

## Ecogeography and utility to plant breeding of the crop wild relatives of sunflower (*Helianthus annuus* L.)

Michael B. Kantar<sup>1,2\*</sup>, Chrystian C. Sosa<sup>3</sup>, Colin K. Khoury<sup>3,4</sup>, Nora P. Castañeda-Álvarez<sup>3</sup>, Harold A. Achicanoy<sup>3</sup>, Vivian Bernau<sup>5</sup>, Nolan Kane<sup>6</sup>, Laura Marek<sup>7</sup>, Gerald Seiler<sup>8</sup>, Loren Rieseberg<sup>2,9</sup>

<sup>1</sup>Agronomy and Plant Genetics, University of Minnesota, USA, <sup>2</sup>Biodiversity Research Centre and Department of Botany, University of British Columbia, Canada, <sup>3</sup>DAPA, International Center for Tropical Agriculture (CIAT), Colombia, <sup>4</sup>Centre for Crop Systems Analysis, Wageningen University, Netherlands, <sup>5</sup>Department of Horticulture and Crop Science, The Ohio State University, USA, <sup>6</sup>Department of Ecology and Evolutionary Biology, University of Colorado at Boulder, USA, <sup>7</sup>North Central Regional Plant Introduction Station, Agronomy Department, Iowa State University and USDA-ARS, USA, <sup>8</sup>Northern Crop Science Laboratory, USDA-ARS, USA, <sup>9</sup>Department of Biology, Indiana University, USA

*Submitted to Journal:*  
Frontiers in Plant Science

*Specialty Section:*  
Crop Science and Horticulture

*ISSN:*  
1664-462X

*Article type:*  
Original Research Article

*Received on:*  
01 Aug 2015

*Accepted on:*  
24 Sep 2015

*Provisional PDF published on:*  
24 Sep 2015

*Frontiers website link:*  
[www.frontiersin.org](http://www.frontiersin.org)

*Citation:*  
Kantar MB, Sosa CC, Khoury CK, Castañeda-álvarez NP, Achicanoy HA, Bernau V, Kane N, Marek L, Seiler G and Rieseberg L (2015) Ecogeography and utility to plant breeding of the crop wild relatives of sunflower (*Helianthus annuus* L.). *Front. Plant Sci.* 6:841. doi:10.3389/fpls.2015.00841

*Copyright statement:*  
© 2015 Kantar, Sosa, Khoury, Castañeda-álvarez, Achicanoy, Bernau, Kane, Marek, Seiler and Rieseberg. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](http://creativecommons.org/licenses/by/4.0/). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

This Provisional PDF corresponds to the article as it appeared upon acceptance, after peer-review. Fully formatted PDF and full text (HTML) versions will be made available soon.

Provisional

1 **Ecogeography and utility to plant breeding of the crop wild relatives of sunflower**  
2 **(*Helianthus annuus* L.)**

3 Michael B. Kantar<sup>1,2,11</sup>, Chrystian C. Sosa<sup>3</sup>, Colin K. Khoury<sup>3,4</sup>, Nora P. Castañeda-Álvarez<sup>3,5</sup>,  
4 Harold A. Achicanoy<sup>3</sup>, Vivian Bernau<sup>3,6</sup>, Nolan C. Kane<sup>7</sup>, Laura Marek<sup>8</sup>, Gerald Seiler<sup>9</sup>, Loren  
5 H. Rieseberg<sup>1,10</sup>

6 <sup>1</sup>Biodiversity Research Centre and Department of Botany, University of British Columbia, 3529-  
7 6270 University Boulevard, Vancouver, British Columbia V6T 1Z4, Canada

8 <sup>2</sup>Department of Agronomy and Plant Genetics, University of Minnesota, 411 Borlaug Hall, 1991  
9 Upper Buford Circle, St. Paul, MN 55108

10 <sup>3</sup>International Center for Tropical Agriculture (CIAT), Km 17, Recta Cali-Palmira, Apartado  
11 Aéreo 6713, Cali, Colombia

12 <sup>4</sup>Centre for Crop Systems Analysis, Wageningen University, Droevendaalsesteeg 1, 6708 PB  
13 Wageningen, Netherlands

14 <sup>5</sup>School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

15 <sup>6</sup>Department of Horticulture and Crop Science, The Ohio State University, 202 Kottman Hall,  
16 2021 Coffey Rd, Columbus, Ohio 43210, USA.

17 <sup>7</sup>Department of Ecology and Evolutionary Biology, University of Colorado at Boulder, Boulder,  
18 CO, USA

19 <sup>8</sup> North Central Regional Plant Introduction Station, Agronomy Department, Iowa State  
20 University and USDA-ARS, Ames, IA 50014, USA

21 <sup>9</sup>USDA-ARS, Northern Crop Science Laboratory, Fargo, ND 58102, USA

22 <sup>10</sup>Department of Biology, Indiana University, Bloomington, IN 47405, USA

23 <sup>11</sup>To whom correspondence should be addressed

24 **Correspondence:**

25 Dr. Michael B. Kantar

26 Department of Agronomy and Plant Genetics

27 University of Minnesota

28 411 Borlaug Hall

29 1991 Upper Buford Circle

30 St. Paul, MN 55108

31 kant0063@umn.edu

32 **Abstract**

33 Crop wild relatives (CWR) are a rich source of genetic diversity for crop improvement.  
34 Combining ecogeographic and phylogenetic techniques can inform both conservation and  
35 breeding. Geographic occurrence, bioclimatic, and biophysical data were used to predict species  
36 distributions, range overlap and niche occupancy in 36 taxa closely related to sunflower  
37 (*Helianthus annuus* L.). Taxa lacking comprehensive *ex situ* conservation were identified. The  
38 predicted distributions for 36 *Helianthus* taxa identified substantial range overlap and asymmetry  
39 and niche conservatism. Specific taxa (e.g., *Helianthus debilis* Nutt., *Helianthus anomalus* Blake,  
40 and *Helianthus divaricatus* L.) were identified as targets for traits of interest, particularly for  
41 abiotic stress tolerance and adaptation to extreme soil properties. The combination of techniques  
42 demonstrates the potential for publicly available ecogeographic and phylogenetic data to  
43 facilitate the identification of possible sources of abiotic stress traits for plant breeding programs.  
44 Much of the primary genepool (wild *H. annuus*) occurs in extreme environments indicating that  
45 introgression of targeted traits may be relatively straightforward. Sister taxa in *Helianthus* have  
46 greater range overlap than more distantly related taxa within the genus. This adds to a growing  
47 body of literature suggesting that in plants (unlike some animal groups), geographic isolation  
48 may not be necessary for speciation.

49 **Key words:** conservation, climate change, crop wild relatives, ecological niche modeling, plant  
50 breeding, plant genetic resources, publicly available data sources

## 51 **Introduction**

52 Plant genetic resources represent the biological foundation for maintaining and improving  
53 crop productivity having played a central role in crop development from antiquity (Porter *et al.*,  
54 2014). Crop wild relatives (CWR) are an important source of useful traits for plant breeding  
55 (Hoisington *et al.*, 1999; Hajjar & Hodgkin, 2007). With the world's population projected to  
56 increase the need to produce more food while using fewer natural resource inputs under  
57 increasingly stochastic climatic conditions is a major challenge (Challinor *et al.*, 2014; Butler &  
58 Huybers, 2013). CWR conservation and utilization focusing on the use of improving  
59 technologies (high throughput phenotyping, genotyping, and geographical information systems),  
60 has been proposed as a way to acquire a greater knowledge of conservation needs and lead to  
61 more targeted use of CWR germplasm (Khoury *et al.*, 2010; Cabrera-Bosquet *et al.*, 2012;  
62 McCouch *et al.*, 2013). Targeted collecting for *ex situ* conservation has become a priority as  
63 rapid changes in both climate and land use patterns increasingly threaten CWR in their natural  
64 habitats (Jarvis *et al.*, 2008; McCouch *et al.*, 2013).

65 Crop wild relatives have traditionally been categorized based on crossing relationships  
66 with domesticates; the primary germplasm contains no crossing barriers, the secondary contains  
67 some meiotic abnormalities, and the tertiary requires special techniques such as embryo rescue  
68 (Harlan & De Wet, 1971; Harlan, 1976). Such classifications may be supplemented by  
69 molecular, bioclimatic and biophysical data to aid in the identification of candidate taxa for  
70 breeding, although such efforts have been constrained by challenges in comprehensively  
71 generating and integrating these data (Ricklefs & Jenkins, 2011).

72 The genus *Helianthus* L. contains 52 species comprising 67 taxa (Schilling, 2006;  
73 Stebbins *et al.*, 2013). Native to North America, the taxa occupy a variety of habitats ranging

74 from open plains to salt marshes (Kane *et al.*, 2013; Seiler & Marek, 2011). Sunflower  
75 (*Helianthus annuus* L.) is the most economically important species from the genus, with ~26  
76 million hectares in production worldwide and a substantial private sector breeding effort,  
77 particularly for oil production (FAOSTAT, 2013). Domesticated approximately 4000 years ago  
78 in east central North America, sunflower has a typical domestication syndrome; i.e., it does not  
79 branch, does not have seed dormancy, has a predictable flowering time, and does not shatter  
80 (Harlan *et al.*, 1973; Harter *et al.*, 2004; Blackman *et al.* 2011). The crop has undergone both  
81 selection and genetic drift during domestication and improvement, which has reduced genetic  
82 diversity (Liu & Burke 2006; Tang & Knapp 2003), with modern cultivars retaining 50-67% of  
83 the diversity present in wild *H. annuus* populations (Kolkman *et al.* 2007; Mandel *et al.*, 2011).

84 Sunflower has often utilized CWR in breeding efforts, with many of the taxa hybridizing  
85 well with the crop (Table S1; Table 1) (Long *et al.*, 1960; Chandler *et al.*, 1986). Despite the  
86 historical use, CWR of sunflower are considered to be relatively untapped, particularly in regard  
87 to adaptation to abiotic stresses. To contribute to an enhanced understanding of the CWR of  
88 sunflower, this studies' objectives were to 1) create geographical distribution models for 36  
89 CWR taxa, and 2) explore niche habitation through comparisons of ecogeographic and  
90 phylogenetic data, to identify taxa occurring in extreme environments of potential interest to  
91 sunflower breeding.

