
250 from silica-gel dried or herbarium material using a modified cetyltrimethylammonium (CTAB) protocol  
251 (Doyle & Doyle (1987)) that included two chloroform washes. DNA was quantified using a Nanodrop 2000  
252 (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and Qubit fluorometer (Thermo Fisher Scientific).

253 **2.5.2.3 | Pollen Genomic DNA Extraction** Pollen genomic DNA was extracted from corbiculae using  
254 a CTAB based protocol modified from Lahlamgiah et al. and Guertler et al. (2014, 2014). A SDS extraction  
255 buffer (350µL , 100mM Tris-HCl, 50 mM EDTA, 50 mM NaCl, 10% SDS v/v., pH 7.5) was added followed by  
256 vortexing to allow dissolution of corbiculae. Pollen grains were then macerated with Kontes Pellet Pestles,  
257 and the tip of these washed with 130 µL of the SDS extraction buffer, samples were then incubated for  
258 1 hour at 30°C. This was followed by the addition of 10% CTAB solution (450ul, of 20 mM Tris-Cl pH.  
259 8.0, 1.4 M NaCl, 10 mM EDTA pH 7.5, 10% CTAB, 5% PVP, ~85% Deionized water) and RNase (10 uL

260 of 10 mg/mL) and samples were incubated for 40 minutes at 37°C, on a heat block (Multi-Blok, Thermo  
261 Fisher Scientific, Waltham Massachusetts) set to 40°C. After 20 minutes incubation, Proteinase K (15 µL of  
262 20mg/ml) and DTT (12.5 µL of 1M in water) were added, and the samples were further incubated at 60°C  
263 for 1 hour. Samples were then incubated overnight at 40°C. 500 µL of Phenol-Chloroform-Isoamyl alcohol  
264 (25:24:1) were added, vortexed, and centrifuged at 10,000 rpm for 10 minutes and the aqueous phase was  
265 pipetted to a 1.5 ml centrifuge tube.

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

277 **2.5.2.4 | Fragmentation, Library Preparation & Target Enrichment** Library preparation was  
278 performed using the NEBNext Ultra II FS-DNA Library Prep Kit for Illumina (New England BioLabs,  
279 Ipswich, Massachusetts, USA) using slightly modified manufacturers recommendation. Fragmentation was  
280 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  
281 DNA. Adapter Ligation and PCR enrichment were performed with  $\frac{1}{2}$  volumes, while cleanup of products  
282 was performed using SPRI beads (Beckman Coulter, Indianapolis, Indiana, USA) and recommended volumes  
283 of 80% v./v. ethanol washes. The exception was the herbarium specimens which were not fragmented and  
284 only end repaired, with similar library preparation of all samples. Products were analysed on 4% agarose  
285 gels, and a Qubit fluorometer. Libraries were pooled and enriched with the Angiosperms 353 probe kit V.4  
286 (Arbor Biosciences myBaits Target Sequence Capture Kit) by following the manufacturer's protocol and  
287 Brewer et al. 2019. Sequencing was performed using an Illumina mi-Seq with 150-bp end reads, (NUSeq  
288 Core, Chicago, Illinois).

289 **2.6 | Computational Processes and Analyses.**

290 **2.6.1 | Reference Library Data Processing** Sequences were processed using Trimmomatic, which  
291 removed sequence adapters, clipped the first 3 bp, discarding reads less than 36 bp, and removing reads  
292 if their average PHRED score dropped beneath 20 over a window of 5 bp (Bolger & Giorgi (2014), Tange  
293 (2021)). Contigs generated were mapped to a reference with HybPiper with using target files created by  
294 M353 (Johnson *et al.* (2016), McLay *et al.* (2021)).

295 **2.6.2 | Sequence Identification** A custom Kraken2 database was created by downloading representative  
296 species indicated as being present in the study area by the spatial analyses from the Sequence Read Archive  
297 (SRA) NCBI (Wood *et al.* (2019)). These sequences were processed in the same manner as our novel  
298 sequences. The Kraken2 database was built using default parameters. Kraken2 was run on sequences using  
299 default parameters (Appendix 9). Following Kraken2, Bracken was used to classify sequences to terminal  
300 taxa (Lu *et al.* (2017)). Finally all reads which could be classified by these databases were passed to a local  
301 BLAST database.

302 **2.6.3 | Identification of Sequence Matching Loci** A local NCBI database was built using the same  
303 processed novel and downloaded sequences as the previous database (Camacho *et al.* (2009)).

304 **2.7 | Integrated Observational, Molecular, and Palynological Corbiculae**

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

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

310 To reclassify the sequence reads, these data were combined with the flora observation data, and mapped  
311 by genus. If more than one species in the genus was flowering at that time and site, than the reads were  
312 split evenly between the taxa. For sequence data which did not match at the genus level, a user subjectively  
313 scored them based on the species composition and phenological activity at each site, the queen interaction  
314 data, and pollen assignments. To estimate the abundance of each of these species in the corbiculae loads,  
315 these data were combined with the microscopy data. For each morphotype detected in pollen, and each  
316 classified sequence read which was not detected via microscopy, they were given a value of 0.5% to indicate

their trace presences. When more than a single species belonged to a morphotype group in a single sample, the quantitative values from the morphological work were multiplied by the relative sequence abundance of each species in the load. All final compositions were standardized to a sum of 100%, by adding or subtracting the differences (induced by classifying records as ‘trace’) to all species with abundances > 1%.

## 3 | RESULTS

### 3.1 | Floral Observations

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

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

[Figure 2 about here.]

### 3.1 | Spatial Analyses

[Table 3 about here.]

[Table 4 about here.]

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

342 (25.009 km) to the third quantile (ca 43.830 km) of the regression distance, were true presences at the field  
343 station. Including these records would have resulted in modelling an additional 222 species distributions of  
344 which 30 are true presences, these taxa were not modelled.

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

347 In the area of the minimum-spanning tree encompassing the field sites, of the 554 vascular plants with biotic  
348 pollination syndromes, the 493 ML ensembles accurately predicted the presence of 362 (65.3%), incorrectly  
349 predicted the presence of 64 (11.6%), incorrectly predicted 34 true presences (6.1%) as being absent, and  
350 correctly predicted the true absence of 33 (6.0%). The balanced accuracy of the ensembled models is  
351 0.627 (Sensitivity = 0.340, Specificity 0.914). Of the 554 vascular plants with biotic pollination, the 475 LM  
352 ensembles accurately predicted the presence of 286 (51.6%), incorrectly predicted the presence of 41 (14.3%),  
353 incorrectly predicted 93 true presences (16.8%) as being absent, and correctly predicted the true absence of  
354 55 (9.9%). The balanced accuracy of the ensembled models is 0.664 (Sensitivity = 0.573, Specificity 0.754).  
355 Of the 554 vascular plants with biotic pollination syndromes in the flora 13 (2.3%) were in the Orchid family  
356 and 41 (7.4%) are non-natives, both of which are restricted from the database, and can only reduce the  
357 number of true predicted presences by roughly 10%.

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

### 363 3.2 | Microscopic Pollen identification

364 Using the fuchsin jelly preparation and light microscopic analyses of grains and scoring of 10 character states  
365 resulted in the establishment of 28 morphotypes which grains could be classified into (Appendix 6). From the  
366 37 samples that were counted and based on rarefaction we identified substantial amounts of the abundance  
367 and morphotype richness of the samples (morphotype richness,  $\bar{x} = 4.5$ , median = 4, min = 1, max = 9)  
368 (Appendix 7 & 8). The number of counted pollen grains in each sample range from (514 - 19924,  $\bar{x} = 3319$ ,  
369 median = 1891).

370

[Figure 3 about here.]

**371 3.3 | Metabarcoding Pollen Identification**

**372 3.3.2 | Temporal Analysis**

373 The first date of modeled snow melt in the Gothic area ( $n = 17$ ,  $\bar{x} = 137.9$ , Mdn = 135, 3<sup>rd</sup> quartile =  
374 151), and the first date of a consistent winter snow base ( $n = 17$ ,  $\bar{x} = 299.9$ , Mdn = 300, 1<sup>st</sup> quartile =  
375 291) from 2000-2017, were used as delimiters for the inclusions of herbarium records in modelling. Of the  
376 439 species predicted likely present in the area via logistic regression, 332 species (64.4%) with more than 10  
377 records in the focal level 4 ecoregions ( $\bar{x} = 35.016$ , Mdn = 35, max = 96) had Weibull estimates calculated,  
378 an additional 56 species (11.2%) with enough contributing records from the “Sedimentary Mid-Elevation  
379 Forests”, a large ecoregion generally just beneath the elevation bands occupied by the five ecoregions around  
380 the study area had Weibull estimates also calculated ( $\bar{x} = 13.868$ , Mdn = 13, max = 24).

381 Only 58 of these 388 species ( $n = 34.6$ , Mdn = 31) were able to be compared to plot based observational  
382 data from the long term (1974–2012) data set (CaraDonna *et al.* (2014)). Of these species relatively high  
383 accord was observed between the long-term ground truthed data set, and the modelled species. There was  
384 very strong evidence that the Weibull estimates were positively associated with the observed onset ( $p <$   
385 0.0001, tau = 0.61), peak ( $p < 0.0001$ , tau = 0.65), and cessation of flowering ( $p < 0.0001$ , tau = 0.49).  
386 There was moderate evidence that the Weibull estimates had a weak positive association with the observed  
387 duration of flowering ( $p = 0.58$ , tau = 0.17).

388 Of the previous 58 species compared, 47 of these could be compared to transect based data from the six  
389 sites observed in 2015. Due to methodological differences, the peak flowering was not compared, and due to  
390 the low performance of attempts to model ‘duration’ in the previous step it was also not compared. There  
391 was very strong evidence that the Weibull estimates were positively associated with the observed onset ( $p$   
392 < 0.0001, tau = 0.58), and cessation of flowering ( $p < 0.0001$ , tau = 0.40).

393 [Figure 4 about here.]

394 [Figure 5 about here.]

**395 3.3.1 | Molecular analysis of corbiculae loads**

396 The 54 corbiculae loads had DNA extracted and underwent various steps towards hyb-seq, in the end a total  
397 of 44 corbiculae samples were sequenced, 7,752,353 reads were recovered from sequencing. The number of  
398 reads per sequence varied widely (range = 76 - 508,795,  $\bar{x} = 176,189.8$ , Mdn = 138,395). Of the possible 353

399 loci, the number which were recovered from each sample, and informative to BLAST were range = 24 - 353,  
400  $\bar{x} = 305.5$ , Mdn = 331. The number of reads per loci from across all samples had a range of 178 - 506,653,  
401  $\bar{x} = 20,688$ , Mdn = 12,616 (Appendix 11).

402 After trimming 7,865,680 sequences remained. 10,682,538 reads were matched using Kraken, of the reads  
403 classified by Kraken 10,160,768 reads were matched using Bracken, of the reads classified by Kraken 7,549,608  
404 reads were matched using BLAST. Based upon subjective review of the three classifiers (Appendix 12)  
405 BLAST was chosen as the classification method which yielded the most probable results by the field ecologist,  
406 and its values were used for all subsequent analyses.

407 [Table 5 about here.]

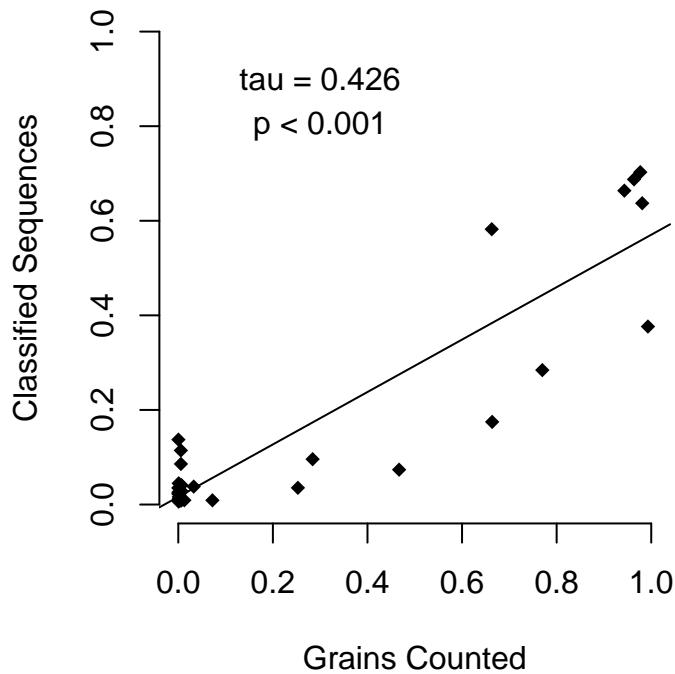
408 The initial classification of sequences which were made by BLAST were reviewed programmatically, using  
409 predicted presence of the species (from spatial modelling), modelled flowering time (from temporal mod-  
410 elling), and taxonomy (from existing sources). A sequential process was utilized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

418 Of the 0 classifications which were assigned to genera without any species predicted by spatial analyses, were  
419 investigated by hand after post-processing steps. These were all assigned via post-processing conditions (:  
420 , Appendix 15). These were manually assigned to a variety of ranks, occasionally to genus - 0, and species  
421 - 0, by consultation of the alpha-taxonomic literature (Sadeghian *et al.* (2015), Sennikov & Kurtto (2017),  
422 Rabeler & Wagner (2016), Pusalkar & Singh (2015), Moore & Bohs (2003), Weber (1998)).

423 To determine at which level species in pollen loads could be detected the results of light microscopy were  
424 compared to the molecular results. The pollen samples contained three morphotypes which could readily  
425 be identified via microscopy. Two of these mapped to the clades (Boraginaceae & Heliantheae Alliance),  
426 and one to a Asteraceae less Heliantheae. Boraginaceae grains were detected in 92.3% of samples where the  
427 proportion of target grains were between 0.01-1 ( $n = 13$  Mdn = 0.663). Asteraceae type 1, non-helianthoids,  
428 were detected in 50% of samples where the proportion of target grains were between 0.001-0.01 ( $n = 4$  Mdn  
429 = 0.001) Asteraceae type 2, Helianthoids, were detected in 33.3% of samples where the proportion of target

430 grains were between 0.001-0.01 ( $n = 6$  Mdn = 0.005); however, Asteraceae were detected in 80% of samples  
431 where the proportion of target grains were between 0.001-0.01 ( $n = 10$  Mdn = 0.003). Both morphotypes  
432 of Asteraceae pollen were detected in 100% of samples where the proportion of target grains were between  
433 0.01-1 ( $n = 2$  Mdn = 0.338).

### Correlation of Proportion Counted Grains and Sequence Reads



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

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

447 To ascertain the extent to which records of multiple species in a family, which were suspected to be sampling

448 artefacts occurred in molecular samples an index of similarity, ala Jaccard, the affinity index was used  
449 to assess co-occurrence (Mainali *et al.* (2022), Mainali & Slud (2022)). Numerous taxa from the family  
450 Ranunculaceae Jussieu (*Caltha* L. sp., *Thalictrum* L. spp., *Trollius* L. sp., *Aquilegia* L. spp.), had  $\alpha$  scores  
451 which indicated that they are only present when a more common confamilial taxa *Delphinium barbeyi* (Huth)  
452 Huth *nuttallianum* Pritz. were recorded. A similar relationship was observed in the Hydrophyllaceae R.Br.  
453 with samples placed in *Nemophila* Nutt., which only occurred when the more abundant *Hydrophyllum* L.  
454 species were present. The size of flower of *Nemophila breviflora* A. Gray make it unlikely to be visited by  
455 Bumble Bees, and it is a false positive. The floral morphology and orientation of flower of *Thalictrum* spp.  
456 also makes them unlikely to be visited, and while evidence of visits to *Caltha* and *Trollius* are lacking, due  
457 to the association between the reads these results appear unlikely.

