frontiers
in Plant Science

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

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

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

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

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

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

[1]Biodiversity Research Centre and Department of Botany, University of British Columbia, 3529-6270 University Boulevard, Vancouver, British Columbia V6T 1Z4, Canada

[2]Department of Agronomy and Plant Genetics, University of Minnesota, 411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108

[3]International Center for Tropical Agriculture (CIAT), Km 17, Recta Cali-Palmira, Apartado Aéreo 6713, Cali, Colombia

[4]Centre for Crop Systems Analysis, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, Netherlands

[5]School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

[6]Department of Horticulture and Crop Science, The Ohio State University, 202 Kottman Hall, 2021 Coffey Rd, Columbus, Ohio 43210, USA.

[7]Department of Ecology and Evolutionary Biology, University of Colorado at Boulder, Boulder, CO, USA

[8] North Central Regional Plant Introduction Station, Agronomy Department, Iowa State University and USDA-ARS, Ames, IA 50014, USA

[9]USDA-ARS, Northern Crop Science Laboratory, Fargo, ND 58102, USA

[10]Department of Biology, Indiana University, Bloomington, IN 47405, USA

[11]To whom correspondence should be addressed

Correspondence:
Dr. Michael B. Kantar

Department of Agronomy and Plant Genetics

University of Minnesota

411 Borlaug Hall

1991 Upper Buford Circle

St. Paul, MN 55108

kant0063@umn.edu

Abstract

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

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

Introduction

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

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

The genus *Helianthus* L. contains 52 species comprising 67 taxa (Schilling, 2006; Stebbins *et al*., 2013). Native to North America, the taxa occupy a variety of habitats ranging from open plains to salt marshes (Kane *et al.*, 2013; Seiler & Marek, 2011). Sunflower (*Helianthus annuus* L.) is the most economically important species from the genus, with ~26 million hectares in production worldwide and a substantial private sector breeding effort, particularly for oil production (FAOSTAT, 2013). Domesticated approximately 4000 years ago in east central North America, sunflower has a typical domestication syndrome; i.e., it does not branch, does not have seed dormancy, has a predictable flowering time, and does not shatter (Harlan *et al.*, 1973; Harter *et al.*, 2004; Blackman *et al.* 2011). The crop has undergone both selection and genetic drift during domestication and improvement, which has reduced genetic diversity (Liu & Burke 2006; Tang & Knapp 2003), with modern cultivars retaining 50-67% of the diversity present in wild *H. annuus* populations (Kolkman *et al.* 2007; Mandel *et al.*, 2011).

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

Materials and Methods

Species distribution modeling

A modified gap analysis (Ramírez-Villegas *et al.*, 2010) was used to determine the conservation status of 36 taxa within *Helianthus* selected based upon their potential to provide useful traits for sunflower breeding. Briefly, 1) target taxa were identified, and geographic occurrence data were gathered and verified, 2) the overall representation of CWR in germplasm collections was estimated, 3) potential distribution models were produced for taxa with sufficient samples with coordinates, 4) the geographic and ecological representation of germplasm collections were assessed for each taxon by comparing potential distribution models to existing germplasm collection locations, 5) taxa were prioritized for further collecting based upon the average of their overall, geographic, and ecological coverage results, and 6) gap analysis results were correlated with the subjective assessments of collection priorities from crop experts.