## 92 **Materials and Methods**

### 93 ***Species distribution modeling***

94 A modified gap analysis (Ramírez-Villegas *et al.*, 2010) was used to determine the conservation  
95 status of 36 taxa within *Helianthus* selected based upon their potential to provide useful traits for  
96 sunflower breeding. Briefly, 1) target taxa were identified, and geographic occurrence data were

97 gathered and verified, 2) the overall representation of CWR in germplasm collections was  
98 estimated, 3) potential distribution models were produced for taxa with sufficient samples with  
99 coordinates, 4) the geographic and ecological representation of germplasm collections were  
100 assessed for each taxon by comparing potential distribution models to existing germplasm  
101 collection locations, 5) taxa were prioritized for further collecting based upon the average of  
102 their overall, geographic, and ecological coverage results, and 6) gap analysis results were  
103 correlated with the subjective assessments of collection priorities from crop experts.

104         The selection of taxa for analysis was based on membership within the primary or  
105 secondary gene pools of sunflower (Vincent *et al.*, 2013) with the addition of all taxa from the  
106 tertiary gene pool indicated in publications to be confirmed or potential trait donors (Table S1). A  
107 total of 12,737 occurrence records for the 36 taxa, sourced from 31 herbaria and five genebanks,  
108 were used for distribution models and conservation analysis (Table S2), including 4,705 records  
109 with geographic coordinates. The overall representation of taxa in genebank collections was  
110 estimated using the ‘Sampling Representativeness Score’ (SRS), calculated as the number of  
111 germplasm samples (GS) divided by the total number of samples (GS plus reference records).  
112 After eliminating duplicate records, potential distributions were calculated using Maxent  
113 (Phillips *et al.*, 2006), with a k-5 cross-validation option and 10,000 background points for model  
114 training over North America (Phillips, 2008; VanDerWal *et al.*, 2009). We included nineteen  
115 bioclimatic variables derived from the WorldClim database (Nix, 1986; Hijmans *et al.*, 2005a;  
116 Hijmans *et al.*, 2005b), seven biophysical variables from the ISRIC – World Soil Information  
117 database (<http://soilgrids1km.isric.org>) at a resolution of 2.5 arc-minutes, and the occurrence  
118 information (coordinates) for each taxon as inputs (Table S3). For edaphic data we calculated a  
119 weighted mean from five depths (0 to 5 cm, 5 to 15 cm, 15 to 30 cm, 30 to 60 cm, 60 to 100 cm) to

120 generate a single value for the first meter of soil for each layer, and then resampled the data from  
121 1 arc minutes to 2.5 arc minutes resolution to match the WorldClim dataset, using the raster  
122 package in R and ArcGIS Desktop 10.1 (Hengl *et al.*, 2014). Distributions were further restricted  
123 by applying a taxon independent threshold, based on the Receiver Operating Characteristic  
124 (ROC) curve (Liu *et al.*, 2005). GRIN distribution data was used to ensure that taxa distributions  
125 were not overinflated beyond known native boundaries (GRIN, 2012). Soil cover data from  
126 GlobCover 2009 (Global Land Cover Map) (<http://due.esrin.esa.int/globcover/>) further refined  
127 the maxent outputs and collecting maps by excluding urban areas, water bodies, bare areas, and  
128 permanent snow and ice regions.

129 Potential distribution models were considered accurate if they complied with the  
130 following conditions: i) 5-fold average area under the test ROC curve (ATAUC) is greater than  
131 0.7, ii) the standard deviation of ATAUC (STAUC) is less than 0.15, and iii) At least 10% of  
132 grids for each model has standard deviation less than 0.15 (ASD15). For taxa whose Maxent  
133 model did not comply, potential distributions were estimated by forming a circular buffer of 50  
134 km around each occurrence point for each species.

135 Geographic representativeness of taxa in genebank collections was calculated using the  
136 ‘Geographic Representativeness Score’ (GRS), comparing the spatial overlap of a circular buffer  
137 surrounding each accession record (50 Km radius as described in Hijmans *et al.*, 2001) against  
138 the potential distribution of the taxon. Ecological gaps in genebank collections were calculated  
139 using the ‘Ecological Representativeness Score’ (ERS), calculated by comparing records to the  
140 full environmental range of the modeled taxon across ecosystem types (Olson *et al.*, 2001). The  
141 overall priority for further collecting for *ex situ* conservation for each taxon was determined by  
142 averaging the SRS, GRS, and ERS with equal weight to obtain a final prioritization score (FPS),



143 classified according to the following ranges: 1., high priority (FPS between 0 and 3); 2., medium  
144 priority (FPS between 3.01 and 5); 3., low priority (FPS between 5.01 and 7.5); and 4., and well  
145 conserved taxa (FPS between 7.51 and 10).

#### 146 ***Expert evaluation of conservation assessment results***

147 Predicted taxon distributions based on genebank and herbarium records were compared to the  
148 knowledge of four crop experts with experience with *Helianthus* distributions, systematics,  
149 conservation and diversity. *Helianthus* experts were asked to evaluate of the adequacy of  
150 germplasm collections per species based on their knowledge of total accessions conserved,  
151 geographic and environmental gaps. This assessment was given an expert priority score (EPS),  
152 analogous to the FPS score. A second score was generated, the contextual EPS, which based on  
153 additional knowledge such as *in situ* threats and utility to crop breeding. After initial evaluation  
154 the experts were asked to review the quantitative results, occurrence data, potential distribution  
155 models, and maps of collecting priorities. Following expert input, occurrence data were refined  
156 through elimination of incorrect points and adjustment native areas. Potential distribution  
157 modeling and gap analyses were then conducted using refined datasets to create more accurate  
158 species distribution maps. Potential zones for collecting were identified for each high priority  
159 taxon, and then combined to create maps depicting areas where multiple taxa of high priority for  
160 conservation could be collected.

#### 161 ***Ecogeographic niche overlap and phylogenetic analyses***

162 Potential distribution probability outputs were used when Maxent models performed well  
163 and CA50 sample buffers when Maxent models did not pass the validation criteria, to calculate  
164 niche overlap based on Schoener's D and Hellinger's I as outlined in Warren *et al.* 2008, and  
165 implemented in the R package Phyloclim (Heibl, 2011). Both indices utilize probability

166 distributions in geographic space, with statistics ranging from 0 (no niche overlap) to 1  
167 (complete niche overlap). First pairwise niche overlap was examined, then niche overlap  
168 between allopatric/sympatric taxa separately, annual/perennial taxa separately, and lastly  
169 allopatric/sympatric sister taxa. Geographic range overlap for all pairwise combinations (630  
170 comparisons) was calculated in two ways, with respect to the larger range  $[(2 * \text{number of shared}$   
171  $\text{grid cells}) / (\text{number of grid cells in taxa A} + \text{number of grid cells in taxa B})]$  and with respect to  
172 the smaller range  $[(2 * \text{number of shared grid cells}) / (\text{Total number of grid cells in taxa A} + \text{Total}$   
173  $\text{number of grid cells in taxa B})] / (\text{Total potential number of shared grid cells}) [2 * \text{total number of}$   
174  $\text{grid cells in species with the smaller range}] / (\text{Total number of species A} + \text{Total number of}$   
175  $\text{species B})]$ .

176 Principal component analyses (PCA) were used to assess the importance of  
177 ecogeographic variables (Table S3) to variation in occurrence data of distribution models per  
178 taxon. A hierarchical cluster of principal components (HCPC) identified climatic clusters using  
179 R package FactoMineR (Husson *et al.*, 2014). Boxplots for each bioclimatic and biophysical  
180 layer were created based on occurrence data points (Fig. S1). Ecogeographic variables for  
181 cultivated sunflower were extracted from the area of species distribution maps (Monfreda *et al.*,  
182 2008) at a resolution of 5 arc-minutes, with a random sample of 1,000 points weighted by  
183 harvested area taken from major production regions.

184 We downloaded the publically available 18S-26S Ribosomal DNA sequence from the  
185 external transcribed spacer (ETS) from GenBank (NCBI-<http://www.ncbi.nlm.nih.gov/>) for 28 of  
186 the 36 *Helianthus* taxa, aligned the sequences using ClustalW, and constructed a maximum  
187 likelihood phylogeny with 1000 bootstrap replications, using MEGA6 with a Jukes-Cantor  
188 nucleotide substitution model (Tamura *et al.*, 2013). We performed a Mantel test in R utilizing

189 the ade4 package to explore the relationship between geography and genetics (Dray & Dufour,  
190 2007). We estimated phylogenetic signal of individual ecogeographic traits utilizing Blomberg's  
191 K (Blomberg, et al, 2003), using the multiphylosignal command with 1000 permutations in  
192 Picante (Kembel *et al.*, 2010).

## 193 **Results**

### 194 ***Geographic distributions of sunflower crop wild relatives***

195 Predicted distribution maps were produced for 36 *Helianthus* taxa, along with taxon  
196 richness and collecting hotspot maps (Fig. 2; Fig. S2). Thirty of the 36 taxa (83%) produced  
197 valid maxent models with utilization of soil pH and percent sand greatly improving the accuracy  
198 of distribution models, as assessed by expert opinion (Fig. 3). Five hotspots (areas of high taxon-  
199 level diversity) were identified in the USA, including the southeastern gulf coast, the south-  
200 central, the midwest, the north central, and the central east coast (Fig. 2a). Our results suggest  
201 that half of the 36 taxa are in urgent need of further collecting (high priority species – HPS),  
202 along with 28% in moderate need (medium priority species – MPS), 6% of low priority (LPS),  
203 and 17% that are well represented in existing germplasm collections and thus do not require  
204 urgent additional collecting (Table 1). While the primary gene pool taxa has been well collected,  
205 only 10% of the taxa in the secondary gene pool are well represented across their geographic,  
206 climatic, and edaphic ranges. Likewise, only 7% of taxa in the tertiary gene pool were assessed as  
207 well-conserved (Fig. 1; Table 1). These results contrasted with those of expert reviewers, who  
208 classified more species as LPS. The discrepancy between the results and expert opinion was due  
209 in part to overly optimistic distribution models regarding likelihood of occurrence, in comparison  
210 to expert realities of existence of populations in these regions. Additionally, experts assessed  
211 some taxa, such as *Helianthus debilis* ssp. *cucumerifolius*, at lower priority because distributions

212 have expanded recently as weedy populations invade new areas, and such regions were not  
213 considered by the experts as of particular priority.