### 458 3.6 | Integrated Observational, Molecular, and Palynological Network

459 While the spatial results were used to declare the taxonomic composition of the sequence database, temporal  
460 results were used in consideration with plant phylogeny to retroactively, reassign the assignment of sequences  
461 to taxa. Essentially, if a sequence was identified to a taxon which was not known from the field site  
462 For example many sequences which mapped to the Asteraceae family, but which was flagged by temporal  
463 filters and is present in both *B. nevadensis* Cresson and *B. rufocinctus* Cresson pollen is most likely *Frasera*  
464 Walter, which failed extractions for the reference library failed (Appendix 11). A similar likely mismatch  
465 could be between what was fide molecular evidence as *Agastache pallidiflora* (A. Heller) Rydb. but where  
466 feeding was infrequently observed on *Pedicularis* L., likely due to the Lamiales order being represented by  
467 only a single molecular reference species.

468 [Figure 6 about here.]

469 Situations where SDM's led to incorrect results at the species level are evident with classification to *Scabrethia*  
470 *Scabra* (Hooker) W.A. Weber, this match almost certainly representing *Wyethia arizonica* A. Gray (Weber  
471 (1998)), a taxon known to be visited by queen bee's via our floral observations. An expected inaccuracy of  
472 the classification scheme is in genus level placements, e.g. were *Epilobium* L. (Onagraceae Juss.) spp. were  
473 classified. However, given the small size of their flowers in the study area, these results more likely indicate  
474 that a species of *Chamaenerion* Seg. (a segregate genus) such as *C. angustifolium* (L.) Scop. or *latifolium*  
475 (L.) Sweet is occasionally utilized, as it supported by limited palynology data. An issue with reclassification  
476 within the family level in combination with time included reclassifying *Parnassia palustris* L. to *Paxistima*

477 *myrsinites* (Pursh) Rafinesque. However, based on flower size it is more likely that the visited taxon was *P.*  
478 *palustris*.

479 Regarding limitations of morphological data we suspect that there were two morphotypes of pollen identified  
480 as Ericaceae Juss. were actually Onagraceae (Samples 19 & 44), based on molecular results.

481 It is not unlikely that much of the difference in the results between the observational and molecular work  
482 are attributable to the challenges in detecting rare events in these smaller sizes. For example, no more than  
483 10 bee corbiculae loads per species were sequenced with the Mdn = 5.5 . . . , and the median of interactions  
484 with the top 5 plant sizes constituted 0.8283385 of the top interactions.

485 Accordingly, combining the results of floral observations, and palynology, molecular sequencing - both pre  
486 and post processing, we subjectively developed re-classifications of the contents of pollen grains . . .

## 487 4 | DISCUSSION

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

499 The SDM's which we generated, with relatively few occurrence records and few modelling iterations, per-  
500 formed beyond expectations, likely due to the utility of the predictor variables and strong alignment of  
501 vegetation by orographic precipitation in the study area. However, we had difficulties in evaluating our  
502 predictions in an operational context. We utilized the database query approach, to only model species with  
503 a high probability of not being dispersal limited to the focal area, and focused on a relevant subset of many  
504 of these species ranges to reduce the contributions of range wide adaptions on habitat (Sork (2018), Joshi  
505 *et al.* (2001)). While the models worked well compared to both test, and validation with external point

506 data, moving from points to polygon features was more difficult. We were able to compare our results to  
507 1) a Flora, 2) lists of plants used by Bumble Bees at plots; the former inappropriate in that it contained  
508 a great number of species which we sought to use modelling to reduce *e.g.* all strictly alpine species, and  
509 the latter inappropriate in that it contained only species relevant to *Bombus* but had no official ‘absence’  
510 data. Further given the size of the minimum spanning tree which we extracted points to, a formal floristic  
511 inventory would still be a time intensive process. Accordingly, we expect the real results of our data lay  
512 somewhere in between these two evaluations; with an excess of species predicted present (Dubuis *et al.*  
513 (2011), Calabrese *et al.* (2014), Pinto-Ledezma & Cavender-Bares (2021)), but few enough that they lend  
514 themselves to metabarcoding. We observe that our models seemed very capable of effectively identifying  
515 alpine species and removing them in binomial contexts. Difficulties in temporal models related to variability  
516 in drivers of flowering phenology, and relatively small appropriate sample sizes per species.

517 Based on our results, we believe the overall results between observations of queen Bumble Bee and barcoding  
518 are largely congruent. Which relaxes concerns regarding differences between the insights gleaned from  
519 observational, as compared to data derived from the pollen records (Barker & Arceo-Gomez (2021), Zhao *et*  
520 *al.* (2019), Alarcón (2010)). This indicates that results from fine resolution survey’s, may serve as calibration  
521 standards to projects which expand geographically outward from them, and seek to make inference to larger  
522 areas. We analyzed pollen loads from all of the most common bumble bee species in the area and were able  
523 to . . . (Pyke (1982)). The species which appeared most often in loads via the molecular records, yet seldom  
524 noted via observations, mostly had affinities to wetlands or stream banks. These habitats are not often  
525 sampled by terrestrial ecologists due to logistics and concerns regarding degrading them. This indicates that  
526 metabarcoding may be applied in studies where areas, or administrative units, prohibit the disturbances  
527 associated sampling.

528 [Table 6 about here.]

529 Results from palynological analyses show that several species of bee show near perfect fidelity to the genus  
530 *Mertensia* on a per visit basis. . . General results show high congruence between foraging and molecular  
531 results, indicating that concerns regarding mismatch between observational networks need not persist with  
532 *Bombus* studies. . .

533 Some foraging preferences of *Bombus*, both at this field site and across a great many localities globally  
534 emerge from this work, which reiterates the needs for land managers to maintain relatively high amounts  
535 of members of the Fabaceae, Boraginaceae, and Ranunculaceae, in Western North American montane land-  
536 scapes (Goulson *et al.* (2005), Goulson (2010), Liang *et al.* (2021), Bontsutsnaja *et al.* (2021)). Numerous

537 historic, and some ongoing, land management practices reduce the ability of many landscapes to support  
538 stable populations of *Bombus*. Historic livestock grazing was often associated with the targeted removal  
539 of many species of plants which are known to have compounds toxic to cattle. In particular, the removal  
540 of locoweeds (Fabaceae: *Astragalus* L. & *Oxytropis* DC.) and larkspurs (Ranunculaceae: *Delphinium*) were  
541 common across public lands administered by the United States Forest Service (Ralphs & Ueckert (1988),  
542 Aldous (1919), Ralphs *et al.* (2003)). Further actions, generally initiated by early settlers, involved the  
543 channelization and incising of streams, culling of beavers, and leaving cattle concentrated on higher order  
544 stream banks for significant periods of time, all processes which lower the water tables and reduced the ex-  
545 tent of stream-associated [riverine] wetlands and the mesic meadows fringes which provide habitat for many  
546 species of tall *Mertensia* (Boraginaceae, e.g. *M. ciliata* Torr. G. Don.) widely distributed across Western  
547 North America, and to an extent *Delphinium barbeyi* and many species of native *Trifolium* L. (Dahl (1990),  
548 Naiman *et al.* (1988), Belsky *et al.* (1999), Cooke & Reeves (1976)). Fire suppression further resulted in the  
549 succession of many Aspen (*Populus tremuloides* Michx.) groves to Conifer stands, decreasing the mosaic of  
550 age structured habitats in many landscapes, adversely effects habitat for tall *Mertensia* species and several  
551 species of *Delphinium* (Brewen *et al.* (2021), Keane (2002)). Finally the effects of Nitrogen deposition, es-  
552 pecially given the West's rapidly growing population still pose adverse effects on the abundance of a variety  
553 of species of Fabaceae at Urban-Rural interfaces (see Stevens *et al.* (2018), Fenn *et al.* (2003)). Current  
554 solutions to these issues, involve targeted burns, reintroduction of beavers and beaver habitat analogs, and  
555 the possibility of re-seeding a variety of 'locoweeds' and 'larkspurs' in areas now seldom used, or only used  
556 for early, grazing. The highly enthusiastic response of land managers, and homeowners, to plant *Asclepias*  
557 L., using genetically appropriate materials, to improve Monarch Butterfly (*Danaus plexippus* L.) habitat  
558 provides an effective framework for the latter (Oberhauser *et al.* (2015), Basey *et al.* (2015)).

559 We have concerns regarding the number of persons training to become and practice botany, and grave  
560 concerns regarding the funding mechanisms for floristic and field based botanical research and for centralized  
561 authorities to produce consensus opinions on alpha taxonomy (Prather *et al.* (2004b), Kramer & Havens  
562 (2015), Prather *et al.* (2004a), Crisci *et al.* (2020), Manzano (2021), Stroud *et al.* (2022)). To reduce  
563 the effects of a low population density of botanists on the maintenance of and production of Flora's and  
564 to foster meta-genomics across landscapes without field stations we utilized Species Distribution Modelling  
565 to generate predictive species lists. In this proof-of-concept example we performed several iterations of  
566 modelling runs, and several approaches (i.e. the 'linear models', and the 'machine learning'), which took  
567 notable amounts of compute power. We suspect the possible deleterious nature of this endeavor may be  
568 reduced by: 1) more field surveying by crews will reduce the need to generate as many species 2) fewer

569 runs of models, 3) only running machine learning models which do not require an explicit process to reduce  
570 spatial autocorrelation. However, given the time required to perform all aspects of a study, even our amount  
571 of computation was negligible. Further, we are very optimistic about the possibility for persons to perform  
572 these tasks, as mentioned we utilized roughly only one quarter of the records which were digitally available  
573 for presence, and we suspect others will have enough records to perform this process nearly anywhere else in  
574 the temperate. In certain scenarios modelling of predicted species via more formally tailored S(tacked)-SDM  
575 or J(oint)-SDM approaches may be beneficial (Wilkinson *et al.* (2021), Pinto-Ledezma & Cavender-Bares  
576 (2021), Schmitt *et al.* (2017)).

577 Tandem to the lack of continued expertise required to generate and maintain species lists, is the expertise  
578 required to continue tracking when major phenological events occur in many plant species at relatively fine  
579 scales or under novel climates. Knowledge of these events is currently limited to general time periods of only  
580 a handful of phenological events and groups of organisms (e.g. flowering initiation, or trees) (Prather *et al.*  
581 (2004a), Li *et al.* (2016)). While many programs and initiatives exist to collect phenological information on  
582 subsets of easily identifiable charismatic species to detect major trends in phenology, these capture only a  
583 subset of the extent diversity (Betancourt *et al.* (2005), Havens *et al.* (2007)). In many instances it appears  
584 that while landscapes respond similarly to environmental variables which predict phenological responses,  
585 that individual species vary widely in their responses to similar environmental cues, or respond to different  
586 cues (Augspurger & Zaya (2020), Xie *et al.* (2015), Xie *et al.* (2018), CaraDonna *et al.* (2014)). As can  
587 be seen here, predictions of when a single, major phenological event occurs is already data limited. A more  
588 promising approach for the tropics may lay in utilizing circular statistics (Park *et al.* (2022)).

589 The nearly complete Plant and Fungal Tree of Life (PAFTOL) will provide a comprehensive phylogenetic  
590 backbone of the entire plant kingdom, and the inclusion of A353 probes with lineage specific probe sets is  
591 common in producing massive genetic datasets (Baker *et al.* (2021b)). We predict that the A353 probes  
592 which it is utilizing to work nearly immediately for DNA barcoding of whole plant material, and that more  
593 elaborate validation studies in controlled metabarcoding settings, utilizing existing experimental designs,  
594 will have favorable results (Bell *et al.* (2017), Bell *et al.* (2019), Bell *et al.* (2021), Lamb *et al.* (2019)). In  
595 particular the harvesting of loci with more variation in certain lineages, and or with more variable flanking  
596 regions, will prove promising for identifying closely related plant material. We suspect that conserved reaches  
597 of genes resulted in the high amounts of reads in somewhat obscure species. Given that the A353 loci are  
598 nuclear, single copy, and a variety are present the possibility of identifying target loci for quantitative  
599 purposes is high, without continual PCR enrichment is possible; this would align with relatively high efficacy  
600 of WGS (Lang *et al.* (2019), Peel *et al.* (2019), Bell *et al.* (2021)). Recent evidence indicates that the

601 potential for identifying nearly cryptic taxa and even infra-specific inference, of either whole plant material,  
602 and perhaps in metagenomic context are possible (Ottenlips *et al.* (2021), Wenzell *et al.* (2021), Loke et  
603 al. in prep, Slimp *et al.* (2021), Beck *et al.* (2021)). We further believe that in synthetic phylogenetic  
604 trees - with incorporation of NGS backbones - will allow in automatic reassignment of reads as a function of  
605 phylogenetic distance with measures of uncertainty (Hinchliff *et al.* (2015), Smith & Brown (2018), Baker  
606 *et al.* (2021a)).

## 607 5 | CONCLUSION

608 We believe that the combination of spatial and temporal models, united and guided by localized natural  
609 history knowledge, provides the essential components of a bayesian framework for approaching the coarse  
610 elucidation of ecological interactions using DNA Barcoding. Herein we crudely utilized this thinking via  
611 binary outcomes, should a species predicted be predicted present or not? Is it unequivocally flowering  
612 or not? Myriad data show biological systems and ecological interactions have more variance than can be  
613 reasonably discretely parsed. We expect that within a bayesian framework studies of pollinator behavior  
614 may be enacted via this approach at a landscape level, e.g. the scale of an entire drainage basin such as the  
615 Gunnison which is quickly becoming one of the worlds few model ecosystems. We hope that the A353 probes  
616 as tools for metabarcoding play a role in these endeavors.

617 **AUTHOR CONTRIBUTIONS:** R.C.B conducted botanical collections, conducted all molecular lab  
618 work, lead all analyses, and writing. J.E.O conceived, designed, and conducted all ecological fieldwork,  
619 assisted with analyses, and writing. E.J.W. prepared, imaged, and collected trait data on pollen reference  
620 slides, and assisted with analysis of trait data and writing a dichotomous key. S.T. assisted with spatial  
621 analyses and writing. P.J.C assisted with ecological analyses and writing. J.B.F. conceived, and designed all  
622 lab work, analyses, and integration of approaches, assisted with writing, and secured funding for molecular  
623 work.

624 **ACKNOWLEDGMENTS:** Nyree Zerega for assistance obtaining herbaria loans and accessioning our  
625 collections at CHIC. Pat Herendeen for assistance with virtually all aspects of preparing pollen vouchers and  
626 the identification process. Hilary Noble, Zoe Diaz-Martinez, Angela McDonnell, & Elena Loke for assistance  
627 with genomic library preparation. Ian Breckheimer for sharing the SDM predictor variables. We thank  
628 the curators at the following herbaria for supplying tissue: Ben Legler at Stillinger (ID), Charles (Rick)  
629 Williams at Ray J. Davis (IDS), (B)Ernie Nelson at Rocky Mountain (RM); and the collectors: D. Knoke,  
630 L. Brummer, J. Boyd, C. Davidson, I. Gilman, M. Kirkpatrick, S. McCauley, J. Smith, K. Taylor, & C.