The selection of taxa for analysis was based on membership within the primary or secondary genepools of sunflower (Vincent *et al.*, 2013) with the addition of all taxa from the tertiary genepool indicated in publications to be confirmed or potential trait donors (Table S1). A total of 12,737 occurrence records for the 36 taxa, sourced from 31 herbaria and five genebanks, were used for distribution models and conservation analysis (Table S2), including 4,705 records with geographic coordinates. The overall representation of taxa in genebank collections was estimated using the ‘Sampling Representativeness Score’ (SRS), calculated as the number of germplasm samples (GS) divided by the total number of samples (GS plus reference records). After eliminating duplicate records, potential distributions were calculated using Maxent (Phillips *et al.*, 2006), with a k-5 cross-validation option and 10,000 background points for model training over North America (Phillips, 2008; VanDerWal *et al.*, 2009). We included nineteen bioclimatic variables derived from the WorldClim database (Nix, 1986; Hijmans *et al.*, 2005a; Hijmans *et al.*, 2005b), seven biophysical variables from the ISRIC – World Soil Information database (http://soilgrids1km.isric.org) at a resolution of 2.5 arc-minutes, and the occurrence information (coordinates) for each taxon as inputs (Table S3). For edaphic data we calculated a weighted mean from five depths (0 to5 cm, 5 to15 cm, 15 to30 cm, 30 to60 cm, 60 to100 cm) to generate a single value for the first meter of soil for each layer, and then resampled the data from 1 arc minutes to 2.5 arc minutes resolution to match the WorldClim dataset, using the raster package in R and ArcGIS Desktop 10.1 (Hengl *et al.*, 2014). Distributions were further restricted by applying a taxon independent threshold, based on the Receiver Operating Characteristic (ROC) curve (Liu *et al.*, 2005). GRIN distribution data was used to ensure that taxa distributions were not overinflated beyond known native boundaries (GRIN, 2012). Soil cover data from GlobCover 2009 (Global Land Cover Map) (http://due.esrin.esa.int/globcover/) further refined the maxent outputs and collecting maps by excluding urban areas, water bodies, bare areas, and permanent snow and ice regions.

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

Geographic representativeness of taxa in genebank collections was calculated using the 'Geographic Representativeness Score' (GRS), comparing the spatial overlap of a circular buffer surrounding each accession record (50 Km radius as described in Hijmans *et al.*, 2001) against the potential distribution of the taxon. Ecological gaps in genebank collections were calculated using the 'Ecological Representativeness Score' (ERS), calculated by comparing records to the full environmental range of the modeled taxon across ecosystem types (Olson *et al.*, 2001). The overall priority for further collecting for *ex situ* conservation for each taxon was determined by averaging the SRS, GRS, and ERS with equal weight to obtain a final prioritization score (FPS), classified according to the following ranges: 1., high priority (FPS between 0 and 3); 2., medium priority (FPS between 3.01 and 5); 3., low priority (FPS between 5.01 and 7.5); and 4., and well conserved taxa (FPS between 7.51 and 10).

Expert evaluation of conservation assessment results

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

Ecogeographic niche overlap and phylogenetic analyses

Potential distribution probability outputs were used when Maxent models performed well and CA50 sample buffers when Maxent models did not pass the validation criteria, to calculate niche overlap based on Schoener's D and Hellinger's I as outlined in Warren *et al.* 2008, and implemented in the R package Phyloclim (Heibl, 2011). Both indices utilize probability distributions in geographic space, with statistics ranging from 0 (no niche overlap) to 1 (complete niche overlap). First pairwise niche overlap was examined, then niche overlap between allopatric/sympatric taxa separately, annual/perennial taxa separately, and lastly allopatric/sympatric sister taxa. Geographic range overlap for all pairwise combinations (630 comparisons) was calculated in two ways, with respect to the larger range [(2*number of shared grid cells)/(number of grid cells in taxa A + number of grid cells in taxa B)] and with respect to the smaller range [(2*number of shared grid cells) / (Total number of grid cells in taxa A + Total number of grid cells in taxa B)] / (Total potential number of shared grid cells) [2*total number of grid cells in species with the smaller range})/(Total number of species A + Total number of species B].

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

We downloaded the publically available 18S-26S Ribosomal DNA sequence from the external transcribed spacer (ETS) from GenBank (NCBI-http://www.ncbi.nlm.nih.gov/) for 28 of the 36 *Helianthus* taxa, aligned the sequences using ClustalW, and constructed a maximum likelihood phylogeny with 1000 bootstrap replications, using MEGA6 with a Jukes-Cantor nucleotide substitution model (Tamura *et al.*, 2013). We performed a Mantel test in R utilizing the ade4 package to explore the relationship between geography and genetics (Dray & Dufour, 2007). We estimated phylogenetic signal of individual ecogeographic traits utilizing Blomberg's K (Blomberg, et al, 2003), using the multiphylosignal command with 1000 permutations in Picante (Kembel *et al.*, 2010).