#### 214 *Ecological niches of sunflower crop wild relatives*

215 Three ecogeographic clusters differentiate the taxa, with the first three PCs accounted for  
216 74.3% of the variation (Fig. 3b; Table S4). Clusters broadly corresponded to plain, desert, and  
217 woodland ecosystems (Table 1). Cluster one was mostly composed of the secondary germplasm  
218 and differentiated by temperature, while cluster two was mostly the tertiary germplasm and  
219 differentiated by precipitation. Cluster three was differentiated by soil and was evenly split  
220 between the secondary and tertiary germplasm (Table S3). It is important to note that PCA can  
221 increase type one error, so ecological niches must be carefully examined and validated (Revel,  
222 2009; Uyeda et al., 2015). Schoener's D and Hellinger's I identified substantial niche overlap  
223 with few taxa showing niche divergence (Fig. 3; Table 1).

224 Potential geographic distributions of crop wild relative taxa were examined for overlap  
225 with wild *H. annuus* (Fig. S1); most (81%) taxa exhibited some geographic range overlap with  
226 *H. annuus* (Table 1). Among CWR taxa, 39% of pairwise comparisons had overlapping  
227 geographic distributions (sympatry), while 61% were allopatric (Table S5; Fig. S3). Eight of the  
228 twelve sister taxa pairs among the CWR showed some level of sympatry (Table S6). There was  
229 considerable range asymmetry between taxa (Fig. S1), with the amount of overlap depending on  
230 the direction of the comparison, where the smaller range showed 26% more overlap on average  
231 than the larger range (Table S5).

232 There was general niche conservatism even for sister-taxa (Fig. 3; Table 2). While  
233 ecogeographic niches were fairly similar for many variables, occasionally there was substantial  
234 divergence (Fig. 4; Fig. S1). Phylogenetic niche conservatism was found in ~54% of variables

235 (Fig. 5). Divergence was found in several soil variables suggesting an important role of soil in  
236 *Helianthus* diversification. A Mantel's test using Mahalanobis distance ( $r=0.1423$ ,  $p=0.01$ ),  
237 indicated that taxa that are geographically close are generally more closely related genetically.  
238 Notable exceptions to this were *H. maximilliani*, *H. grosseserratus*, and *H. giganteus*, which are  
239 sympatric with *H. annuus*, but are distantly related.

## 240 **Discussion**

241 There has been increased effort to digitize data related to plant species in general and  
242 CWR in particular. The public databases (GBIF, ISRIC, WorldClim, National Germplasm  
243 repositories, DivSeek) that archive these data are an increasingly important tool to  
244 conservationists, evolutionary biologists and plant breeders. Utilizing public data can reduce the  
245 research costs in terms of people hours and consumables to achieve desired environmental and  
246 food production goals. Exploring public databases can provide a targeted way to identify  
247 accessions for introgression that can then be used to validate predicted extreme variation. This  
248 may be a way to more quickly utilize germplasm collections and provide a link to international  
249 initiatives aimed at facilitating more use of plant genetic resources ([www.DivSeek.org](http://www.DivSeek.org)). Here we  
250 have used geographic occurrence, bioclimatic, and biophysical data to predict species  
251 distributions, range overlap and niche occupancy in 36 *Helianthus* taxa that are cross-compatible  
252 with cultivated sunflower and thus likely to be useful in crop breeding. As discussed briefly  
253 below, our results not only have implications for conservation genetics and breeding in  
254 *Helianthus*, but they also impact our understanding of the role of geography in the origin of  
255 species in this group.

### 256 ***Implications for conservation and plant breeding***

257 Our approach is both new and complementary to previous work on *Helianthus* species  
258 distributions and CWR in the literature (Thompson *et al.*, 1981; Rogers *et al.*, 1982). The method

259 of constraining ranges to known native distributions may have limited our identification of some  
260 the extreme variation. Despite this, many taxa that diverge ecologically from cultivated  
261 sunflower were identified (Fig. 4; Table 1). It was also possible to identify extreme populations  
262 within taxa that showed potential adaptation to different ecological niches.

263 Taxa with larger ranges tend to have greater resilience to changes in environmental  
264 conditions than taxa with more limited distributions (Sheth & Angert, 2014; Sexton *et al.*, 2014).  
265 Thus, the latter may be considered a primary priority for conservation. Several taxa have  
266 expanded far beyond their historical ranges, including *H. annuus*, *H. petiolaris* Nutt., *H.*  
267 *argophyllus* Torrey & Gray, *H. giganteus* L. and *H. tuberosus* L.. While taxa from the non-native  
268 parts of their ranges have not been prioritized, existing accessions from such ranges are  
269 acknowledged, and may be worthwhile for exploration for traits useful in crop breeding.

270 Clustering of CWR by environmental variables has great utility by allowing genetic  
271 resources to be exploited in a more targeted manner. For example, with respect to soil pH the  
272 taxa *H. atrorubens*, *H. resinosus*, and *H. deserticola* occupy different ecological space from  
273 cultivated *H. annuus* (Fig. 4). These taxa represent potential candidates for tolerance to acid or  
274 alkaline soils, particularly to improve the ability of the crop to accumulate heavy metals for  
275 phytoremediation (Fassler *et al.*, 2010). Surprisingly, when examining the properties of the  
276 primary, secondary and tertiary germplasm, often extreme profiles are found in the primary  
277 germplasm. This is fortuitous since introgression from primary germplasm is more likely to be  
278 successful (Fig. 4; Fig. S1; Table S7). Approximately 650 wild *H. annuus* accessions are  
279 conserved in genebanks which occur outside the ecological parameters of the cultivar (Table S7).  
280 The general reduction of environmental diversity occupied by the cultivated sunflower relative to  
281 wild *H. annuus* may indicate the reduction in genetic diversity occurring through domestication.

282           Recent advances in plant and animal breeding (e.g. marker assisted selection, genomic  
283 selection) have been facilitated by low cost molecular marker technologies resulting in new tools  
284 that can be used to broaden the genetic base in crops (Tester & Langridge, 2010). These methods  
285 can shorten breeding cycles, increasing genetic gain per unit time, and allow for wider crosses to  
286 be utilized by minimizing linkage drag (Bernardo, 2008). The recent development of genome  
287 wide marker sets (Bowers *et al.*, 2012; Renaut *et al.* 2013) and release of the *H. annuus* genome  
288 (Kane *et al.*, 2011; <http://www.sunflowergenome.org>) facilitate the use of marker assisted  
289 selection (Iftekharuddaula *et al.*, 2011) by decreasing costs and increasing data resolution.  
290 Further, if germplasm collections are genotyped, these data can be used to associate particular  
291 allelic variants with environmental adaptation (Fang *et al.*, 2014).

#### 292 ***Range overlap of wild relatives of sunflower***

293           Sister species in *Helianthus* often have overlapping ranges, an observation that is  
294 consistent with sympatric and “budding” speciation (parapatric or peripheral range speciation).  
295 Substantial range asymmetry among some (but not all) sister species is also consistent with a  
296 budding speciation scenario (Table S6). The amount of range overlap between sister taxa in  
297 *Helianthus* is similar to recent reports from other plant genera, but different from many animal  
298 groups, where allopatry tends to be the rule in speciation (Mayr, 1954; Soltis *et al.*, 2004;  
299 Quenouille *et al.*, 2011; Anacker & Strauss, 2014). This may suggest that geographic isolation is  
300 less critical to plant than animal speciation, perhaps because of the low vagility of many plant  
301 species.

302           Unlike sympatric congeners in other plant groups (Grossenbacher *et al.*, 2014; Anacker &  
303 Strauss, 2014), *Helianthus* sister taxa typically lack strong ecological divergence. This  
304 observation is inconsistent with most models of speciation involving gene flow, which assume

305 divergent ecological selection (Via, 2009). Possibly, our analyses lacked sufficient resolution or  
306 focus on key ecological attributes to detect real differences between the ecological niches of  
307 these species. For example, it is possible that there has been pollinator and phenological  
308 divergence between sister species that was not included in our analyses. Alternatively, local  
309 niche differences between sympatric populations may have been masked by substantial  
310 ecological heterogeneity among populations of the more widely ranging species. Additionally,  
311 the approach used was designed to analyze potential habitat in the historical, native range, rather  
312 than recent range expansions, which in many cases may be recent introductions facilitated by  
313 humans, perhaps accounting for observations of limited ecological divergence.

314 Our analyses imply that many *Helianthus* taxa have similar ecological niches and exhibit  
315 niche conservatism. Under niche conservatism, greater allopatric and parapatric speciation is  
316 predicted, as habitat fragmentation is expected to contribute to reproductive isolation (Loera *et*  
317 *al.*, 2012). While such a speciation strategy would be surprising given the overlap in geographic  
318 range of sister species within *Helianthus*, this trend has been observed in North American  
319 *Ephedra* (Loera *et al.*, 2012). That larger amount of niche conservatism observed here than in  
320 other systems may be due to properties of the K-statistic, which can have inflated values in  
321 polyphyletic phylogenies and in the presence of incomplete lineage sorting, both of which occur  
322 in *Helianthus* (Rosenthal *et al.*, 2002; Gross & Rieseberg, 2005; Horandl & Stuessey, 2010;  
323 Davies *et al.*, 2012).

## 324 **Conclusions**

325 Using a combination of gap analysis, environmental niche modeling and phylogenetic  
326 approaches 36 CWR of sunflower were examined. Taxa that are under-represented in germplasm  
327 collections as well as species and populations inhabiting environmental niches with extreme



328 phenotypes that may possess traits of value to crop improvement were identified. In *Helianthus*,  
329 sister taxa appear to occur more frequently in sympatry than allopatry, possibly suggesting that  
330 speciation may occur in the presence of gene flow. Finally, much of the primary gene pool occurs  
331 in extreme environments indicating that utilization of wild *H. annuus* for the breeding of abiotic  
332 stress tolerance may produce quick gains with minimal effort.

### 333 **Acknowledgments**

334 This work was undertaken as part of the initiative “Adapting Agriculture to Climate Change:  
335 Conserving, Protecting and Preparing Crop Wild Relatives” which is supported by the  
336 Government of Norway. The project is managed by the Global Crop Diversity Trust with the  
337 Millennium Seed Bank of the Royal Botanic Gardens, Kew UK and implemented in partnership  
338 with national and international genebanks and plant breeding institutes around the world. For  
339 further information, go to the project website: <http://www.cwrdiversity.org/>. Funding was  
340 provided by the aforementioned initiative, The National Sunflower Association, The U.S.  
341 National Science Foundation, The Natural Sciences and Engineering Research Council of  
342 Canada, Genome BC, and Genome Canada.