631 Williams. David Giblin & Mare Nazaire for sharing relevant sections of an advanced draft of FNA V. 15.  
632 The Bureau of Land Management is thanked as many plant specimens were collected by R.C.B as a partner  
633 or contractor to the agency; Sarah Burnett and Lauren Price are thanked for sharing AIM data. Sanda and  
634 New England Biotech are gratefully acknowledged for technical support and generously sharing samples.  
635 T.C.H. Cole for sharing the Angiosperm Phylogeny 4 colour palette. The Program in Plant Biology and  
636 Conservation is thanked for funding. The holdings of the following herbaria were essential for this project:  
637 AK, ALTA, ASU, BABY, BC, BM, BMO, BOON, BRIT, CANB, CAS, CHSC, CM, CMN, CNS, COLO,  
638 CONN, CS, CSU, DAV, DBG, DES, ENCB, F, FR, G, GH, GZU, IAC, K, KR, KSP, KSTC, KU, LD, LOB,  
639 LSU, MA, MACF, MEL, MICH, MIL, MIN, MNHN, MO, MT, MW, NCSC, NSW, NY, O, OBI, PI, RBG,  
640 RSA, SD, SDSU, SFV, TENN, TRT, UA, UAC, UAM, UAZ, UBC, UBC, UCR, UCS, UCSB, UMO, UNM,  
641 UPS, US, USCH, USF, USU, UTEP, UWBM, V, VT, W, WSCO, WU, XAL, YPM, Z.

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

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

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

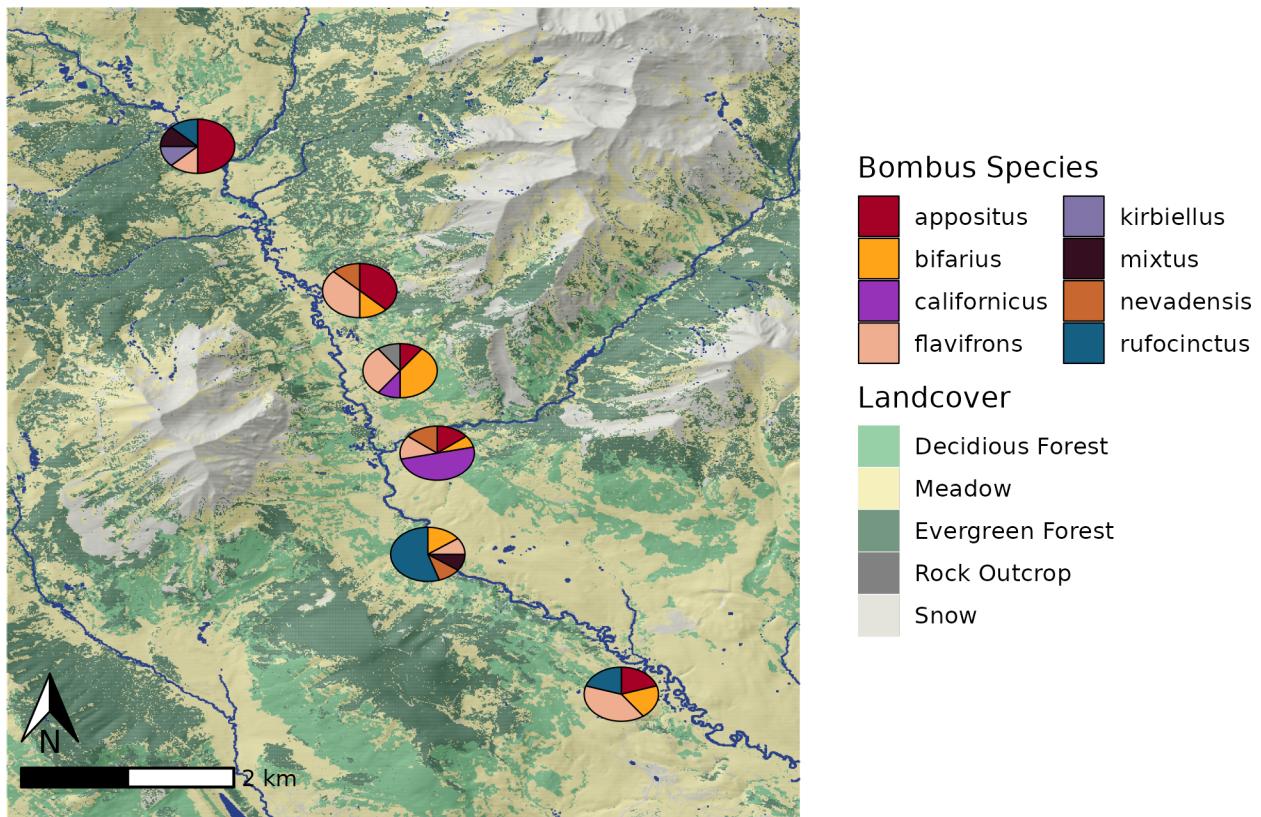
646 **ORCID**

647 Paul CaraDonna <https://orcid.org/0000-0003-3517-9090>  
648 Jeremie Fant <https://orcid.org/0000-0001-9276-1111>  
649 Jane Ogilvie <https://orcid.org/0000-0001-8546-0417>  
650 Sophie Taddeo <https://orcid.org/0000-0002-7789-1417>

651 **References**

652 **Supporting**

## Origins of Corbiculae Loads



Upper East River Valley, Colorado

## 655 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 <sub>2</sub> 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>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 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>Campanula rotundifolia</i> L.	Campanulaceae	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	982.5
<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

<sup>a</sup> Accession includes both Herbarium and Accession number

<sup>b</sup> Pres. refers to Preservation method. 'S' denotes silica gel dried, 'P' denotes pressed

<sup>c</sup> All Localities are in the United States of America

## 659 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.

663 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)

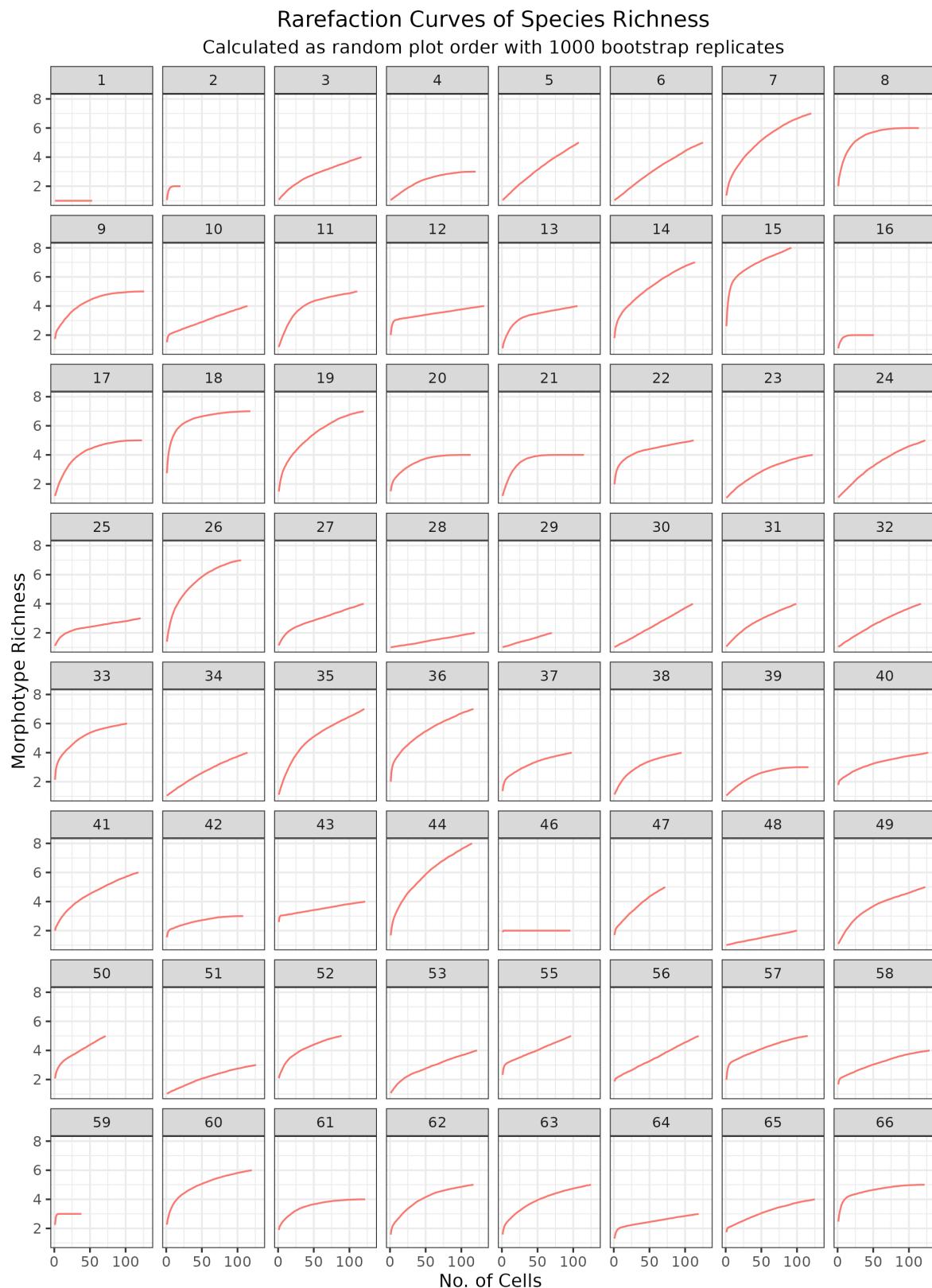
34b: Apertures not annulate – grains appear more or less circular (~ERICACEAE)

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)

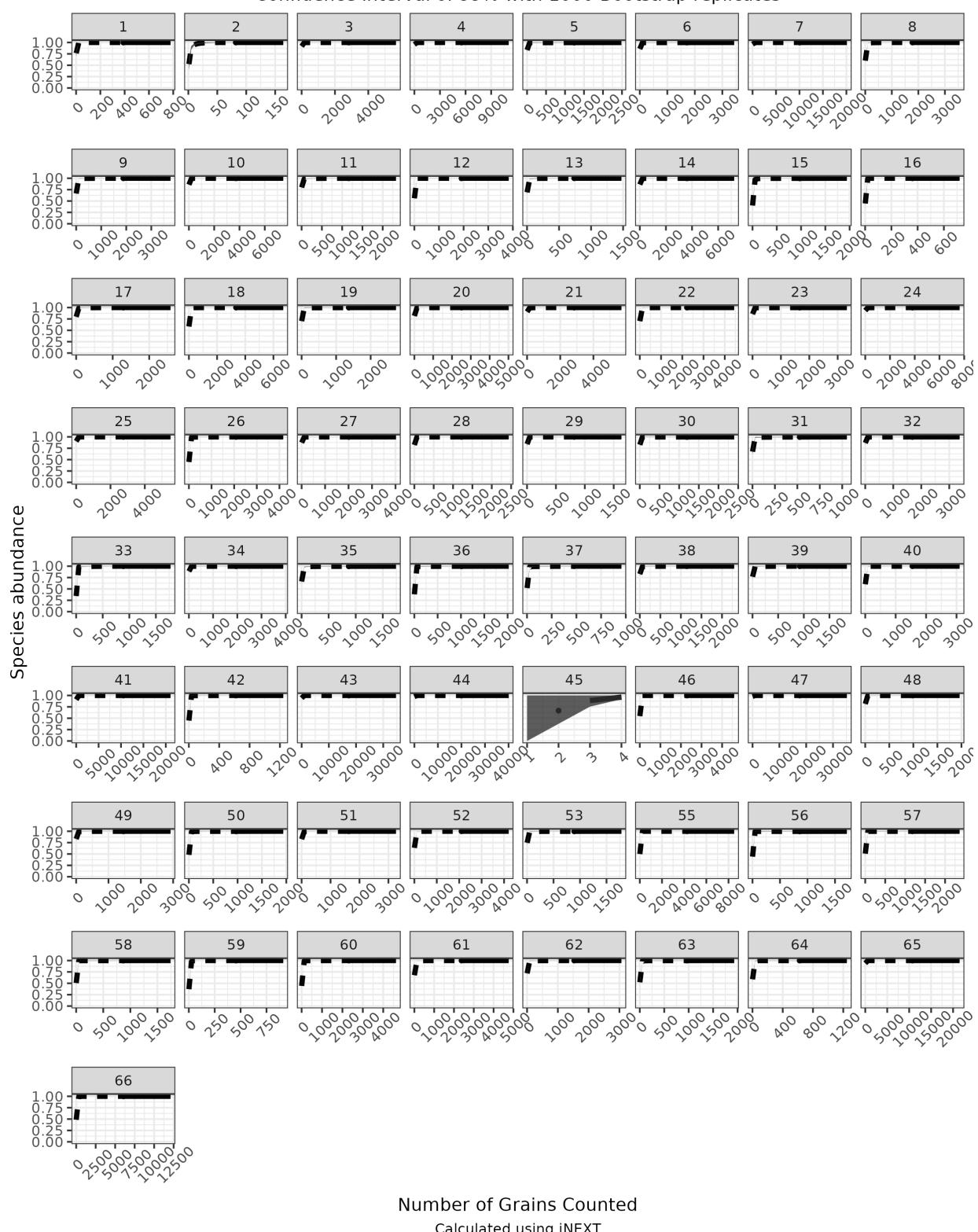
Morphotype D: Salix, Boechera



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



Number of Grains Counted

Calculated using iNEXT

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>

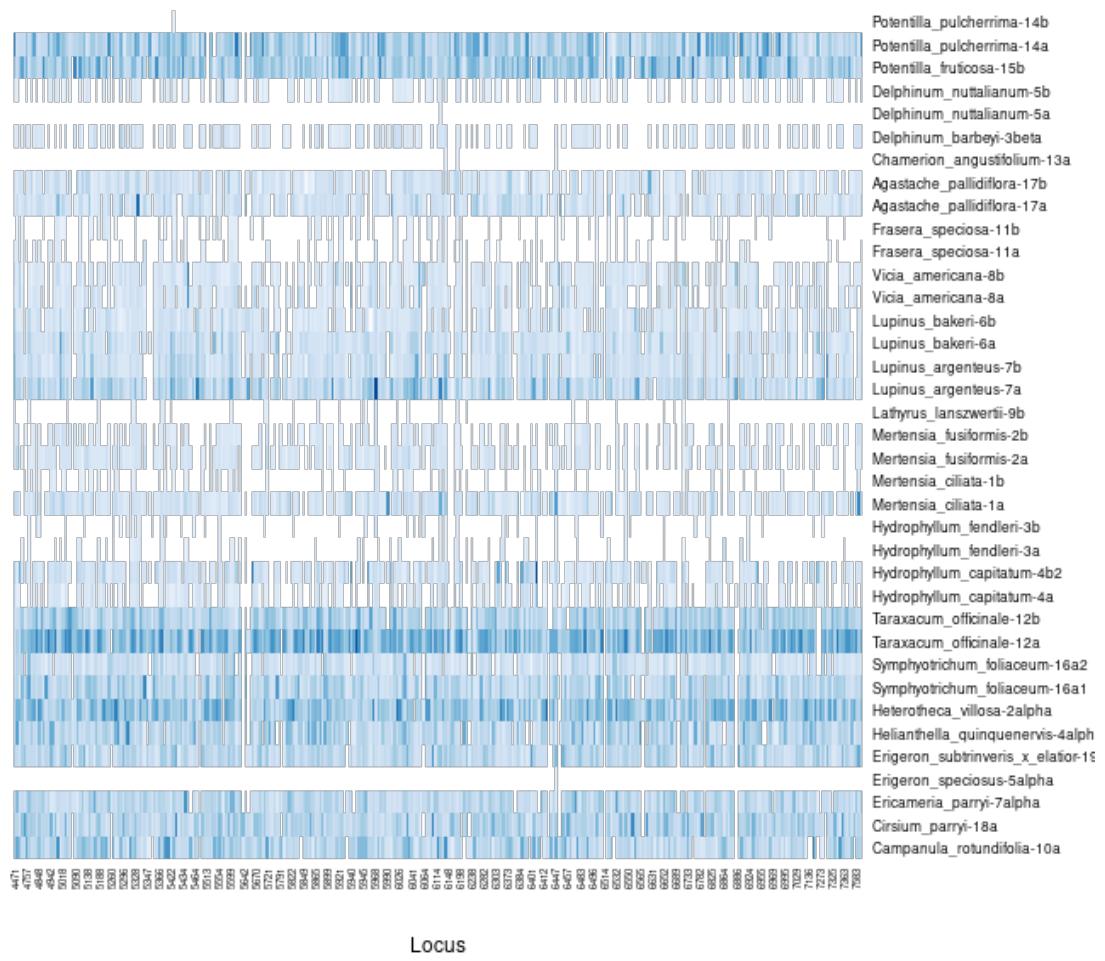
(Continued on Next Page)