Results

Geographic distributions of sunflower crop wild relatives

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

Ecological niches of sunflower crop wild relatives

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

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

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

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

Discussion

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

Implications for conservation and plant breeding

Our approach is both new and complementary to previous work on *Helianthus* species distributions and CWR in the literature (Thompson *et al.*, 1981; Rogers *et al.*, 1982). The method of constraining ranges to known native distributions may have limited our identification of some the extreme variation. Despite this, many taxa that diverge ecologically from cultivated sunflower were identified (Fig. 4; Table 1). It was also possible to identify extreme populations within taxa that showed potential adaptation to different ecological niches.

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

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

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

Range overlap of wild relatives of sunflower

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

Unlike sympatric congeners in other plant groups (Grossenbacher *et al.*, 2014; Anacker & Strauss, 2014), *Helianthus* sister taxa typically lack strong ecological divergence. This observation is inconsistent with most models of speciation involving gene flow, which assume divergent ecological selection (Via, 2009). Possibly, our analyses lacked sufficient resolution or focus on key ecological attributes to detect real differences between the ecological niches of these species. For example, it is possible that there has been pollinator and phenological divergence between sister species that was not included in our analyses. Alternatively, local niche differences between sympatric populations may have been masked by substantial ecological heterogeneity among populations of the more widely ranging species. Additionally, the approach used was designed to analyze potential habitat in the historical, native range, rather than recent range expansions, which in many cases may be recent introductions facilitated by humans, perhaps accounting for observations of limited ecological divergence.

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

Conclusions

Using a combination of gap analysis, environmental niche modeling and phylogenetic approaches 36 CWR of sunflower were examined. Taxa that are under-represented in germplasm collections as well as species and populations inhabiting environmental niches with extreme phenotypes that may possess traits of value to crop improvement were identified. In *Helianthus,* sister taxa appear to occur more frequently in sympatry than allopatry, possibly suggesting that speciation may occur in the presence of gene flow. Finally, much of the primary genepool occurs in extreme environments indicating that utilization of wild *H. annuus* for the breeding of abiotic stress tolerance may produce quick gains with minimal effort.

Acknowledgments

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

References

Anacker, B.L., and Strauss, S.Y. (2014). The geography and ecology of plant speciation: range overlap and niche divergence in sister species. *Proc. R. Soc. B*, 281, 20132980. doi:10.1098/rspb.2013.2980

Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Sci*, 48, 1649-1664.

Blackman B.K., Scascitelli M., Kane N.C., Luton H.H., Rasmussen D.A., Bye R.A. et al. (2011). Sunflower domestication alleles support single domestication center in eastern North America. *Proc Natl Acad Sci USA*, 108, 14360-14365.

Blomberg S.P., Garland T., and Ives A.R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*, 57, 717–745.

Bowers, J. E., Nambeesan, S., Corbi, J., Barker, M.S., Rieseberg, L.H., Knapp, S.J. et al. (2012). Development of an Ultra-Dense Genetic Map of the Sunflower Genome Based on Single-Feature Polymorphisms. *PloS One*, 7, e51360.

Butler, E.E., and Huybers, P. (2013). Adaptation of US maize to temperature variations. *Nat Clim Change*, 3, 68-72.

Cabrera-Bosquet, L., Crossa, J., von Zitzewitz, J., Serret, M.D., and Luis Araus, J. (2012). High-throughput phenotyping and genomic selection: The frontiers of crop breeding converge. *Journal of integrative plant biology*, 54, 312-320.

Challinor, A.J., Watson, J., Lobell, D.B., Howden, S.M., Smith, D.R., and Chhetri, N. (2014). A meta-analysis of crop yield under climate change and adaptation. *Nat Clim Change*, 4, 287-291.