343 **References**

- 344 Anacker, B.L., and Strauss, S.Y. (2014). The geography and ecology of plant speciation: range  
345 overlap and niche divergence in sister species. *Proc. R. Soc. B*, 281, 20132980.  
346 doi:10.1098/rspb.2013.2980
- 347 Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: learning from  
348 the last 20 years. *Crop Sci*, 48, 1649-1664.
- 349 Blackman B.K., Scascitelli M., Kane N.C., Luton H.H., Rasmussen D.A., Bye R.A. et al. (2011).  
350 Sunflower domestication alleles support single domestication center in eastern North America.  
351 *Proc Natl Acad Sci USA*, 108, 14360-14365.
- 352 Blomberg S.P., Garland T., and Ives A.R. (2003). Testing for phylogenetic signal in comparative  
353 data: behavioral traits are more labile. *Evolution*, 57, 717–745.
- 354 Bowers, J. E., Nambeesan, S., Corbi, J., Barker, M.S., Rieseberg, L.H., Knapp, S.J. et al. (2012).  
355 Development of an Ultra-Dense Genetic Map of the Sunflower Genome Based on Single-Feature  
356 Polymorphisms. *PLoS One*, 7, e51360.
- 357 Butler, E.E., and Huybers, P. (2013). Adaptation of US maize to temperature variations. *Nat*  
358 *Clim Change*, 3, 68-72.
- 359 Cabrera-Bosquet, L., Crossa, J., von Zitzewitz, J., Serret, M.D., and Luis Araus, J. (2012). High-  
360 throughput phenotyping and genomic selection: The frontiers of crop breeding converge. *Journal*  
361 *of integrative plant biology*, 54, 312-320.
- 362 Challinor, A.J., Watson, J., Lobell, D.B., Howden, S.M., Smith, D.R., and Chhetri, N. (2014). A  
363 meta-analysis of crop yield under climate change and adaptation. *Nat Clim Change*, 4, 287-291.
- 364 Chandler J.M., Jan C., and Beard B.H. (1986). Chromosomal differentiation among the annual  
365 *Helianthus* species. *Systematic Botany*, 11, 354-371.
- 366 Davies, T.J., Kraft, N.J.B., Salamin, N., and Wolkovich, E.M. (2012). Incompletely resolved  
367 phylogenetic trees inflate estimates of phylogenetic conservatism. *Ecology*, 93, 242–247.
- 368 Dray, S. and Dufour, A.B. (2007). The ade4 package: implementing the duality diagram for  
369 ecologists. *Journal of Statistical Software*, 22, 1-20.
- 370 Fang, Z., Gonzales, A.M., Clegg, M.T., Smith, K.P., Muehlbauer, G.J., Steffenson, B.J. et al.,  
371 (2014) Two genomic regions contribute disproportionately to geographic differentiation in wild  
372 barley. *G3: Genes/ Genomes/ Genetics*, 4, 1193-1203.
- 373 Fassler, E., Robinson, B.H. Stauffer, W., Gupta, S.K., Papritz, A., and Schulin, R. (2010).  
374 Phytomanagement of metal contaminated agricultural land using sunflower, maize and tobacco.  
375 *Agriculture Ecosystems and Environment*, 136, 49–58.
- 376 FAOSTAT. *Final Data 2013*. Retrieved May, 2015. <http://faostat.fao.org>.

377 Grossenbacher, D.L., Veloz, S.D., and Sexton, J.P. (2014). Niche and range size patterns suggest  
378 that speciation begins in small, ecologically diverged populations in North American  
379 monkeyflowers (*Mimulus* spp.). *Evolution*, 68, 1270-1280.

380 Gross, B.L., and Rieseberg, L.H. (2005). The ecological genetics of homoploid hybrid  
381 speciation. *The Journal of Heredity*, 96, 241–52. doi:10.1093/jhered/esi026

382 Hajjar, R., and Hodgkin T. (2007). The use of wild relatives in crop improvement: A survey of  
383 developments over the last 20 years. *Euphytica*, 156:1–13.

384 Harlan, J.R. (1976). Genetic resources in wild relatives of crops. *Crop Sci*, 16, 329–333.

385 Harlan, J.R., and de Wet, J.M.J. (1971). Toward a rational classification of cultivated plants.  
386 *Taxon*, 20, 509–517.

387 Harlan, J.R., De Wet, J.M.J., and Price, E. G. (1973). Comparative evolution of cereals.  
388 *Evolution*, 27, 311-325.

389 Harter A.V., Gardner K.A., Falush D., Lentz D.L., Bye R., and Rieseberg L.H. (2004). Origin of  
390 extant domesticated sunflowers in eastern North America. *Nature*, 430, 201-205.

391 Heibl C. (2011). [<http://cran.r-project.org/web/packages/phyloclim/index.html>] webcite  
392 phyloclim: Integrating phylogenetics and climatic niche modelling. OpenURL

393 Hengl T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M, Ribeiro, E., et al.,  
394 (2014). SoilGrids1km — Global Soil Information Based on Automated Mapping. *PLoS ONE*  
395 9(8): e105992. doi: 10.1371/journal.pone.0105992

396 Hijmans, R.J., Guarino, L., Cruz, M., and Rojas, E. (2001) Computer tools for spatial analysis of  
397 plant genetic resources data: 1. DIVA-GIS. *Plant Genetic Resource Newsletter*. 127, 15–19.

398 Hijmans, R.J., Guarino, L., Jarvis, A., O'Brien, R., Mathur, P., Bussink, C., et al. (2005a). *DIVA-*  
399 *GIS version 5.2 manual*. Available: <http://www.diva-gis.org/Materials.htm>.

400 Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. and Jarvis, A. (2005b). Very high  
401 resolution interpolated climate surfaces for global land areas. *International Journal of*  
402 *Climatology*, 25, 1965-1978.

403 Hoisington D., Khairallah, M., Reeves, T., Ribault, J.M., Skovmand, B., Taba, S., et al. (1999).  
404 Plant genetic resources: what can they contribute toward increased crop productivity? *Proc Natl*  
405 *Acad Sci USA*, 96, 5937-5943.

406 Horandl, E., and Stuessy, T. (2010). Paraphyletic groups as natural units of biological  
407 classification. *Taxon*, 59, 1641–1653.

408 Hulke, B.S., Miller, J.F., Gulya, T.J., and Vick, B.A. (2010). Registration of the oilseed sunflower  
409 genetic stocks HA 458, HA 459, and HA 460 possessing genes for resistance to downy mildew.  
410 *Journal of Plant Registrations*, 4, 1-5.

411 Husson, F., Josse, J., Le, S., and Mazet J. (2014). FactoMineR: Multivariate Exploratory Data  
412 Analysis and Data Mining with R. R package version 1.27. <http://CRAN.R-project.org/package>  
413 = FactoMineR

414 Iftekharuddaula, K.M., Newaz, M.A., Salam, M.A., Ahmed, H.U., Mahbub, M.A.A.,  
415 Septiningsih, E.M., et al. (2011). Rapid and high-precision marker assisted backcrossing to  
416 introgress the SUB1 QTL into BR11, the rainfed lowland rice mega variety of Bangladesh.  
417 *Euphytica*, 178, 83-97.

418 Jarvis, A., Lane, A., and Hijmans, R.J. (2008). The effect of climate change on crop wild  
419 relatives. *Agriculture, Ecosystems & Environment*, 126, 13-23.

420 Loera, I., Sosa, V., and Ickert-Bond, S.M. (2012). Diversification in North American arid lands:  
421 Niche conservatism, divergence and expansion of habitat explain speciation in the genus  
422 *Ephedra*. *Molecular Phylogenetics and Evolution*, 65, 437-450.

423 Kane, N.C., Gill, N., King, M, Bowers, J.E., Berges, H., Gouzy, J., et al., (2011) Progress  
424 towards a reference genome for sunflower. *Botany*, 89, 429-437.

425 Kane N.C., Burke, J.M., Marek, L.F., Seiler, G.J., Vear, F., Knapp, S.J., et al. (2013). Sunflower  
426 genetic, genomic, and ecological resources. *Molecular Ecology Resources*, 13, 10-20.

427 Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., et al.,  
428 (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26, 1463-1464.

429 Khoury, C., Laliberté, B., and Guarino, L. (2010). Trends in ex situ conservation of plant genetic  
430 resources: a review of global crop and regional conservation strategies. *Genetic Resources and*  
431 *Crop Evolution*, 57, 625-639.

432 Kolkman, J.M., Berry, S.T., Leon, A.J., Slabaugh, M.B., Tang, S., Gao, W., et al. (2007). Single  
433 nucleotide polymorphisms and linkage disequilibrium in sunflower. *Genetics*, 177, 457-68.

434 Kozak, K. H., and J. J. Wiens. (2006). Does niche conservatism promote speciation? A case  
435 study in North American salamanders. *Evolution* 60, 2604–2621.

436 Liu C., Berry, P.M., Dawson, T.P., and Pearson, R.G. (2005). Selecting thresholds of occurrence  
437 in the prediction of species distributions. *Ecography*, 28, 385-393.

438 Liu, A., and Burke, J.M. (2006). Patterns of nucleotide diversity in wild and cultivated  
439 sunflower. *Genetics*, 173, 321-330.

440 Long, R.W. (1960). Biosystematics of two perennial species of *Helianthus* (Compositae). I.  
441 Crossing relationships and transplant studies. *American Journal of Botany*, 47,729-735.

442 Loera, I., Sosa, V., and Ickert-Bond, S. M. (2012). Diversification in North American arid lands:  
443 Niche conservatism, divergence and expansion of habitat explain speciation in the genus *Ephedra*.  
444 *Molecular phylogenetics and evolution*, 65, 437-450.

445 Losos, J.B. (2008). Phylogenetic niche conservatism, phylogenetic signal and the relationship  
446 between phylogenetic relatedness and ecological similarity among species. *Ecology Letters*, 11,  
447 995–1003.

448 Mandel, J.R., Dechaine, J.M., Marek, L.F., & Burke, J.M. (2011) Genetic diversity and population  
449 structure in cultivated sunflower and comparison to its wild progenitor *Helianthus annuus* L.  
450 *Theoretical and Applied Genetics*, 123, 693-704.

451 Maxted, N., Ford-Lloyd, B.V., Jury, S.L., Kell, S.P., and Scholten, M.A. (2006) Towards a  
452 definition of a crop wild relative. *Biodiversity and Conservation*, 15, 2673-2685.