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

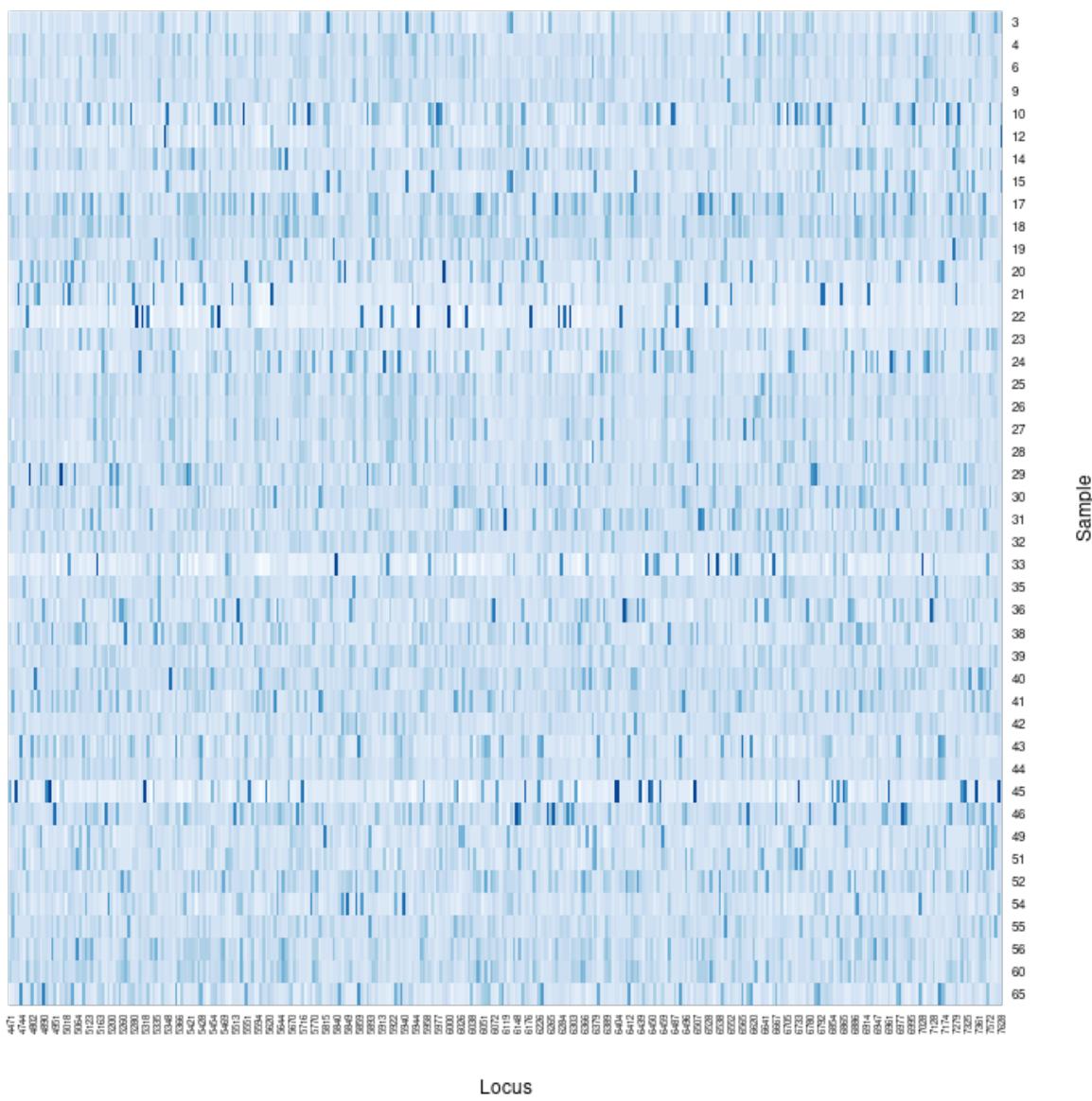
Order	Family	Taxon
		Saxifraga maderensis
		Saxifraga oppositifolia
		Saxifraga portosanctana
		Saxifraga x geum

x geum\\* \end{longtable}

## Loci & Nucleotides Returned per Reference Sample

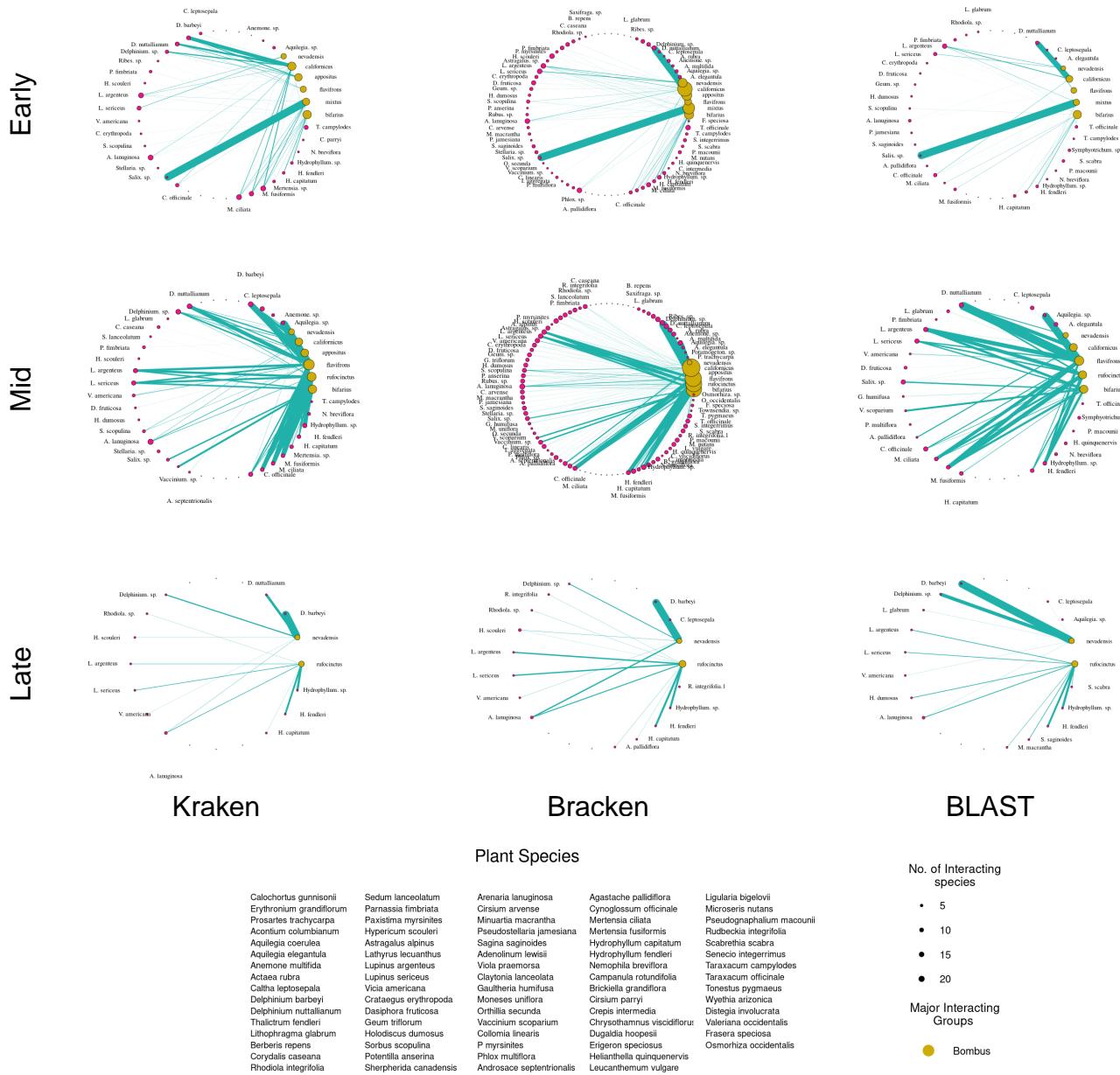


### Percent matched reads



680

## Comparision of Foraging Patterns from Three Sequence Alignment Algorithms



683 Appendix 13 - Models used for Species Distribution Model Ensembles

684 The two machine learning models utilize Ensemble learning.

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

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

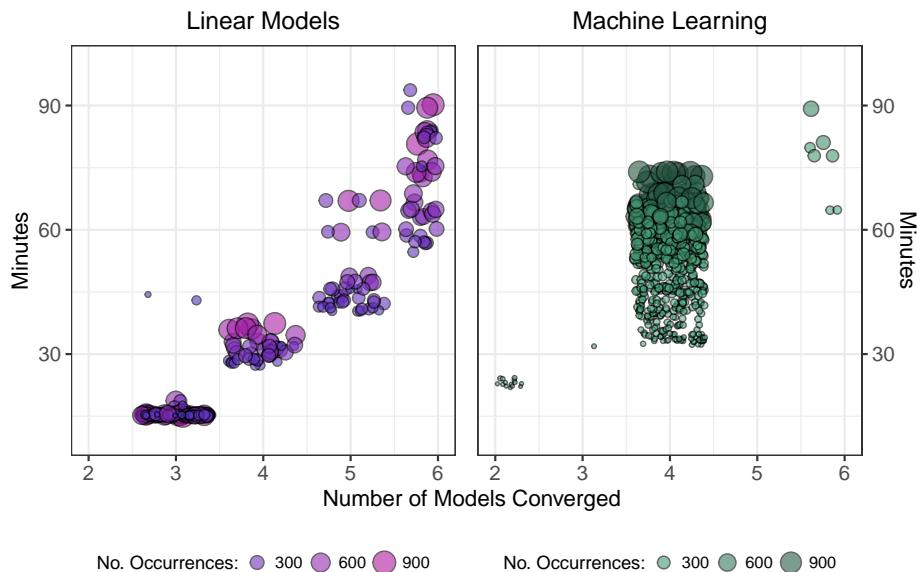
697 **Boosted Regression Tree (BRT)** (or Gradient Boosted tree) An initial tree is grown, and all other trees  
698 are derived sequentially from it, as each new tree is grown the errors in responses from the last tree are  
699 weighed more heavily so that the model focuses on selecting dependent variables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

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

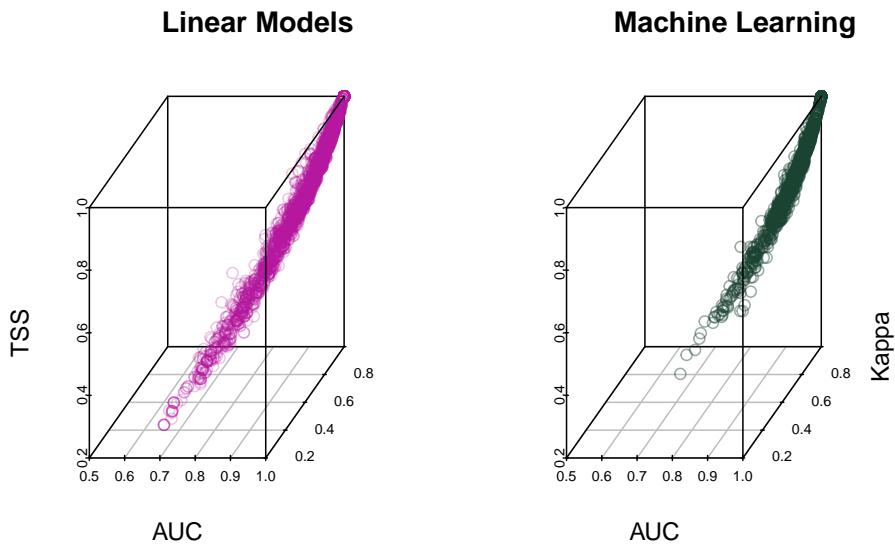
703 **Variance** errors in models due to an over-reliance and sensitivity of training to outliers in training data.

704 In general, Random Forest models have high bias and low variance, where boosted regressions trees have lower  
705 bias and higher variance. Theoretically, the weaknesses and strengths of bootstrap aggregation (bagging) as  
706 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.