Chandler J.M., Jan C., and Beard B.H. (1986). Chromosomal differentiation among the annual *Helianthus* species. *Systematic Botany*, 11, 354-371.

Davies, T.J., Kraft, N.J.B., Salamin, N., and Wolkovich, E.M. (2012). Incompletely resolved phylogenetic trees inflate estimates of phylogenetic conservatism. *Ecology*, 93, 242–247.

Dray, S. and Dufour, A.B. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22, 1-20.

Fang, Z., Gonzales, A.M., Clegg, M.T., Smith, K.P., Muehlbauer, G.J., Steffenson, B.J. et al., (2014) Two genomic regions contribute disproportionately to geographic differentiation in wild barley. *G3: Genes| Genomes| Genetics*, 4, 1193-1203.

Fassler, E., Robinson, B.H. Stauffer, W., Gupta, S.K., Papritz, A., and Schulin, R. (2010). Phytomanagement of metal contaminated agricultural land using sunflower, maize and tobacco. *Agriculture Ecosystems and Environment*, 136, 49–58.

FAOSTAT. *Final Data 2013*. Retrieved May, 2015. http://faostat.fao.org.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Revell LJ. 2009. Size-correction and principal components for interspecific comparative studies. *Evolution* 63: 3258-326

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Tester, M., and Langridge, P. (2010). Breeding Technologies to Increase Crop Production in a Changing World. *Science*, 327, 818.

Thompson, T. E., Zimmerman, D. C., and Rogers, C. E. (1981). Wild *Helianthus* as a genetic resource. *Field Crops Research*, 4, 333-343.

Timme, R. E., Simpson, B. B., and Linder, C. R. (2007). High-resolution phylogeny for *Helianthus* (*Asteraceae*) using the 18S-26S ribosomal DNA external transcribed spacer. *American Journal of Botany*, 94, 1837-1852.

Uyeda J.C., Caetano D.S., Pennell M.W. 2015. Comparative analysis of principal components can be misleading. *Systematic Biology* 64: 677-689.

VanDerWal, J., Shoo, L.P., Graham, C.H., Williams, S.E. (2009). Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecological Modeling*, 220, 589-594.

Via, S. (2009). Natural selection in action during speciation. *Proc Natl Acad Sci USA*, 106, 9939-9946.

Vincent, H., Wiersema, J., Kell, S., Fielder, H., Dobbie, S., Castañeda-Álvarez, N. P., et al. (2013). A prioritized crop wild relative inventory to help underpin global food security. *Biological Conservation*, 167, 265-275.

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

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

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

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

	Perfect Overlap (%)	D or I Greater than 0.5 (%)	D or I Less than 0.2 (%, Divergent Niche)
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

Figure Legends

Fig. 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).

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

Fig. 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.

Fig. 4 Climatic niches 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.

Fig. 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).