453 Mayr, E. (1954) Geographic speciation in tropical echinoids. *Evolution*, 8, 1–18.

454 McCouch, S., Baute, G.J., Bradeen, J., Bramel, P., Bretting, P.K., Buckler, E., et al., (2013).  
455 Agriculture: Feeding the future. *Nature*, 499, 23-24.

456 Monfreda, C., Ramankutty, N., and Foley, J.A. (2008). Farming the planet: 2. Geographic  
457 distribution of crop areas, yields, physiological types, and net primary production in the year  
458 2000. *Global Biogeochemical Cycles* 22: GB1022. Data available online at  
459 <http://www.geog.mcgill.ca/landuse/pub/Data/175crops2000/>.

460 Nix, H.A. (1986). A biogeographic analysis of Australian elapid snakes. In R. Longmore, ed.  
461 Atlas of Elapid Snakes of Australia. Canberra: *Australian Government Publishing Service*, pp.  
462 4–15.

463 Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood,  
464 E.C., et al. (2001). Terrestrial ecoregions of the world: a new map of life on earth. *BioScience*,  
465 51, 933-938.

466 Phillips, S.J., Anderson, R.P., and Schapire, R.E. (2006) Maximum entropy modeling of species  
467 geographic distributions. *Ecological Modelling*, 190, 231-259.

468 Phillips, S.J. (2008). Transferability, sample selection bias and background data in presence-only  
469 modeling: a response to Peterson *et al.* (2007). *Ecography*, 31, 272-278.

470 Porter, J.R., Xie, L., Challinor, A.J., Cochrane, K., Howden, S.M., Iqbal, M.M., Lobell, D.B., et  
471 al. (2014). Food security and food production systems. In C. B. Field *et al.*, eds. *Climate Change*  
472 *2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution*  
473 *of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate*  
474 *Change*. Cambridge, United Kingdom and New York, NY: Cambridge University Press.

475 Putt, E.D. (1978). History and present world status. In: *Sunflower and science and technology*  
476 (ed. Carter, J.P.). pp. 1–29. American Society of Agronomy, Madison, WI (USA).

477 Quenouille, B., Hubert, N., Bermingham, E., and Planes, S. (2011). Speciation in tropical seas:  
478 allopatry followed by range change. *Molecular phylogenetics and evolution*, 58, 546-552.

479 Ramírez-Villegas, J., Khoury, C., Jarvis, A., Debouck, D.G., and Guarino, L. (2010). A gap  
480 analysis methodology for collecting crop gene pools: a case study with *Phaseolus* beans. *PLoS*  
481 *ONE*, 5, e13497.

482 Revell L.J. 2009. Size-correction and principal components for interspecific comparative studies.  
483 *Evolution* 63: 3258-326

484 Ricklefs, R.E., and Jenkins, D.G. (2011). Biogeography and ecology: towards the integration of  
485 two disciplines. *Phil Trans R Soc B*, 366, 2438-2448.

486 Rieseberg, L.H., Carter, R., and Zona, S. (1990). Molecular tests of the hypothesized hybrid  
487 origin of two diploid *Helianthus* species (Asteraceae). *Evolution*, 44, 1498-1511.

488 Rieseberg, L.H., Van Fossen, C., and Desrochers, A.M. (1995). Hybrid speciation accompanied  
489 by genomic reorganization in wild sunflowers. *Nature*, 375, 313-316.

490 Rieseberg, L.H., Raymond, O., Rosenthal, D.M., Lai, Z., Livingstone, K., Nakazato, T., et al.  
491 (2003). Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, 301,  
492 1211-1216.

493 Rogers, C.E., Thompson, T.E., and Seiler, G.J. (1982). Sunflower species of the United States  
494 (pp. 1-75). Bismarck, ND: National Sunflower Association.

495 Rosenthal, D.M., Schwarzbach, A.E., Donovan, L.A., Raymond, O., and Rieseberg, L.H. (2002).  
496 Phenotypic Differentiation between Three Ancient Hybrid Taxa and Their Parental Species.  
497 *International Journal of Plant Sciences*, 163, 387-398.

498 Schilling, E.E. (2006). *Helianthus*. In Flora of North America Committee, eds. *Flora of North*  
499 *America North of Mexico*. New York and Oxford. 21, 141-169.

500 Seiler, G., and Marek, L.F. (2011). Germplasm resources for increasing the genetic diversity of  
501 global cultivated sunflower. *Helia*, 34, 1-20.

502 Sexton, J. P., Hangartner, S.B., and Hoffmann, A.A. (2014). Genetic isolation by environment or  
503 distance: which pattern of gene flow is most common? *Evolution*, 68, 1-15.

504 Sheth, S.N., and Angert, A.L. (2014). The evolution of environmental tolerance and range size: a  
505 comparison of geographically restricted and widespread *Mimulus*. *Evolution*, 68, 2917-2931.

506 Soltis, D.E., Soltis, P.S., & Tate, J.A. (2004) Advances in the study of polyploidy since plant  
507 speciation. *New Phytologist*, 161, 173-191. (doi:10.1046/j.1469-8137. 2003.00948.

508 Stebbins, J.C., Winchell, C.J., and Constable, J.V.H. (2013). *Helianthus winteri* (Asteraceae), a  
509 new perennial species from the southern Sierra Nevada foothills, California. *Aliso* 31: 19-24.

510 Tamura, K., Stecher, G., Peterson, D., Filipowski, A., and Kumar, S. (2013). MEGA6: Molecular  
511 Evolutionary Genetics Analysis Version 6.0. *Molecular biology and evolution*, 30, 2725-2729.

512 Tang, S., Yu, J.K., Slabaugh, M.B., Shintani, D.K., and Knapp, S.J. (2002). Simple Sequence  
513 repeat map of the sunflower genome. *Theoretical and Applied Genetics*, 105, 1124-1136.



- 514 Tester, M., and Langridge, P. (2010). Breeding Technologies to Increase Crop Production in a  
515 Changing World. *Science*, 327, 818.
- 516 Thompson, T. E., Zimmerman, D. C., and Rogers, C. E. (1981). Wild *Helianthus* as a genetic  
517 resource. *Field Crops Research*, 4, 333-343.
- 518 Timme, R. E., Simpson, B. B., and Linder, C. R. (2007). High-resolution phylogeny for  
519 *Helianthus* (Asteraceae) using the 18S-26S ribosomal DNA external transcribed spacer.  
520 *American Journal of Botany*, 94, 1837-1852.
- 521 Uyeda J.C., Caetano D.S., Pennell M.W. 2015. Comparative analysis of principal components  
522 can be misleading. *Systematic Biology* 64: 677-689.
- 523 VanDerWal, J., Shoo, L.P., Graham, C.H., Williams, S.E. (2009). Selecting pseudo-absence data  
524 for presence-only distribution modeling: How far should you stray from what you know?  
525 *Ecological Modeling*, 220, 589-594.
- 526 Via, S. (2009). Natural selection in action during speciation. *Proc Natl Acad Sci USA*, 106, 9939-  
527 9946.
- 528 Vincent, H., Wiersema, J., Kell, S., Fielder, H., Dobbie, S., Castañeda-Álvarez, N. P., et al.  
529 (2013). A prioritized crop wild relative inventory to help underpin global food security.  
530 *Biological Conservation*, 167, 265-275.

531  
532

**Table 1. Taxa examined in this study, recommendation, position in germplasm, environmental cluster, life history, and potential extreme characteristics.**

<u>Taxa</u>	<u>Recommendation for Collection</u>	<u>Position in Germplasm</u>	<u>Range overlap with <i>H. annuus</i></u>	<u>Environmental Cluster Assignment</u>	<u>Life History</u>	<u>Potential Extreme Characteristics Based on Different Ecological Niche Relative to <i>H. annuus</i></u>
<i>H. annuus</i> (wild)	Assessed to be well represented	Primary	NA	Cluster 1	Annual	NA
<i>H. anomalus</i>	High priority	Secondary	Utah New Mexico	Cluster 3	Annual	Low precipitation tolerance Tolerance to high pH
<i>H. argophyllus</i>	Medium priority	Secondary	Texas	Cluster 1	Annual	High temperature tolerance Tolerance to high clay content
<i>H. arizonensis</i>	Medium priority	Tertiary	Arizona New Mexico	Cluster 3	Perennial	Response to stochastic climate Low precipitation tolerance Tolerance to low bulk density
<i>H. atrorubens</i>	Medium priority	Tertiary	No overlap	Cluster 2	Perennial	Tolerance to low Cation-exchange capacity Tolerance of high precipitation Tolerance to low pH
<i>H. bolanderi</i>	High priority	Secondary	California	Cluster 1	Annual	Tolerance to erratic precipitation Low precipitation tolerance
<i>H. debilis</i> subsp. <i>cucmerifolius</i>	High priority	Secondary	East Texas	Cluster 2	Annual	High temperature tolerance
<i>H. debilis</i> subsp. <i>debilis</i>	Medium priority	Secondary	No overlap	Cluster 2	Annual	High temperature tolerance Tolerance of high precipitation Tolerance to low clay content
<i>H. debilis</i> subsp. <i>silvestris</i>	Medium priority	Secondary	No overlap	Cluster 2	Annual	Tolerance to high clay content
<i>H. debilis</i> subsp. <i>tardiflorus</i>	Assessed to be well represented	Secondary	No overlap	Cluster 2	Annual	Tolerance of high precipitation Tolerance to low clay content
<i>H. debilis</i> subsp. <i>vestitus</i>	Low priority	Secondary	No overlap	Cluster 2	Annual	High temperature tolerance Tolerance of high precipitation Tolerance to low clay content
<i>H. deserticola</i>	High priority	Secondary	Nevada Utah New Mexico	Cluster 3	Annual	Response to stochastic climate Low precipitation tolerance
<i>H. divaricatus</i>	High priority	Tertiary	Central US	Cluster 2	Perennial	Perennial habit Tolerance to low pH
<i>H. exilis</i>	Medium priority	Secondary	California	Cluster 1	Annual	Tolerance to erratic precipitation Low precipitation tolerance Low bulk density
<i>H. giganteus</i>	High priority	Tertiary	No overlap	Cluster 2	Perennial	Tolerance of high precipitation
<i>H. grosseserratus</i>	Medium priority	Tertiary	Central US	Cluster 3	Perennial	Tolerance to erratic temperature
<i>H. hirsutus</i>	High priority	Tertiary	Central US	Cluster 2	Perennial	Tolerance to low pH
<i>H. maximilliani</i>	High priority	Tertiary	Central US	Cluster 3	Perennial	Low temperature tolerance Tolerance to erratic temperature
<i>H. neglectus</i>	Assessed to be well represented	Secondary	New Mexico	Cluster 1	Annual	Low organic carbon content
<i>H. niveus</i> subsp. <i>canescens</i>	High priority	Secondary	California Arizona New Mexico	Cluster 1	Annual Rarely Perennial	High temperature tolerance Low precipitation tolerance
<i>H. niveus</i> subsp. <i>niveus</i>	High priority	Secondary	Baja California	Cluster 1	Perennial	Low precipitation tolerance
<i>H. niveus</i> subsp. <i>tephrodes</i>	High priority	Secondary		Cluster 1	Perennial	High temperature tolerance low Precipitation tolerance