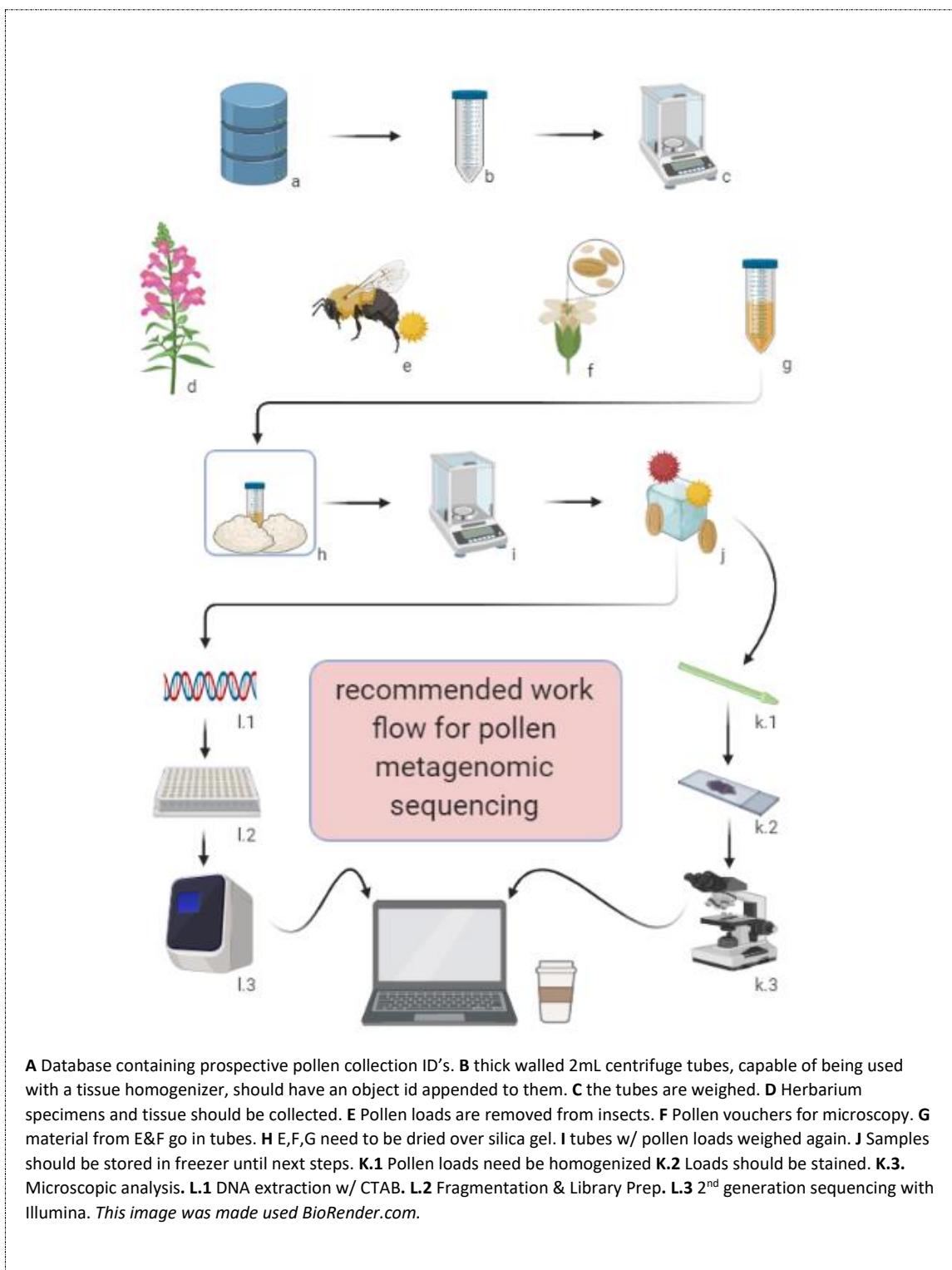
$$\begin{aligned} \text{Spatial} == 1 \& \text{ Temporal} == 1 \sim \mathbf{A} \\ \text{Spatial} == 1 \& \text{ Temporal} == 0 \& \text{ Congener} = 1 \sim \mathbf{B} \end{aligned}$$

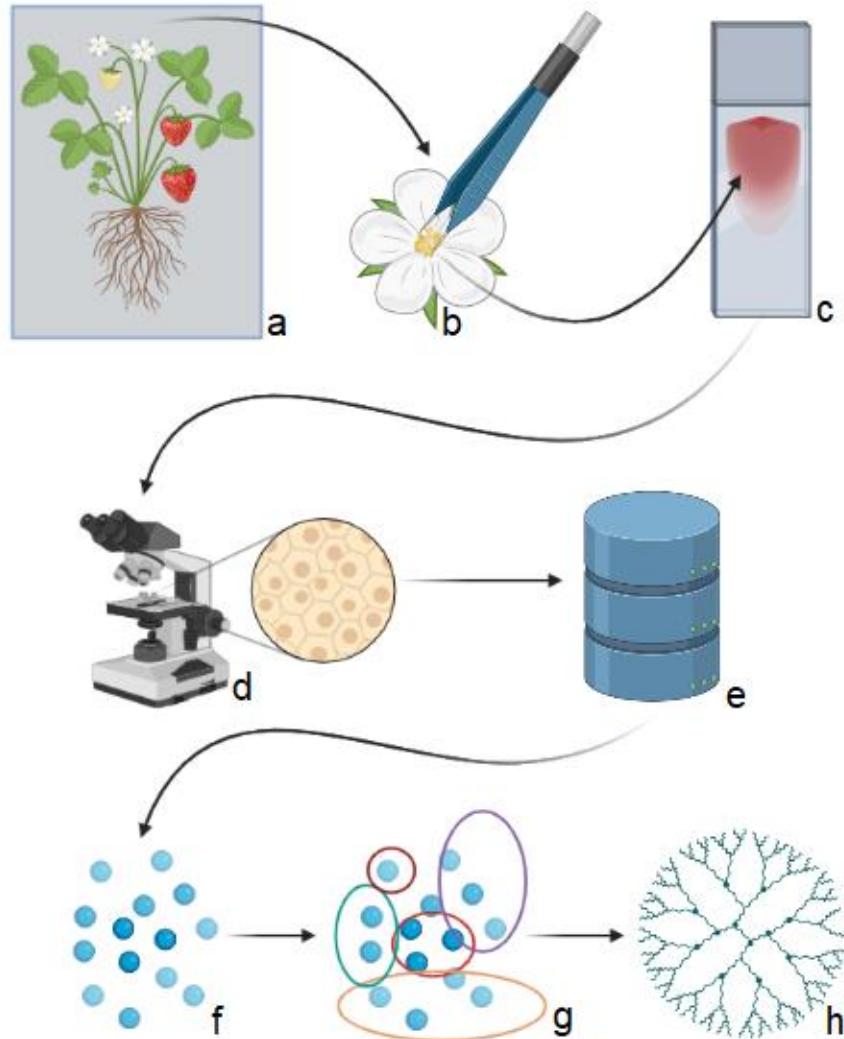
*The temporal dimension is now buffered and a form of  $\mathbf{A}$  is employed*  
 $\text{Spatial} == 1 \& \text{ Temporal} +/- \text{ Buffer} == 1 \sim \mathbf{X}$

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

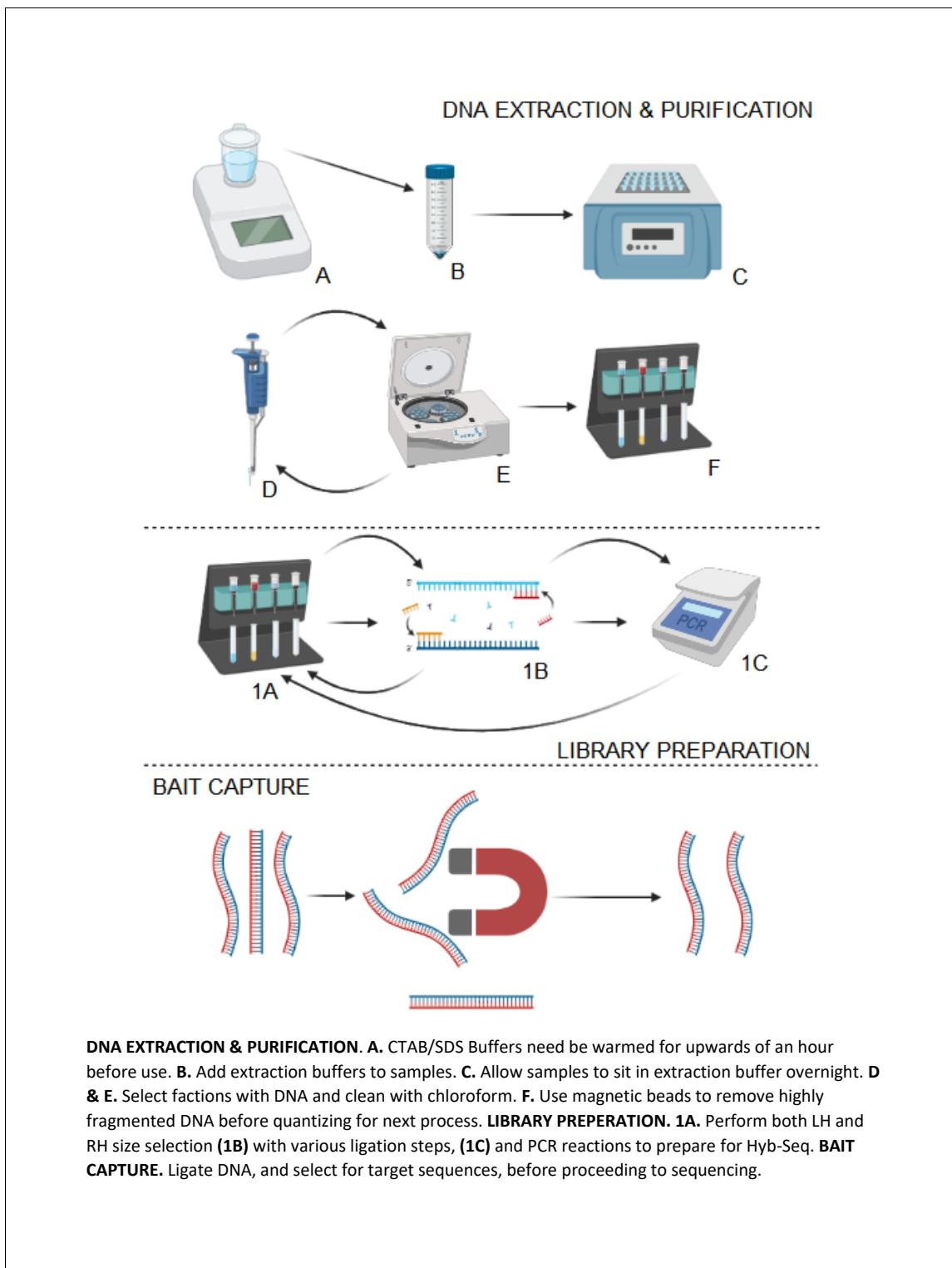
$$\begin{aligned} \text{Spatial} == 0 \& \text{ Temporal} == 0 \& \text{ Congener} == 1 \sim \mathbf{G} \\ \text{Spatial} == 0 \& \text{ Temporal} == 0 \& \text{ Congeners} == 1 \sim \mathbf{H} \\ \text{Spatial} == 0 \& \text{ Temporal} == 0 \& \text{ Confamilial} == 1 \sim \mathbf{I} \\ \text{Spatial}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

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.



## CTAB-DNA POLLEN EXTRACTIONS

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

### SAMPLE PREPARATION AND GRINDING

- a1) Add 380  $\mu$ 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  $\mu$ 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  $\mu$ L 10% CTAB buffer.
- b3) Add 10  $\mu$ 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  $\mu$ L proteinase K (20mg/mL) & 12.5  $\mu$ L DTT (1 molar in H<sub>2</sub>O); 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  $\mu$ 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  $\mu$ 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  $\mu$ L of 75% EtOH, invert tube x3, centrifuge at 13,000 rpm for 4 minutes; discard supernatant
- c4) Add 400  $\mu$ 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  $\mu$ L of dna free H<sub>2</sub>O 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 H<sub>2</sub>O)**

#### 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 deH<sub>2</sub>O

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

Fill to 50 mL with deH<sub>2</sub>O

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 DiH<sub>2</sub>O)**

#### For 100 samples (50 mL solution)

add ~30 mL deH<sub>2</sub>O,

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 deH<sub>2</sub>O

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 deH<sub>2</sub>O

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

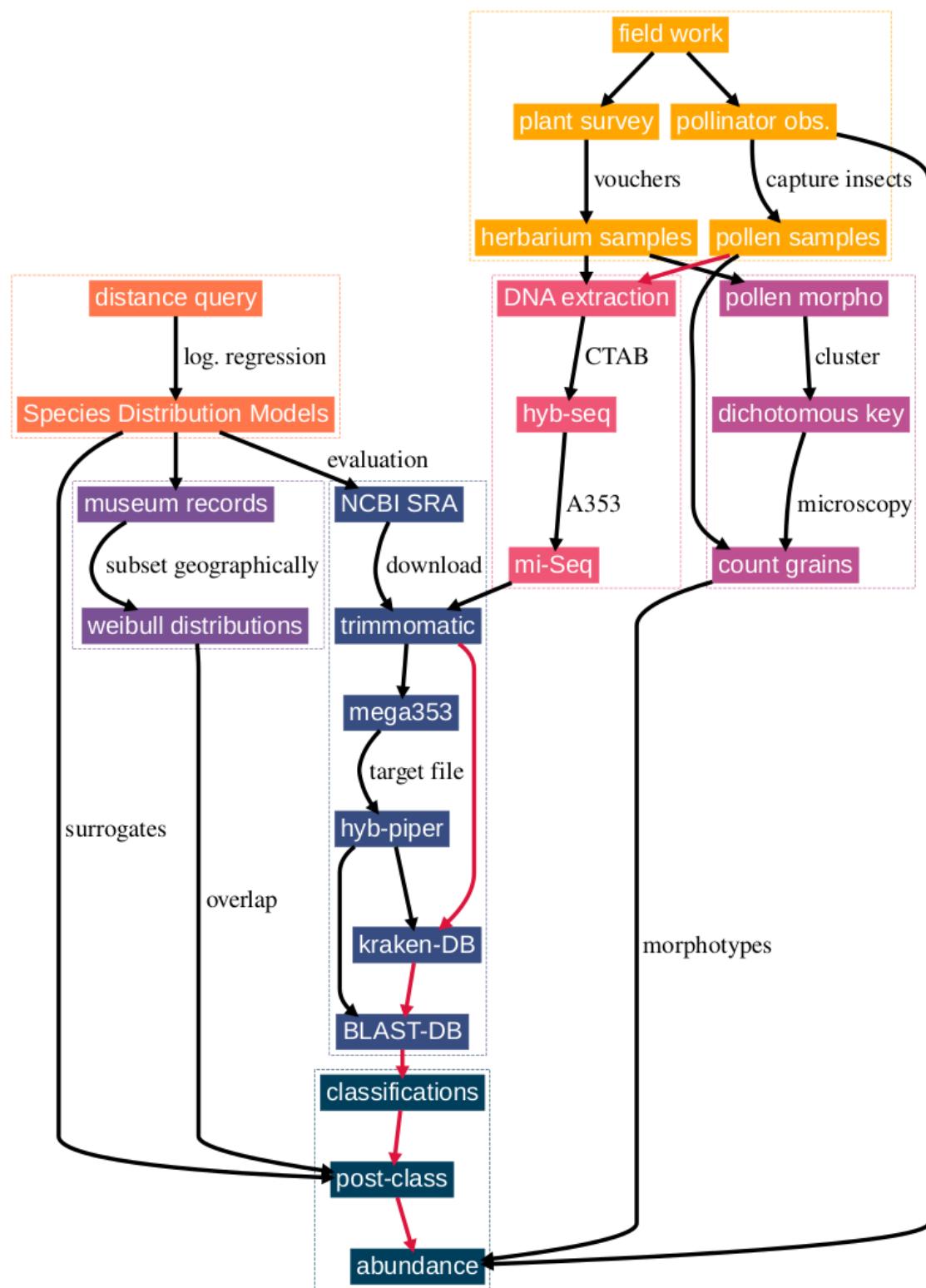
717 Appendix 17 Samples Per Bee

Table 1: Queen Bee Pollen Loads examined

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

<sup>a</sup> All subgenera follow the system of Williams et al. 2008, and placements were found from the NMH website.