			California, Mexico (Sonora)		Sometime Annual	
<i>H. paradoxus</i>	Assessed to be well represented	Secondary	Texas, New Mexico	Cluster 1	Annual	Low organic carbon content
<i>H. pauciflorus</i> subsp. <i>pauciflorus</i>	High priority	Tertiary	Central US	Cluster 3	Perennial	Tolerance to erratic temperature
<i>H. pauciflorus</i> subsp. <i>subrhomboideus</i>	High priority	Tertiary	Central US	Cluster 3	Perennial	Low temperature tolerance Tolerance to erratic temperature
<i>H. petiolaris</i> subsp. <i>fallax</i>	High priority	Secondary	Western US	Cluster 3	Annual	Tolerance to erratic temperature
<i>H. petiolaris</i> subsp. <i>petiolaris</i>	High priority	Secondary	Central US	Cluster 3	Annual	Tolerance to erratic temperature Low temperature tolerance
<i>H. praecox</i> subsp. <i>hirtus</i>	Assessed to be well represented	Secondary	West Texas	Cluster 1	Annual	High temperature tolerance
<i>H. praecox</i> subsp. <i>praecox</i>	Assessed to be well represented	Secondary	East Texas	Cluster 2	Annual	Tolerance to erratic temperature
<i>H. praecox</i> subsp. <i>runyonii</i>	Low priority	Secondary	Texas	Cluster 1	Annual	Tolerance of high bulk density
<i>H. resinus</i>	Medium priority	Tertiary	No overlap	Cluster 2	Perennial	Tolerance of high precipitation Tolerance to low Cation exchange capacity Tolerance to low pH
<i>H. salicifolius</i>	Medium priority	Tertiary	Oklahoma Kansas Arkansas Missouri	Cluster 3	Perennial	Tolerance to high clay content
<i>H. silphoides</i>	Assessed to be well represented	Tertiary	Oklahoma Arkansas Missouri	Cluster 2	Perennial	Tolerance to low cation-exchange capacity Tolerance to low pH
<i>H. strumosus</i>	High priority	Tertiary	Central US	Cluster 2	Perennial	Tolerance of high precipitation
<i>H. tuberosus</i>	Medium priority	Secondary	Central US	Cluster 2	Perennial	Low temperature tolerance
<i>H. winteri</i>	High priority	Primary	California	Cluster 1	Perennial	High temperature tolerance

534 **Table 2. Environmental Niche occupancy based on Schoener's D (1968) and a modified**  
 535 **Hellinger's I (Warren *et al.*, 2008).**

	<u>Perfect Overlap</u>	<u>D or I Greater than 0.5</u>	<u>D or I Less than 0.2</u>
	<u>(%)</u>	<u>(%)</u>	<u>(%, Divergent Niche)</u>
All taxa	36.9	69.4	4.7
Annual taxa	32.2	36.6	6.6
Perennial taxa	19.8	85.7	2.2
Allopatric taxa	54.2	62.5	4.3
Sympatric taxa	3.3	83.3	2.6
Sister taxa	33.3	57.7	2.6

536

537

538

Provisional

539 **Figure Legends**

540 Fig. 1 Synthesis of gap analysis results and expert assessments for each of the 36 *Helianthus*  
541 CWR taxa surveyed. Taxa are listed by descending priority for further collecting by category:  
542 HPS, high priority taxa; MPS, medium priority taxa; LPS, low priority taxa; NFCR, no further  
543 collecting recommended. The final priority scores (FPS, black circle) is the mean of the sampling  
544 representativeness score (SRS, blue circle), geographic representativeness score (GRS, red  
545 circle), and ecological representativeness score (ERS, green circle).

546 Fig. 2 Map of North America showing A) taxon richness of sunflower and B) hotspots for further  
547 collecting of high priority taxa.

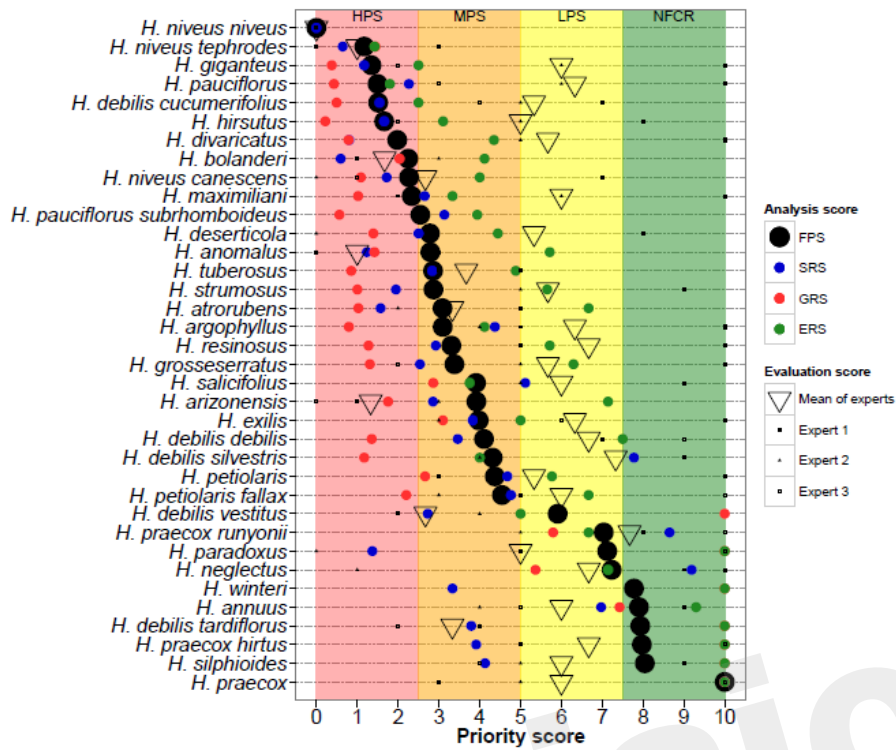
548 Fig. 3 Geographic niche overlap based on bioclimatic and biophysical variables, both calculated  
549 by D (above diagonal) and I (below diagonal) . Taxa are grouped by the phylogenetic  
550 relationship identified from the ETS sequences retrieved from NCBI. Values closer to 0 (no  
551 overlap = niche divergence) are purple while values closer to 1 (complete overlap = niche  
552 convergence) are orange; B) Occurrence points for each taxa grouped based on the first three  
553 principle components of biophysical and bioclimatic variables. Clusters share homogeneous  
554 bioclimatic and biophysical conditions.

555 Fig. 4 Climatic niches for A) mean diurnal range and annual precipitation, B) Soil pH and mean  
556 annual precipitation, C) mean diurnal range and annual precipitation. Niches per taxa represent  
557 the middle 90% of occurrence points, i.e., 10% outliers are not included. Red boxes show the  
558 niche of wild *H. annuus* and black boxes show the niche of cultivated *H. annuus* in North  
559 America.

560 Fig. 5 Test of phylogenetic signal utilizing the K for 25 of 36 taxa analyzed with complete  
561 genetic and environmental information (Blomberg, et al, 2003). K measures phylogenetic signal  
562 in traits, where K values below 1 indicates low dependence of traits on evolutionary history (not  
563 conserved between taxa) and K values above 1 indicates trait conservation over evolutionary  
564 history (traits conserved over evolutionary time). \*indicates K significantly greater than 1 ( $p <$   
565 0.05).

566

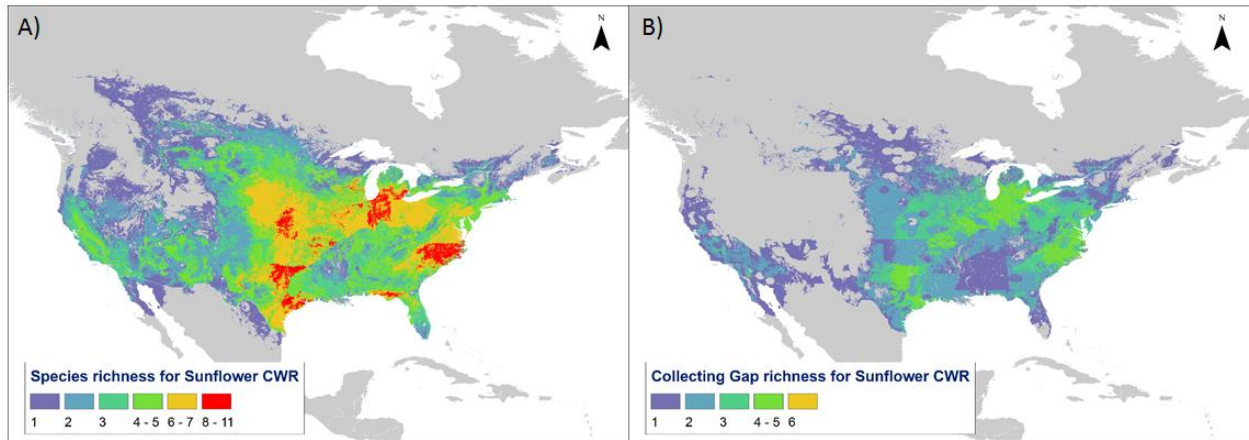
567 **Fig. 1**



568  
569

Provisional

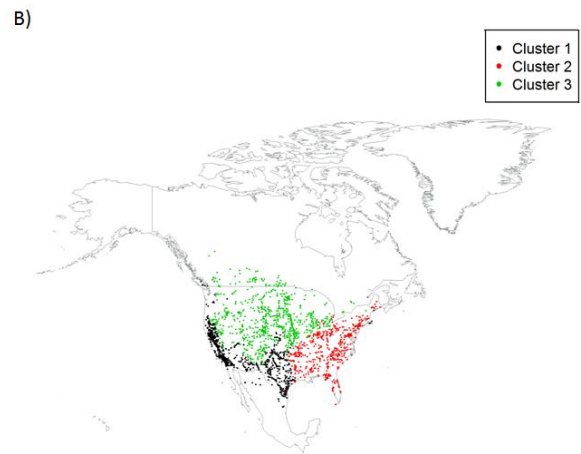
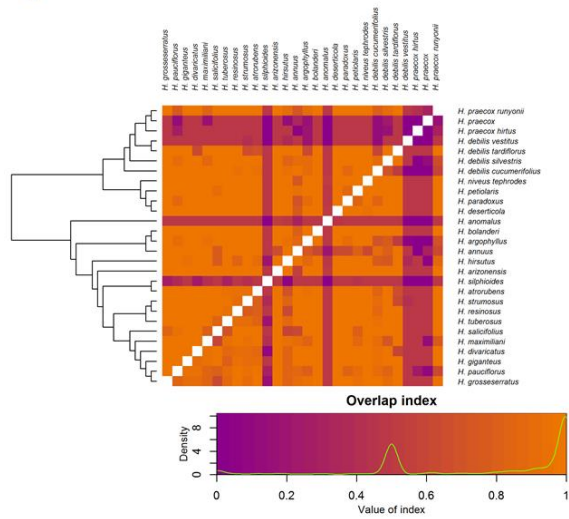
570 **Fig. 2.**  
571



572  
573

Provisional

574 **Fig. 3**  
A)

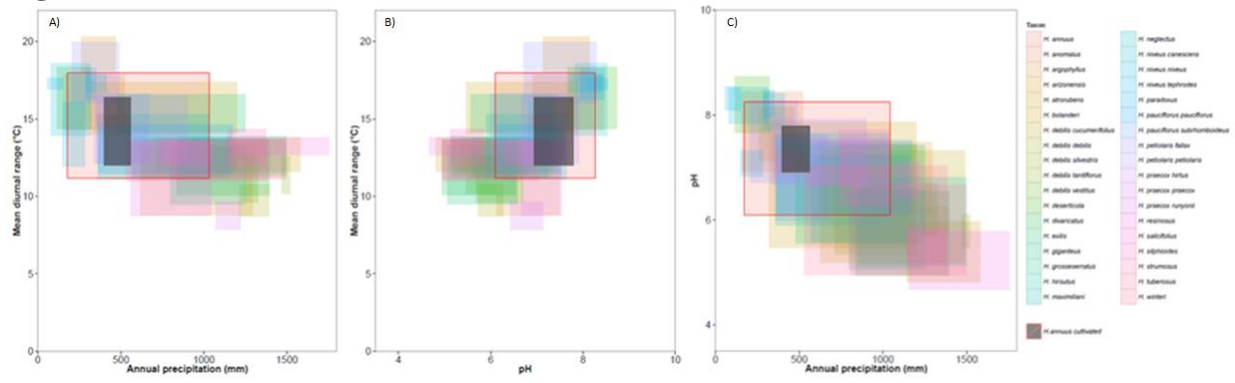


575

Provisional

576

**Fig. 4**

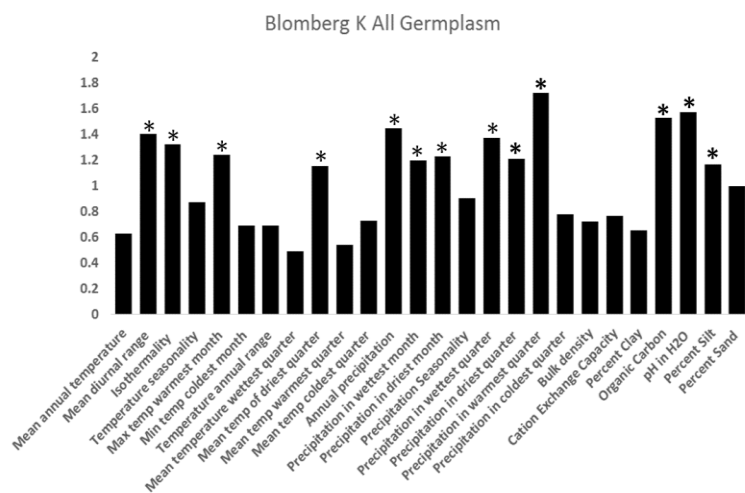


577

578

Provisional

579 **Fig. 5**



580

Provisional



581 **Supplementary information**

582

583 Table S1. *Helianthus* taxa which have provided useful traits for cultivated sunflower.

584 Table S2. Name and location of the 31 herbaria and five germplasm Institutes from which taxa  
585 data were sourced.

586 Table S3. Bioclimatic and biophysical variables examined and correlation between climatic  
587 variables and selected principal components.

588 Table S4. Bioclimatic and biophysical variables partitioned into clusters using the R package  
589 FactoMineR variables. All of the cluster 1 variables are related to temperature and cluster 1 can  
590 be defined by dry climatic conditions. Cluster 2 is defined by precipitation variables, and is  
591 associated with humid climatic conditions and high soil organic matter. Cluster 3 contains a  
592 combination of soil and temperature variables. This cluster has soils with higher than average silt  
593 content, a higher capacity for cation exchange, neutral pH, and higher soil porosity.

594 Table S5. A) Geographic overlap as determined with respect to the smaller (minor range) in the  
595 bottom left, and larger range (major) in the top right. B) Difference between minor and major  
596 range overlap. Red indicates no geographic overlap, white indicates a small amount of overlap  
597 and blue indicates a larger amount of overlap.

598 Table S6. Geographic overlap of 12 sister taxa pairs present in our data represented as percent of  
599 shared grid cells.

600 Table S7. Populations of wild *H. annuus* that are outliers relative to domestic *H. annuus* so that  
601 they may be useful for abiotic stress breeding, (yellow indicates lower than 2.5% of the domestic  
602 *H. annuus* distribution, blue indicates outside the 97.5% of the domestic *H. annuus* distribution).

603 Fig. S1 Climatic niches of *Helianthus* species per bioclimatic variable.

604 Fig. S2 Species distribution maps for the 36 *Helianthus* taxa examined in this study.

605 Fig. S3 Heat map of geographic overlap as determined with respect to the smaller (minor range)  
606 in the bottom left, and larger range (major) in the top right. Red indicates no geographic overlap,  
607 white indicates a small amount of overlap and blue indicates a larger amount of overlap.

608 Fig. S4 Predicted Niche Occupancy (PNO) for all 19 bioclimatic and 7 biophysical variables.  
609 Horizontal axes represent the bioclim parameter space divided into 50 equally spaced bins;  
610 vertical axes denote the total suitability of the mean annual temperature index of each species  
611 over its entire geographic range. Overlapping peaks of PNO profiles indicate similar tolerances,  
612 while the overall breadth of the profile denotes the degree of specificity in tolerance. Black  
613 profiles indicate the primary germplasm, red indicates the secondary germplasm pool, blue  
614 indicates the tertiary germplasm pool.

Figure 1.TIF

Figure 1. Synthesis of gap analysis results and expert assessments for each of the 36 *Helianthus* CWR taxa surveyed. Taxa are listed by descending priority for further collecting by category: HPS, high priority taxa; MPS, medium priority taxa; LPS, low priority taxa; NFCR, no further collecting recommended. The final priority scores (FPS, black circle) is the mean of the sampling representativeness score (SRS, blue circle), geographic representativeness score (GRS, red circle), and ecological representativeness score (ERS, green circle).

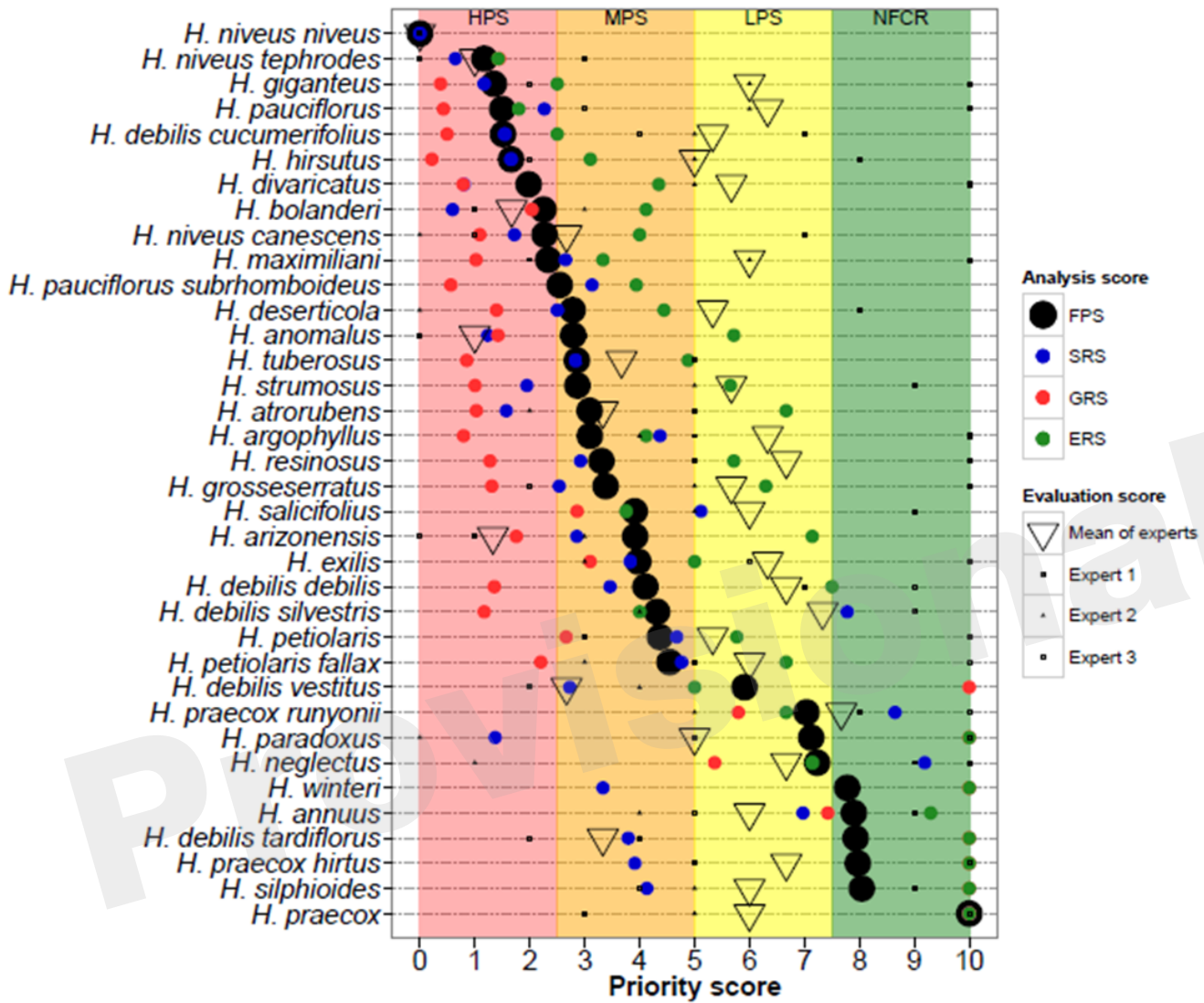
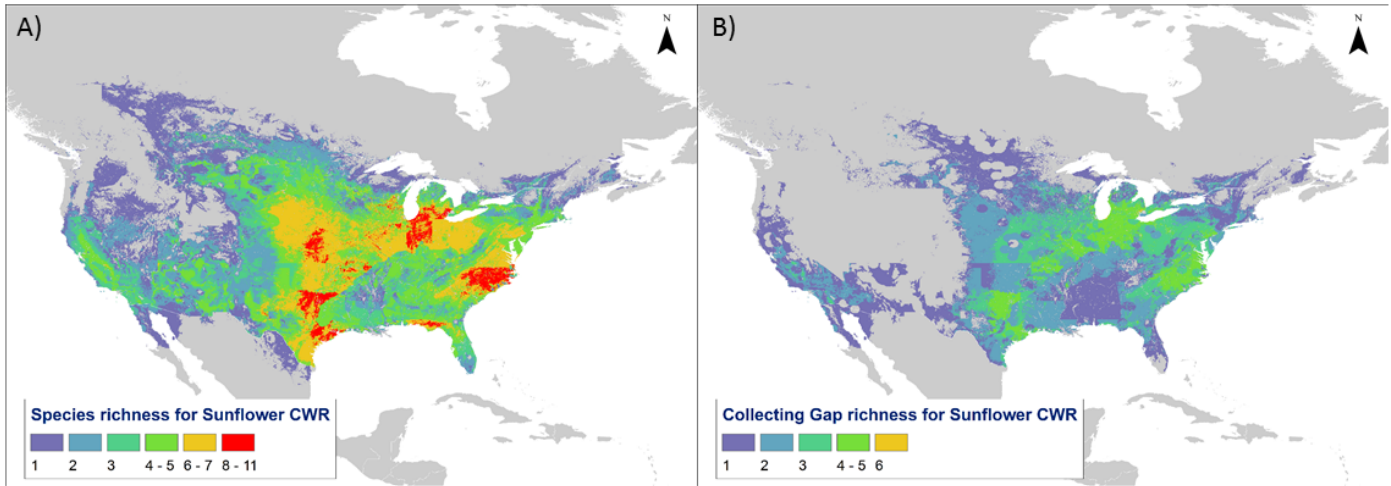