<sup>b</sup> Tongue Lengths collected from Pyke et al. 2012



721 **References**

- 722 Ackerfield, J. (2015). *Flora of colorado*. BRIT Press Fort Worth.
- 723 Alarcón, R. (2010). Congruence between visitation and pollen-transport networks in a California plant–  
724 pollinator community. *Oikos*, **119**, 35–44. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0706.2009.17694.x>
- 725
- 726 Aldous, A.E. (1919). *Eradicating tall larkspur on cattle ranges in the national forest*. US Department of  
727 Agriculture.
- 728 Allouche, O., Tsoar, A. & Kadmon, R. (2006). Assessing the accuracy of species distribution models:  
729 Prevalence, kappa and the true skill statistic (TSS). *Journal of applied ecology*, **43**, 1223–1232.
- 730 Allred, K.W. & Ivey, R. (2012). Flora neomexicana III: An illustrated identification manual. *Lulu. com*.
- 731 Araujo, M.B. & New, M. (2007). Ensemble forecasting of species distributions. *Trends in ecology & evolution*,  
732 **22**, 42–47.
- 733 Augspurger, C.K. & Zaya, D.N. (2020). Concordance of long-term shifts with climate warming varies among  
734 phenological events and herbaceous species. *Ecological Monographs*, **90**, e01421.
- 735 Baker, W.J., Bailey, P., Barber, V., Barker, A., Bellot, S., Bishop, D., Botigué, L.R., Brewer, G., Carruthers,  
736 T., Clarkson, J.J., Cook, J., Cowan, R.S., Dodsworth, S., Epitawalage, N., Françoso, E., Gallego, B.,  
737 Johnson, M.G., Kim, J.T., Leempoel, K., Maurin, O., Mcginnie, C., Pokorny, L., Roy, S., Stone, M.,  
738 Toledo, E., Wickett, N.J., Zuntini, A.R., Eiserhardt, W.L., Kersey, P.J., Leitch, I.J. & Forest, F. (2021a).  
739 A Comprehensive Phylogenomic Platform for Exploring the Angiosperm Tree of Life. *Systematic Biology*,  
740 **71**, 301–319. Retrieved from <https://doi.org/10.1093/sysbio/syab035>
- 741 Baker, W., Dodsworth, S., Forest, F., Graham, S., Johnson, M., McDonnell, A., Pokorny, L., Tate, J., Wicke,  
742 S. & Wickett, N. (2021b). Exploring Angiosperms353: An open, community toolkit for collaborative  
743 phylogenomic research on flowering plants. *American Journal of Botany*, **108**.
- 744 Banerjee, P., Stewart, K.A., Dey, G., Antognazza, C.M., Sharma, R.K., Maity, J.P., Saha, S., Doi, H., Vere,  
745 N. de, Chan, M.W. & others. (2022). Environmental DNA analysis as an emerging non-destructive  
746 method for plant biodiversity monitoring: A review. *AoB Plants*, **14**, plac031.
- 747 Barbet-Massin, M., Jiguet, F., Albert, C.H. & Thuiller, W. (2012). Selecting pseudo-absences for species  
748 distribution models: How, where and how many? *Methods in ecology and evolution*, **3**, 327–338.
- 749 Barker, D.A. & Arceo-Gomez, G. (2021). Pollen transport networks reveal highly diverse and temporally  
750 stable plant–pollinator interactions in an Appalachian floral community. *AoB PLANTS*, **13**. Retrieved  
751 from <https://doi.org/10.1093/aobpla/plab062>
- 752 Basey, A.C., Fant, J.B. & Kramer, A.T. (2015). Producing native plant materials for restoration: 10 rules

- 753 to collect and maintain genetic diversity. *Native Plants Journal*, **16**, 37–53.
- 754 Beattie, A. (1971). A technique for the study of insect-borne pollen. *The Pan-Pacific Entomologist*, **47**, 82.
- 755 Beck, J.B., Markley, M.L., Zielke, M.G., Thomas, J.R., Hale, H.J., Williams, L.D. & Johnson, M.G. (2021).
- 756 Are palmer's elm-leaf goldenrod and the smooth elm-leaf goldenrod real? The Angiosperms353 kit  
757 provides within-species signal in solidago ulmifolia sl. *Systematic Botany*, **46**, 1107–1113.
- 758 Belitz, M.W., Larsen, E.A., Ries, L. & Guralnick, R.P. (2020). The accuracy of phenology estimators for use  
759 with sparsely sampled presence-only observations. *Methods in Ecology and Evolution*, **11**, 1273–1285.
- 760 Bell, K.L., Burgess, K.S., Botsch, J.C., Dobbs, E.K., Read, T.D. & Brosi, B.J. (2019). Quantitative and  
761 qualitative assessment of pollen DNA metabarcoding using constructed species mixtures. *Molecular  
762 Ecology*, **28**, 431–455.
- 763 Bell, K.L., Fowler, J., Burgess, K.S., Dobbs, E.K., Gruenewald, D., Lawley, B., Morozumi, C. & Brosi, B.J.  
764 (2017). Applying pollen DNA metabarcoding to the study of plant–pollinator interactions. *Applications  
765 in plant sciences*, **5**, 1600124.
- 766 Bell, K.L., Petit III, R.A., Cutler, A., Dobbs, E.K., Macpherson, J.M., Read, T.D., Burgess, K.S. & Brosi,  
767 B.J. (2021). Comparing whole-genome shotgun sequencing and DNA metabarcoding approaches for  
768 species identification and quantification of pollen species mixtures. *Ecology and Evolution*, **11**, 16082–  
769 16098.
- 770 Bell, K.L., Turo, K.J., Lowe, A., Nota, K., Keller, A., Encinas-Viso, F., Parducci, L., Richardson, R.T.,  
771 Leggett, R.M., Brosi, B.J. & others. (2022). Plants, pollinators and their interactions under global  
772 ecological change: The role of pollen DNA metabarcoding. *Molecular ecology*.
- 773 Belsky, A.J., Matzke, A. & Uselman, S. (1999). Survey of livestock influences on stream and riparian  
774 ecosystems in the western united states. *Journal of Soil and water Conservation*, **54**, 419–431.
- 775 Bergman, P., Molau, U. & Holmgren, B. (1996). Micrometeorological impacts on insect activity and plant  
776 reproductive success in an alpine environment, swedish lapland. *Arctic and alpine research*, **28**, 196–202.
- 777 Betancourt, J.L., Schwartz, M.D., Breshears, D.D., Cayan, D.R., Dettinger, M.D., Inouye, D.W., Post, E.  
778 & Reed, B.C. (2005). Implementing a US national phenology network.
- 779 Bingham, R.A. & Orthner, A.R. (1998). Efficient pollination of alpine plants. *Nature*, **391**, 238–239.
- 780 Bolger, A. & Giorgi, F. (2014). Trimmomatic: A flexible read trimming tool for illumina NGS data. *Bioin-  
781 formatics*, **30**, 2114–2120.
- 782 Bontsutsnaja, A., Karise, R., Mand, M. & Smagghe, G. (2021). Bumble bee foraged pollen analyses in spring  
783 time in southern estonia shows abundant food sources. *Insects*, **12**, 922.
- 784 Bortolus, A. (2008). Error cascades in the biological sciences: The unwanted consequences of using bad  
785 taxonomy in ecology. *AMBIO: A journal of the human environment*, **37**, 114–118.

- 786 Brewen, C.J., Berrill, J.-P., Ritchie, M.W., Boston, K., Dagley, C.M., Jones, B., Coppoletta, M. & Burnett,  
787 C.L. (2021). 76-year decline and recovery of aspen mediated by contrasting fire regimes: Long-unburned,  
788 infrequent and frequent mixed-severity wildfire. *Plos one*, **16**, e0232995.
- 789 Brito-Morales, I., Molinos, J.G., Schoeman, D.S., Burrows, M.T., Poloczanska, E.S., Brown, C.J., Ferrier,  
790 S., Harwood, T.D., Klein, C.J., McDonald-Madden, E. & others. (2018). Climate velocity can inform  
791 conservation in a warming world. *Trends in ecology & evolution*, **33**, 441–457.
- 792 Brosi, B.J. & Briggs, H.M. (2013). Single pollinator species losses reduce floral fidelity and plant reproductive  
793 function. *Proceedings of the National Academy of Sciences*, **110**, 13044–13048.
- 794 Calabrese, J.M., Certain, G., Kraan, C. & Dormann, C.F. (2014). Stacking species distribution models and  
795 adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography*, **23**, 99–112.
- 796 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T.L. (2009).  
797 BLAST+: Architecture and applications. *BMC bioinformatics*, **10**, 1–9.
- 798 Cameron, S.A. & Sadd, B.M. (2020). Global trends in bumble bee health. *Annual review of entomology*, **65**,  
799 209–232.
- 800 CaraDonna, P.J., Burkle, L.A., Schwarz, B., Resasco, J., Knight, T.M., Benadi, G., Bluthgen, N., Dormann,  
801 C.F., Fang, Q., Frund, J. & others. (2021). Seeing through the static: The temporal dimension of  
802 plant–animal mutualistic interactions. *Ecology Letters*, **24**, 149–161.
- 803 CaraDonna, P.J., Iler, A.M. & Inouye, D.W. (2014). Shifts in flowering phenology reshape a subalpine plant  
804 community. *Proceedings of the National Academy of Sciences*, **111**, 4916–4921.
- 805 CaraDonna, P.J., Petry, W.K., Brennan, R.M., Cunningham, J.L., Bronstein, J.L., Waser, N.M. & Sanders,  
806 N.J. (2017). Interaction rewiring and the rapid turnover of plant–pollinator networks. *Ecology letters*,  
807 **20**, 385–394.
- 808 Chao, A., Gotelli, N.J., Hsieh, T.C., Sande, E.L., Ma, K.H., Colwell, R.K. & Ellison, A.M. (2014). Rarefac-  
809 tion and extrapolation with hill numbers: A framework for sampling and estimation in species diversity  
810 studies. *Ecological Monographs*, **84**, 45–67.
- 811 Cheng, S., Melkonian, M., Smith, S.A., Brockington, S., Archibald, J.M., Delaux, P.-M., Li, F.-W., Melko-  
812 nian, B., Mavrodiev, E.V., Sun, W., Fu, Y., Yang, H., Soltis, D.E., Graham, S.W., Soltis, P.S., Liu,  
813 X., Xu, X. & Wong, G.K.-S. (2018). 10KP: A phylogenetic genome sequencing plan. *GigaScience*, **7**.  
814 Retrieved from <https://doi.org/10.1093/gigascience/giy013>
- 815 Coissac, E., Hollingsworth, P.M., Lavergne, S. & Taberlet, P. (2016). From barcodes to genomes: Extending  
816 the concept of DNA barcoding.
- 817 Coissac, E., Riaz, T. & Puillandre, N. (2012). Bioinformatic challenges for DNA metabarcoding of plants  
818 and animals. *Molecular ecology*, **21**, 1834–1847.

- 819 Colla, S.R., Gadallah, F., Richardson, L., Wagner, D. & Gall, L. (2012). Assessing declines of north american  
820 bumble bees (*bombus* spp.) Using museum specimens. *Biodiversity and Conservation*, **21**, 3585–3595.
- 821 Cooke, R.U. & Reeves, R.W. (1976). *Arroyos and environmental change in the american south-west*. Claren-  
822 don Press.
- 823 Crisci, J.V., Katinas, L., Apodaca, M.J. & Hoch, P.C. (2020). The end of botany. *Trends in Plant Science*,  
824 **25**, 1173–1176.
- 825 Cronquist, A., Holmgren, A.H., Holmgren, N.H., Reveal, J.L., Holmgren, P.K., Barneby, R & others.  
826 (1977+). *Intermountain flora. Vascular plants of the intermountain west, USA volume six. The mono-*  
827 *cotyledons*. Columbia University.
- 828 Dahl, T.E. (1990). *Wetlands losses in the united states, 1780's to 1980's*. US Department of the Interior,  
829 Fish; Wildlife Service.
- 830 Davis, C.C., Lyra, G.M., Park, D.S., Asprino, R., Maruyama, R., Torquato, D., Cook, B.I. & Ellison, A.M.  
831 (2022). New directions in tropical phenology. *Trends in Ecology & Evolution*.
- 832 Dobrowski, S.Z. & Parks, S.A. (2016). Climate change velocity underestimates climate change exposure in  
833 mountainous regions. *Nature Communications*, **7**, 1–8.
- 834 Doyle, J.J. & Doyle, J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue.  
835 *Phytochemical Bulletin*, **19**, 11–15.
- 836 Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.-P. & Guisan, A. (2011). Predicting spatial  
837 patterns of plant species richness: A comparison of direct macroecological and species stacking modelling  
838 approaches. *Diversity and Distributions*, **17**, 1122–1131.
- 839 Elith\*, J., H. Graham\*, C., P. Anderson, R., Dudik, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann,  
840 F., R. Leathwick, J., Lehmann, A. & others. (2006). Novel methods improve prediction of species'  
841 distributions from occurrence data. *Ecography*, **29**, 129–151.
- 842 Fazekas, A.J., Kesanakurti, P.R., Burgess, K.S., Percy, D.M., Graham, S.W., Barrett, S.C., Newmaster,  
843 S.G., Hajibabaei, M. & Husband, B.C. (2009). Are plant species inherently harder to discriminate than  
844 animal species using DNA barcoding markers? *Molecular Ecology Resources*, **9**, 130–139.
- 845 Fenn, M.E., Baron, J.S., Allen, E.B., Rueth, H.M., Nydick, K.R., Geiser, L., Bowman, W.D., Sickman, J.O.,  
846 Meixner, T., Johnson, D.W. & others. (2003). Ecological effects of nitrogen deposition in the western  
847 united states. *BioScience*, **53**, 404–420.
- 848 Flora of North America Editorial Committee, eds. (1993+). *Flora of north america north of mexico [online]*.  
849 Oxford University Press on Demand.
- 850 Frase, Barbara A. & Buck, P. (2007). Vascular Plants of the Gothic Area. Retrieved from [https://www.digitalrmb.org/wp-content/uploads/2016/05/vascularplantlist\\_20071.pdf](https://www.digitalrmb.org/wp-content/uploads/2016/05/vascularplantlist_20071.pdf)

- 852 Gage, E. & Cooper, D.J. (2013). Historical range of variation assessment for wetland and riparian ecosystems,  
853 u.s. Forest service rocky mountain region
- 854 Goulson, D. (2010). *Bumblebees: Behaviour, ecology, and conservation*. Oxford University Press on Demand.
- 855 Goulson, D., Hanley, M.E., Darvill, B., Ellis, J. & Knight, M.E. (2005). Causes of rarity in bumblebees.  
856 *Biological conservation*, **122**, 1–8.
- 857 Goulson, D., Lye, G. & Darvill, B. (2008). The decline and conservation of bumblebees. *Annual review of  
858 entomology*, **53**, 191–208.
- 859 Grixti, J.C., Wong, L.T., Cameron, S.A. & Favret, C. (2009). Decline of bumble bees (*bombus*) in the north  
860 american midwest. *Biological conservation*, **142**, 75–84.
- 861 Group, C.P.W., Hollingsworth, P.M., Forrest, L.L., Spouge, J.L., Hajibabaei, M., Ratnasingham, S., Bank,  
862 M. van der, Chase, M.W., Cowan, R.S., Erickson, D.L. & others. (2009). A DNA barcode for land  
863 plants. *Proceedings of the National Academy of Sciences*, **106**, 12794–12797.
- 864 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.,  
865 Chen, S.-L. & others. (2011). Comparative analysis of a large dataset indicates that internal transcribed  
866 spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National  
867 Academy of Sciences*, **108**, 19641–19646.
- 868 Havens, K., Vitt, P., Schwarz, J., Orr, B. & Crimmins, T. (2007). Chicago botanic garden's conservation  
869 and outreach efforts on climate change. *BGjournal*, **4**, 13–16.
- 870 Hebert, P.D., Cywinska, A., Ball, S.L. & DeWaard, J.R. (2003). Biological identifications through DNA  
871 barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**, 313–321.
- 872 Hengl, T., Mendes de Jesus, J., Heuvelink, G.B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A.,  
873 Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B. & others. (2017). SoilGrids250m:  
874 Global gridded soil information based on machine learning. *PLoS one*, **12**, e0169748.
- 875 Hennig, C. (2020). *Fpc: Flexible procedures for clustering*. Retrieved from <https://CRAN.R-project.org/>  
876 package=fpc
- 877 Hinchliff, C.E., Smith, S.A., Allman, J.F., Burleigh, J.G., Chaudhary, R., Coghill, L.M., Crandall, K.A.,  
878 Deng, J., Drew, B.T., Gazis, R. & others. (2015). Synthesis of phylogeny and taxonomy into a compre-  
879 hensive tree of life. *Proceedings of the National Academy of Sciences*, **112**, 12764–12769.
- 880 Hitchcock, C.L. & Cronquist, A. (2018). *Flora of the pacific northwest: An illustrated manual*. University  
881 of Washington Press.
- 882 Hollingsworth, P.M., Li, D.-Z., Bank, M. van der & Twyford, A.D. (2016). Telling plant species apart with  
883 DNA: From barcodes to genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*,  
884 **371**, 20150338.