Figure 2.TIF

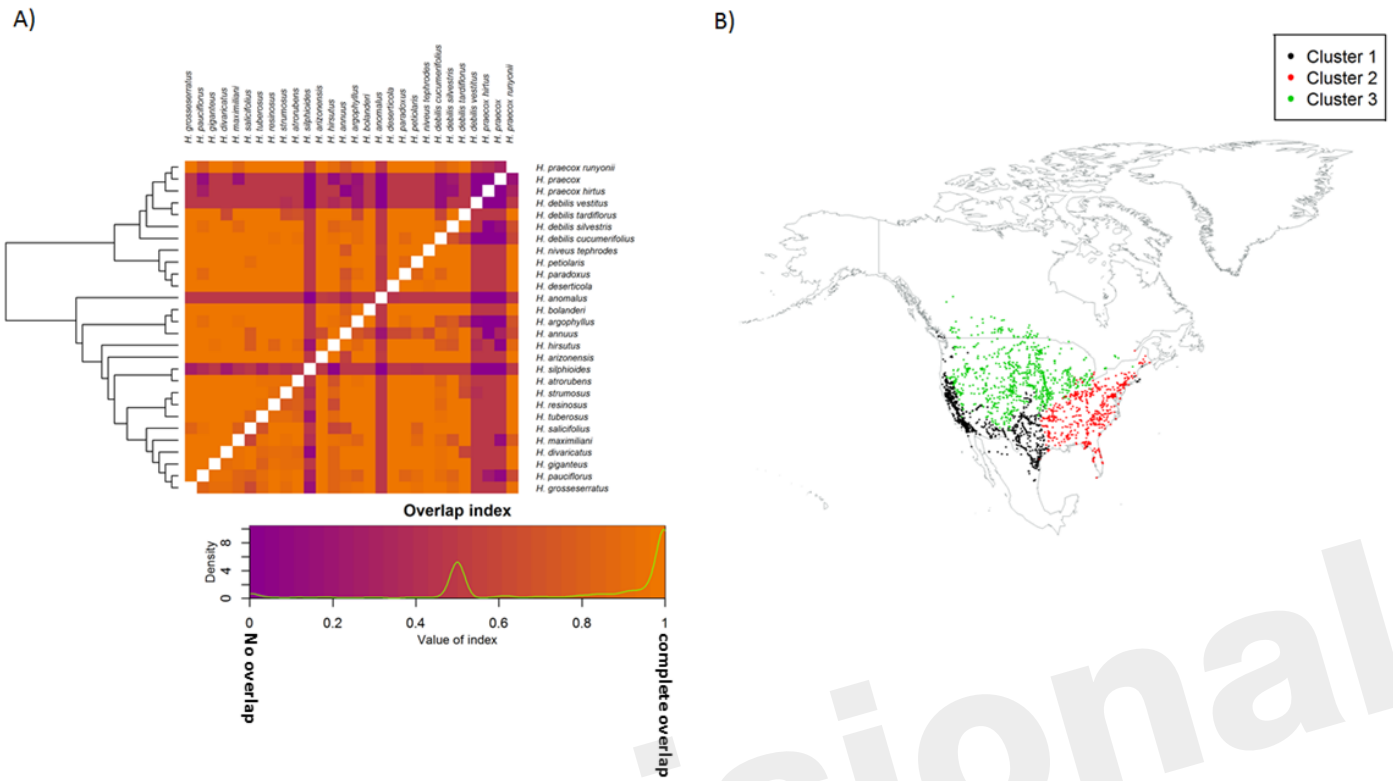
Figure 2. A) Map of North America showing the species richness of sunflower. B) Map of North America showing collection gaps for sunflower; in both maps lower numbers (bluish colors) indicates low species numbers and high numbers (reddish) indicate high species numbers in a given location, all areas colored require collection they differ only in the number of species that need to be collected within the geographic location.



Provisional

Figure 3.TIF

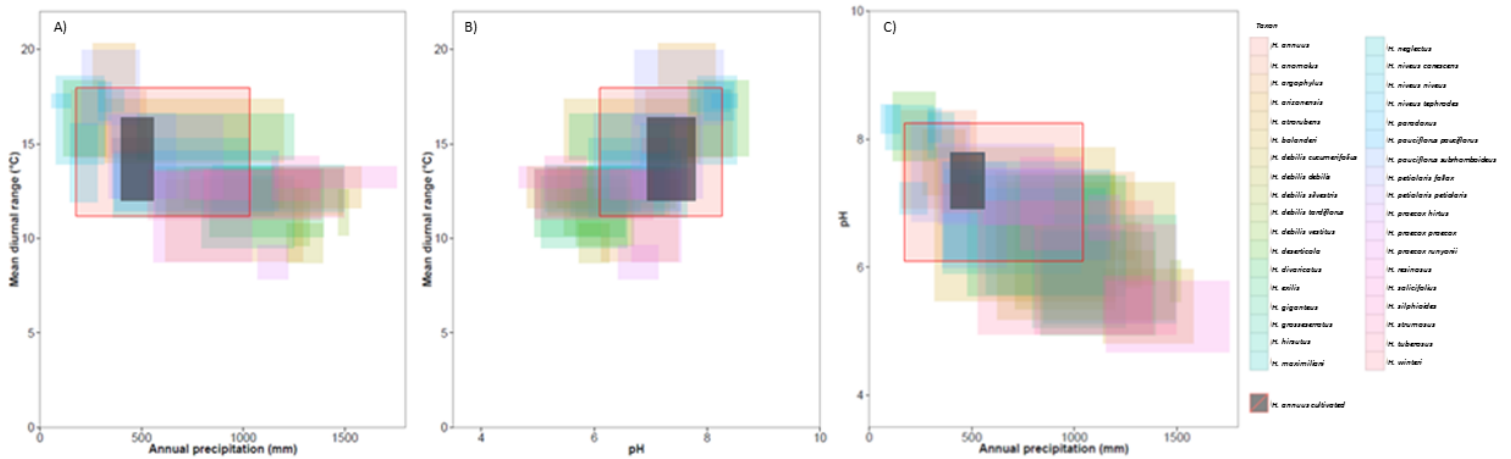
Figure 3. Geographic niche overlap based on bioclimatic and biophysical variables, both calculated by D (above diagonal) and I (below diagonal). Taxa are grouped by the phylogenetic relationship identified from the ETS sequences retrieved from NCBI. Values closer to 0 (no overlap = niche divergence) are purple while values closer to 1 (complete overlap = niche convergence) are orange; B) Occurrence points for each taxa grouped based on the first three principle components of biophysical and bioclimatic variables. Clusters share homogeneous bioclimatic and biophysical conditions.



Provisional

Figure 4.TIF

Figure 4. Climatic niches of *Helianthus* crop wild relatives for A) Mean diurnal range and annual precipitation, B) Soil pH and mean annual precipitation, C) Mean diurnal range and annual precipitation. Niches per taxa represent the middle 90% of occurrence points, i.e., 10% outliers are not included. Red boxes show the niche of wild *H. annuus* and black boxes show the niche of cultivated *H. annuus* in North America.



Provisional

Figure 5.TIF

Figure 5. Test of phylogenetic signal utilizing the K for 25 of 36 taxa analyzed with complete genetic and environmental information (Blomberg, et al, 2003). K measures phylogenetic signal in traits, where K values below 1 indicates low dependence of traits on evolutionary history (not conserved between taxa) and K values above 1 indicates trait conservation over evolutionary history (traits conserved over evolutionary time). \*indicates K significantly greater than 1 ( $p < 0.05$ ).