- 885 Hsieh, T.C., Ma, K.H. & Chao, A. (2020). *iNEXT: Interpolation and extrapolation for species diversity*.  
886 Retrieved from [http://chao.stat.nthu.edu.tw/wordpress/software\\_download/](http://chao.stat.nthu.edu.tw/wordpress/software_download/)
- 887 Iler, A.M., Humphrey, P.T., Ogilvie, J.E. & CaraDonna, P.J. (2021). Conceptual and practical issues limit  
888 the utility of statistical estimators of phenological events. *Ecosphere*, **12**, e03828.
- 889 Janzen, D.H. (1967). Synchronization of sexual reproduction of trees within the dry season in central america.  
890 *Evolution*, **21**, 620–637.
- 891 Janzen, D.H., Burns, J.M., Cong, Q., Hallwachs, W., Dapkey, T., Manjunath, R., Hajibabaei, M., Hebert,  
892 P.D. & Grishin, N.V. (2017). Nuclear genomes distinguish cryptic species suggested by their DNA  
893 barcodes and ecology. *Proceedings of the National Academy of Sciences*, **114**, 8313–8318.
- 894 *Jepson flora project*. (2020).
- 895 Johnson, M.D., Fokar, M., Cox, R.D. & Barnes, M.A. (2021). Airborne environmental DNA metabarcoding  
896 detects more diversity, with less sampling effort, than a traditional plant community survey. *BMC Ecology*  
897 and *Evolution*, **21**, 1–15.
- 898 Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J. & Wickett,  
899 N.J. (2016). HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput  
900 sequencing reads using target enrichment. *Applications in plant sciences*, **4**, 1600016.
- 901 Johnson, M.G., Pokorny, L., Dodsworth, S., Botigue, L.R., Cowan, R.S., Devault, A., Eiserhardt, W.L.,  
902 Epitawalage, N., Forest, F., Kim, J.T. & others. (2019). A universal probe set for targeted sequencing  
903 of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic biology*,  
904 **68**, 594–606.
- 905 Joshi, J., Schmid, B., Caldeira, M., Dimitrakopoulos, P., Good, J., Harris, R., Hector, A., Huss-Danell, K.,  
906 Jumpponen, A., Minns, A. & others. (2001). Local adaptation enhances performance of common plant  
907 species. *Ecology Letters*, **4**, 536–544.
- 908 Keane, R.E. (2002). Cascading effects of fire exclusion in rocky mountain ecosystems: A literature review.
- 909 Kearns, C.A., Thomson, J.D. & others. (2001). *Natural history of bumblebees*. University Press of Colorado.
- 910 Kramer, A.T. & Havens, K. (2015). Report in brief: Assessing botanical capacity to address grand challenges  
911 in the united states. *Natural Areas Journal*, **35**, 83–89.
- 912 Kress, W.J. (2017). Plant DNA barcodes: Applications today and in the future. *Journal of systematics and*  
913 *evolution*, **55**, 291–307.
- 914 Kress, W.J. & Erickson, D.L. (2007). A two-locus global DNA barcode for land plants: The coding rbcL  
915 gene complements the non-coding trnH-psbA spacer region. *PLoS one*, **2**, e508.
- 916 Kuhn, M. (2022). *Caret: Classification and regression training*. Retrieved from <https://CRAN.R-project.org/package=caret>

- 918 Lamb, P.D., Hunter, E., Pinnegar, J.K., Creer, S., Davies, R.G. & Taylor, M.I. (2019). How quantitative is  
919 metabarcoding: A meta-analytical approach. *Molecular ecology*, **28**, 420–430.
- 920 Land Management, B. of. (2019). U.s. Department of interior bureau of land management, BLM - assess-  
921 ment, inventory, and monitoring (AIM) terrestrial indicators raw dataset. Retrieved from <https://gbp->  
922 [blm-egis.hub.arcgis.com/pages/aim](https://blm-egis.hub.arcgis.com/pages/aim)
- 923 Lang, D., Tang, M., Hu, J. & Zhou, X. (2019). Genome-skimming provides accurate quantification for pollen  
924 mixtures. *Molecular Ecology Resources*, **19**, 1433–1446.
- 925 Lesica, P., Lavin, M. & Stickney, P.F. (2012). *Manual of montana vascular plants*. Brit Press.
- 926 Lewin, H.A., Richards, S., Aiden, E.L., Allende, M.L., Archibald, J.M., Bálint, M., Barker, K.B., Baumgart-  
927 ner, B., Belov, K., Bertorelle, G., Blaxter, M.L., Cai, J., Caperello, N.D., Carlson, K., Castilla-Rubio,  
928 J.C., Chaw, S.-M., Chen, L., Childers, A.K., Coddington, J.A., Conde, D.A., Corominas, M., Crandall,  
929 K.A., Crawford, A.J., DiPalma, F., Durbin, R., Ebenezer, T.E., Edwards, S.V., Fedrigo, O., Flieck, P.,  
930 Formenti, G., Gibbs, R.A., Gilbert, M.T.P., Goldstein, M.M., Graves, J.M., Greely, H.T., Grigoriev,  
931 I.V., Hackett, K.J., Hall, N., Haussler, D., Helgen, K.M., Hogg, C.J., Isobe, S., Jakobsen, K.S., Janke,  
932 A., Jarvis, E.D., Johnson, W.E., Jones, S.J.M., Karlsson, E.K., Kersey, P.J., Kim, J.-H., Kress, W.J.,  
933 Kuraku, S., Lawniczak, M.K.N., Leebens-Mack, J.H., Li, X., Lindblad-Toh, K., Liu, X., Lopez, J.V.,  
934 Marques-Bonet, T., Mazard, S., Mazet, J.A.K., Mazzoni, C.J., Myers, E.W., O'Neill, R.J., Paez, S.,  
935 Park, H., Robinson, G.E., Roquet, C., Ryder, O.A., Sabir, J.S.M., Shaffer, H.B., Shank, T.M., Sherkow,  
936 J.S., Soltis, P.S., Tang, B., Tedersoo, L., Uliano-Silva, M., Wang, K., Wei, X., Wetzer, R., Wilson,  
937 J.L., Xu, X., Yang, H., Yoder, A.D. & Zhang, G. (2022). The earth BioGenome project 2020: Start-  
938 ing the clock. *Proceedings of the National Academy of Sciences*, **119**, e2115635118. Retrieved from  
939 <https://www.pnas.org/doi/abs/10.1073/pnas.2115635118>
- 940 Liang, H., Zhao, Y.-H., Rafferty, N.E., Ren, Z.-X., Zhong, L., Li, H.-D., Li, D.-Z. & Wang, H. (2021).  
941 Evolutionary and ecological factors structure a plant–bumblebee network in a biodiversity hotspot, the  
942 himalaya–hengduan mountains. *Functional Ecology*, **35**, 2523–2535.
- 943 Life Project Consortium, D.T. of, Blaxter, M., Mieszkowska, N., Palma, F.D., Holland, P., Durbin, R.,  
944 Richards, T., Berriman, M., Kersey, P., Hollingsworth, P., Wilson, W., Twyford, A., Gaya, E., Lawniczak,  
945 M., Lewis, O., Broad, G., Howe, K., Hart, M., Flieck, P. & Barnes, I. (2022). Sequence locally, think glob-  
946 ally: The darwin tree of life project. *Proceedings of the National Academy of Sciences*, **119**, e2115642118.  
947 Retrieved from <https://www.pnas.org/doi/abs/10.1073/pnas.2115642118>
- 948 Li, X., Jiang, L., Meng, F., Wang, S., Niu, H., Iler, A.M., Duan, J., Zhang, Z., Luo, C., Cui, S. & others.  
949 (2016). Responses of sequential and hierarchical phenological events to warming and cooling in alpine  
950 meadows. *Nature Communications*, **7**, 1–8.

- 951 Liu, J., Shi, L., Han, J., Li, G., Lu, H., Hou, J., Zhou, X., Meng, F. & Downie, S.R. (2014). Identification  
952 of species in the angiosperm family apiaceae using DNA barcodes. *Molecular ecology resources*, **14**,  
953 1231–1238.
- 954 Li, X., Yang, Y., Henry, R.J., Rossetto, M., Wang, Y. & Chen, S. (2015). Plant DNA barcoding: From gene  
955 to genome. *Biological Reviews*, **90**, 157–166.
- 956 Loarie, S.R., Duffy, P.B., Hamilton, H., Asner, G.P., Field, C.B. & Ackerly, D.D. (2009). The velocity of  
957 climate change. *Nature*, **462**, 1052–1055.
- 958 Lu, J., Breitwieser, F.P., Thielen, P. & Salzberg, S.L. (2017). Bracken: Estimating species abundance in  
959 metagenomics data. *PeerJ Computer Science*, **3**, e104.
- 960 Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. (2022). *Cluster: Cluster analysis basics  
961 and extensions*. Retrieved from <https://CRAN.R-project.org/package=cluster>
- 962 Mainali, K. & Slud, E. (2022). *CooccurrenceAffinity: Affinity in cooccurrence data*.
- 963 Mainali, K.P., Slud, E., Singer, M.C. & Fagan, W.F. (2022). A better index for analysis of co-occurrence  
964 and similarity. *Science Advances*, **8**, eabj9204.
- 965 Maitner, B. (2022). *BIEN: Tools for accessing the botanical information and ecology network database*.  
966 Retrieved from <https://CRAN.R-project.org/package=BIEN>
- 967 Manzano, S. (2021). Flippant attitudes towards plant identification jeopardize early career botanists. *Trends  
968 in Plant Science*, **26**, 987–988.
- 969 McLay, T.G., Birch, J.L., Gunn, B.F., Ning, W., Tate, J.A., Nauheimer, L., Joyce, E.M., Simpson, L.,  
970 Schmidt-Lebuhn, A.N., William J & others. (2021). New targets acquired: Improving locus recovery  
971 from the Angiosperms353 probe set. *Applications in plant sciences*, **9**.
- 972 Mohlenbrock, R.H. (2002). *Vascular flora of illinois*. SIU Press.
- 973 Moore, A.J. & Bohs, L. (2003). An ITS phylogeny of balsamorhiza and wyethia (asteraceae: heliantheae).  
974 *American Journal of Botany*, **90**, 1653–1660.
- 975 Naiman, R.J., Johnston, C.A. & Kelley, J.C. (1988). Alteration of north american streams by beaver.  
976 *BioScience*, **38**, 753–762.
- 977 Naimi, B. & Araujo, M.B. (2016). Sdm: A reproducible and extensible r platform for species distribution  
978 modelling. *Ecography*, **39**, 368–375.
- 979 Naimi, B., Hamm, N. a.s., Groen, T.A., Skidmore, A.K. & Toxopeus, A.G. (2014). Where is positional  
980 uncertainty a problem for species distribution modelling. *Ecography*, **37**, 191–203.
- 981 Nazaire, M. & Hufford, L. (2014). Phylogenetic systematics of the genus mertensia (boraginaceae). *System-  
982 atic Botany*, **39**, 268–303.
- 983 Nevado, B., Atchison, G.W., Hughes, C.E. & Filatov, D.A. (2016). Widespread adaptive evolution during

- repeated evolutionary radiations in new world lupins. *Nature communications*, **7**, 1–9.
- Newstrom, L.E., Frankie, G.W. & Baker, H.G. (1994). A new classification for plant phenology based on flowering patterns in lowland tropical rain forest trees at la selva, costa rica. *Biotropica*, **26**, 141–159.
- Oberhauser, K.S., Nail, K.R. & Altizer, S. (2015). *Monarchs in a changing world: Biology and conservation of an iconic butterfly*. Cornell University Press.
- Occownload Gbif.Org. (2021). Occurrence download. Retrieved from <https://www.gbif.org/occurrence/download/0206948-200613084148143>
- Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Solymos, P., Stevens, M.H.H., Szoechs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H.B.A., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M.O., Lahti, L., McGlinn, D., Ouellette, M.-H., Ribeiro Cunha, E., Smith, T., Stier, A., Ter Braak, C.J.F. & Weedon, J. (2022). *Vegan: Community ecology package*. Retrieved from <https://CRAN.R-project.org/package=vegan>
- Oliver, P.M., Adams, M., Lee, M.S., Hutchinson, M.N. & Doughty, P. (2009). Cryptic diversity in vertebrates: Molecular data double estimates of species diversity in a radiation of australian lizards (diplodactylus, gekkota). *Proceedings of the Royal Society B: Biological Sciences*, **276**, 2001–2007.
- Omernik, J.M. (1987). Ecoregions of the conterminous united states. *Annals of the Association of American geographers*, **77**, 118–125.
- Ottenlips, M.V., Mansfield, D.H., Buerki, S., Feist, M.A.E., Downie, S.R., Dodsworth, S., Forest, F., Plunkett, G.M. & Smith, J.F. (2021). Resolving species boundaries in a recent radiation with the Angiosperms353 probe set: The lomatium packardiae/l. Anomalum clade of the l. Triternatum (apiaceae) complex. *American journal of botany*, **108**, 1217–1233.
- Park, D.S., Lyra, G.M., Ellison, A.M., Maruyama, R.K.B., Torquato, D. dos R., Asprino, R.C., Cook, B.I. & Davis, C.C. (2022). Herbarium records provide reliable phenology estimates in the understudied tropics. *Journal of Ecology*.
- Pearse, W.D., Davis, C.C., Inouye, D.W., Primack, R.B. & Davies, T.J. (2017). A statistical estimator for determining the limits of contemporary and historic phenology. *Nature Ecology & Evolution*, **1**, 1876–1882.
- Peel, N., Dicks, L.V., Clark, M.D., Heavens, D., Percival-Alwyn, L., Cooper, C., Davies, R.G., Leggett, R.M. & Yu, D.W. (2019). Semi-quantitative characterisation of mixed pollen samples using MinION sequencing and reverse metagenomics (RevMet). *Methods in Ecology and Evolution*, **10**, 1690–1701.
- Pepin, N., Arnone, E., Gobiet, A., Haslinger, K., Kotlarski, S., Notarnicola, C., Palazzi, E., Seibert, P., Serafin, S., Schöner, W. & others. (2022). Climate changes and their elevational patterns in the mountains

- 1017 of the world. *Reviews of geophysics*, **60**, e2020RG000730.
- 1018 Pinto-Ledezma, J.N. & Cavender-Bares, J. (2021). Predicting species distributions and community compo-  
1019 sition using satellite remote sensing predictors. *Scientific Reports*, **11**, 1–12.
- 1020 Poron, A., Andalo, C., Burrus, M. & Escaravage, N. (2017). DNA metabarcoding data unveils invisible  
1021 pollination networks. *Scientific Reports*, **7**, 1–11.
- 1022 Prather, L.A., Alvarez-Fuentes, O., Mayfield, M.H. & Ferguson, C.J. (2004a). Implications of the decline in  
1023 plant collecting for systematic and floristic research. *Systematic Botany*, **29**, 216–220.
- 1024 Prather, L.A., Alvarez-Fuentes, O., Mayfield, M.H. & Ferguson, C.J. (2004b). The decline of plant collecting  
1025 in the united states: A threat to the infrastructure of biodiversity studies. *Systematic Botany*, **29**, 15–28.
- 1026 Prim, R.C. (1957). Shortest connection networks and some generalisations. *Bell System Technical Journal*,  
1027 **36**, 1389–1401.
- 1028 Pusalkar, P.K. & Singh, D.K. (2015). Taxonomic rearrangement of arenaria (caryophyllaceae) in indian  
1029 western himalaya. *Journal of Japanese Botany*, **90**, 77–91.
- 1030 Pyke, G.H. (1982). Local geographic distributions of bumblebees near crested butte, colorado: Competition  
1031 and community structure. *Ecology*, **63**, 555–573.
- 1032 Qiao, H., Soberon, J. & Peterson, A.T. (2015). No silver bullets in correlative ecological niche modelling:  
1033 Insights from testing among many potential algorithms for niche estimation. *Methods in Ecology and  
1034 Evolution*, **6**, 1126–1136.
- 1035 Rabeler, R.K. & Wagner, W.L. (2016). New combinations in odontostemma (caryophyllaceae). *PhytoKeys*,  
1036 77.
- 1037 Ralphs, M.H. & Ueckert, D.N. (1988). Herbicide control of locoweeds: A review. *Weed Technology*, **2**,  
1038 460–465.
- 1039 Ralphs, M., Woolsey, L. & Bowns, J. (2003). Mechanism by which ammonium fertilizers kill tall larkspur.  
1040 *Rangeland Ecology & Management/Journal of Range Management Archives*, **56**, 524–528.
- 1041 Robinson, N., Regetz, J. & Guralnick, R.P. (2014). EarthEnv-DEM90: A nearly-global, void-free, multi-  
1042 scale smoothed, 90m digital elevation model from fused ASTER and SRTM data. *ISPRS Journal of  
1043 Photogrammetry and Remote Sensing*, **87**, 57–67.
- 1044 Ruppert, K.M., Kline, R.J. & Rahman, M.S. (2019). Past, present, and future perspectives of environmental  
1045 DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global  
1046 eDNA. *Global Ecology and Conservation*, **17**, e00547.
- 1047 Sadeghian, S., Zarre, S., Rabeler, R.K. & Heubl, G. (2015). Molecular phylogenetic analysis of arenaria  
1048 (caryophyllaceae: Tribe arenarieae) and its allies inferred from nuclear DNA internal transcribed spacer  
1049 and plastid DNA rps16 sequences. *Botanical Journal of the Linnean Society*, **178**, 648–669.

- 1050 Sarro, E., Tripodi, A. & Woodard, S.H. (2022). Bumble bee (*bombus vosnesenskii*) queen nest searching  
1051 occurs independent of ovary developmental status. *Integrative Organismal Biology*, **4**, obac007.
- 1052 Schmitt, S., Pouteau, R., Justeau, D., De Boissieu, F. & Birnbaum, P. (2017). Ssdm: An r package to  
1053 predict distribution of species richness and composition based on stacked species distribution models.  
1054 *Methods in Ecology and Evolution*, **8**, 1795–1803.
- 1055 Sennikov, A.N. & Kurtto, A. (2017). A phylogenetic checklist of sorbus sl (rosaceae) in europe.
- 1056 Sickel, W., Ankenbrand, M.J., Grimmer, G., Holzschuh, A., Hartel, S., Lanzen, J., Steffan-Dewenter, I. &  
1057 Keller, A. (2015). Increased efficiency in identifying mixed pollen samples by meta-barcoding with a  
1058 dual-indexing approach. *BMC ecology*, **15**, 1–9.
- 1059 Slimp, M., Williams, L.D., Hale, H. & Johnson, M.G. (2021). On the potential of Angiosperms353 for  
1060 population genomic studies. *Applications in Plant Sciences*, **9**.
- 1061 Smith, S.A. & Brown, J.W. (2018). Constructing a broadly inclusive seed plant phylogeny. *American journal*  
1062 *of botany*, **105**, 302–314.
- 1063 Sork, V.L. (2018). Genomic studies of local adaptation in natural plant populations. *Journal of Heredity*,  
1064 **109**, 3–15.
- 1065 Stevens, C.J., David, T.I. & Storkey, J. (2018). Atmospheric nitrogen deposition in terrestrial ecosystems: Its  
1066 impact on plant communities and consequences across trophic levels. *Functional Ecology*, **32**, 1757–1769.
- 1067 Stroud, S., Fennell, M., Mitchley, J., Lydon, S., Peacock, J. & Bacon, K.L. (2022). The botanical education  
1068 extinction and the fall of plant awareness. *Ecology and Evolution*, **12**, e9019.
- 1069 Suchan, T., Talavera, G., Saez, L., Ronikier, M. & Vila, R. (2019). Pollen metabarcoding as a tool for  
1070 tracking long-distance insect migrations. *Molecular Ecology Resources*, **19**, 149–162.
- 1071 Tange, O. (2021). GNU parallel 20220322 (savannah). Retrieved from <https://doi.org/10.5281/zenodo.6377950>
- 1072
- 1073 Thompson, J.N. (1994). *The coevolutionary process*. University of Chicago press.
- 1074 Tran, H., Nguyen, P., Ombadi, M., Hsu, K., Sorooshian, S. & Qing, X. (2019). A cloud-free MODIS snow  
1075 cover dataset for the contiguous united states from 2000 to 2017. *Scientific data*, **6**, 1–13.
- 1076 Wang, T., Hamann, A., Spittlehouse, D. & Carroll, C. (2016). Locally downscaled and spatially customizable  
1077 climate data for historical and future periods for north america. *PloS one*, **11**, e0156720.
- 1078 Weber, W. (1998). New names and combinations in asteraceae: Heliantheae-ecliptinae. *Phytologia*, **85**,  
1079 19–21.
- 1080 Wenzell, K.E., McDonnell, A.J., Wickett, N.J., Fant, J.B. & Skogen, K.A. (2021). Incomplete reproductive  
1081 isolation and low genetic differentiation despite floral divergence across varying geographic scales in  
1082 *castilleja*. *American Journal of Botany*, **108**, 1270–1288.

- 1083 Wilkinson, D.P., Golding, N., Guillera-Arroita, G., Tingley, R. & McCarthy, M.A. (2021). Defining and  
1084 evaluating predictions of joint species distribution models. *Methods in Ecology and Evolution*, **12**, 394–  
1085 404.
- 1086 Williams, P.H. (1982). The distribution and decline of british bumble bees (*bombus latr.*). *Journal of*  
1087 *Apicultural Research*, **21**, 236–245. Retrieved from <https://doi.org/10.1080/00218839.1982.11100549>
- 1088 Wilson, A.M. & Jetz, W. (2016). Remotely sensed high-resolution global cloud dynamics for predicting  
1089 ecosystem and biodiversity distributions. *PLoS biology*, **14**, e1002415.
- 1090 Wood, D.E., Lu, J. & Langmead, B. (2019). Improved metagenomic analysis with kraken 2. *Genome biology*,  
1091 **20**, 1–13.
- 1092 Xie, Y., Wang, X. & Silander Jr, J.A. (2015). Deciduous forest responses to temperature, precipitation, and  
1093 drought imply complex climate change impacts. *Proceedings of the National Academy of Sciences*, **112**,  
1094 13585–13590.
- 1095 Xie, Y., Wang, X., Wilson, A.M. & Silander Jr, J.A. (2018). Predicting autumn phenology: How deciduous  
1096 tree species respond to weather stressors. *Agricultural and Forest Meteorology*, **250**, 127–137.
- 1097 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,  
1098 D.-Z. & Wang, H. (2019). The topological differences between visitation and pollen transport networks:  
1099 A comparison in species rich communities of the himalaya–hengduan mountains. *Oikos*, **128**, 551–562.  
1100 Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1111/oik.05262>

1101 **List of Figures**

1102      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. . . . .	66
1109      2	Number of the ten most commonly visited plants which are also in the top ten most common sequences . . . . .	67
1111      3	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. . . . .	68
1116      4	Modelled dates of when major flowering events occurred compared between long term and modelled data . . . . .	69
1118      5	Modelled dates of when major flowering events occurred compared between 2015 and modelled data . . . . .	70
1120      6	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 . . . . .	71

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

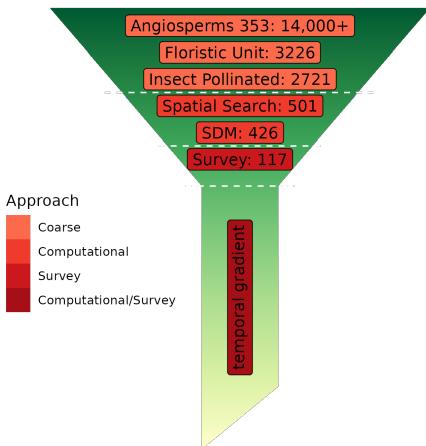


Figure 1: Simplified Conceptual Diagram of three approaches leading towards classification of sequencing results, and the number of species associated with them in our area. The upper three boxes indicate a common coarse approach, assuming one has a digitized Flora, which is not always the case. The 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.

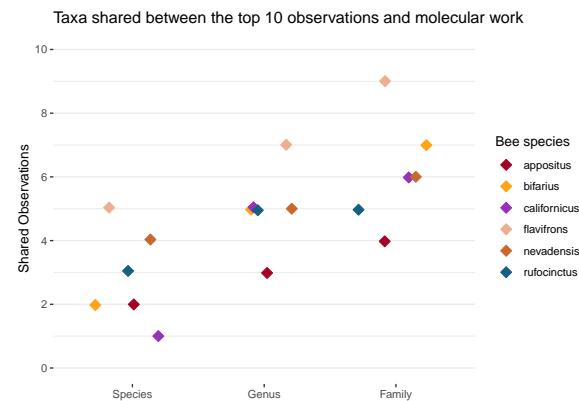
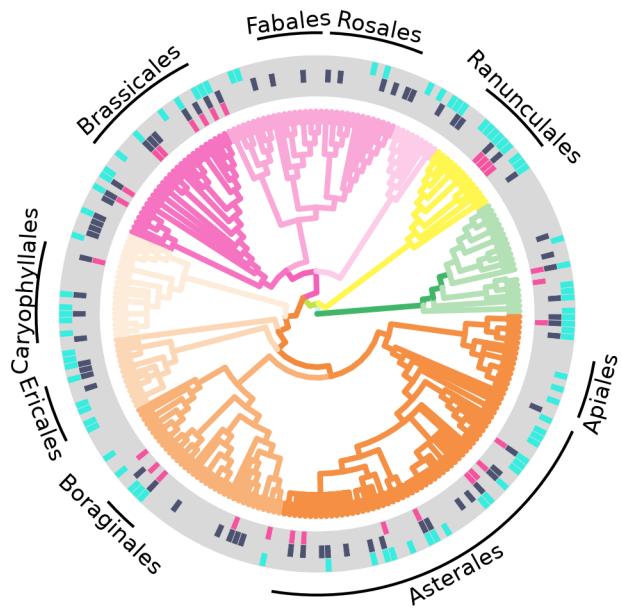


Figure 2: Number of the ten most commonly visited plants which are also in the top ten most common sequences

## Biotically pollinated plant genera with morphological or molecular data



Status     lacking     observed     sequenced     slide

Figure 3: 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.

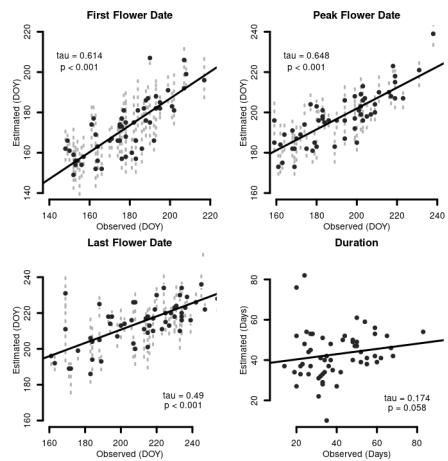


Figure 4: Modelled dates of when major flowering events occurred compared between long term and modelled data

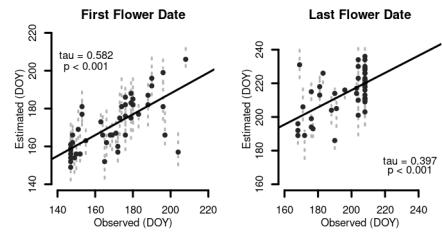


Figure 5: Modelled dates of when major flowering events occurred compared between 2015 and modelled data

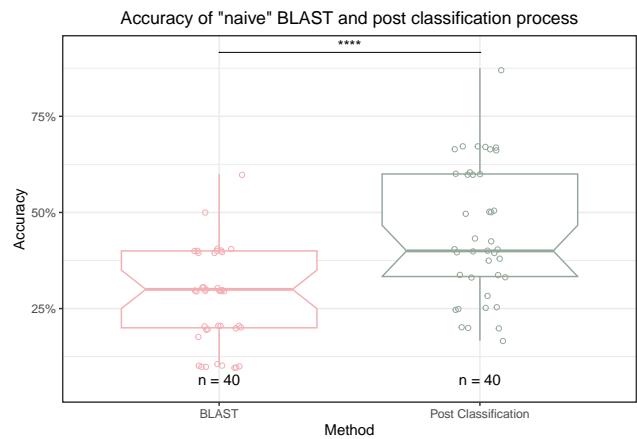


Figure 6: Comparision of Accuracy between the initial output data from BLAST, and these same data subjected to the post-classification process which removes surrogate, and temporally restricted species

<sub>1123</sub> **List of Tables**

<sub>1124</sub> 2	Applications of Plant Metabarcoding . . . . .	73
<sub>1125</sub> 3	Current Issues Facing Plant Metagenomics . . . . .	74
<sub>1126</sub> 4	Logistic regression assessing accuracy of SDMs . . . . .	75
<sub>1127</sub> 5	Species Distribution Modeling evaluation contingency table . . . . .	76
<sub>1128</sub> 6	Post classification of Sequences via Taxonomy and Ecology, top 15 most abundant reads . . .	77
<sub>1129</sub> 7	Limitations Faced and Possible Solutions . . . . .	78

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: 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 5: Species Distribution Modeling evaluation contingency table

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

Table 6: 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 7: 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 Genome Sizes