Fig. 1

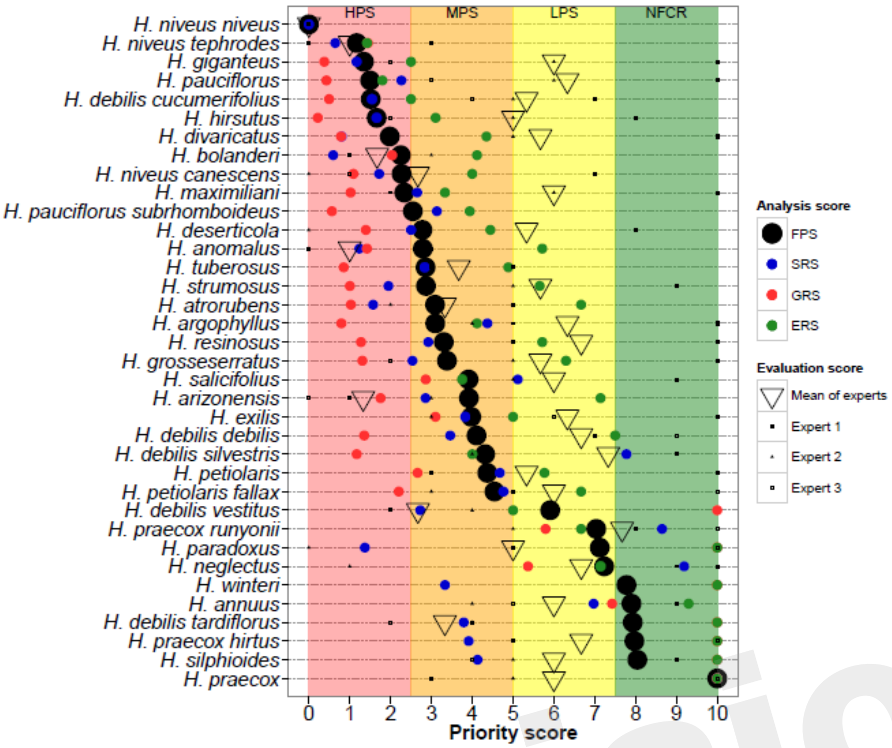

H. niveus niveus
H. niveus tephrodes
H. giganteus
H. pauciflorus
H. debilis cucumerifolius
H. hirsutus
H. divaricatus
H. bolanderi
H. niveus canescens
H. maximiliani
H. pauciflorus subrhomboideus
H. deserticola
H. anomalus
H. tuberosus
H. strumosus
H. atrorubens
H. argophyllus
H. resinosus
H. grosseserratus
H. salicifolius
H. arizonensis
H. exilis
H. debilis debilis
H. debilis silvestris
H. petiolaris
H. petiolaris fallax
H. debilis vestitus
H. praecox runyonii
H. paradoxus
H. neglectus
H. winteri
H. annuus
H. debilis tardiflorus
H. praecox hirtus
H. silphioides
H. praecox
HPS
MPS
LPS
NFCR
0
1
2
3
4
5
6
7
8
9
10
Priority score
Analysis score
FPS
SRS
GRS
ERS
Evaluation score
Mean of experts
Expert 1
Expert 2
Expert 3

Fig. 2.

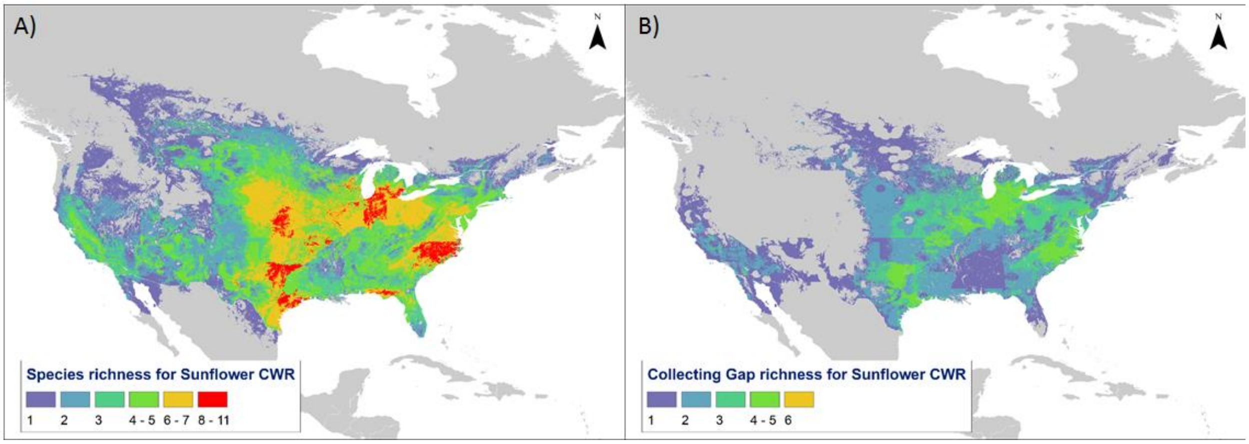

A)
N
Species richness for Sunflower CWR
1
2
3
4 - 5
6 - 7
8 - 11
B)
N
Collecting Gap richness for Sunflower CWR
1
2
3
4 - 5
6

Fig. 3

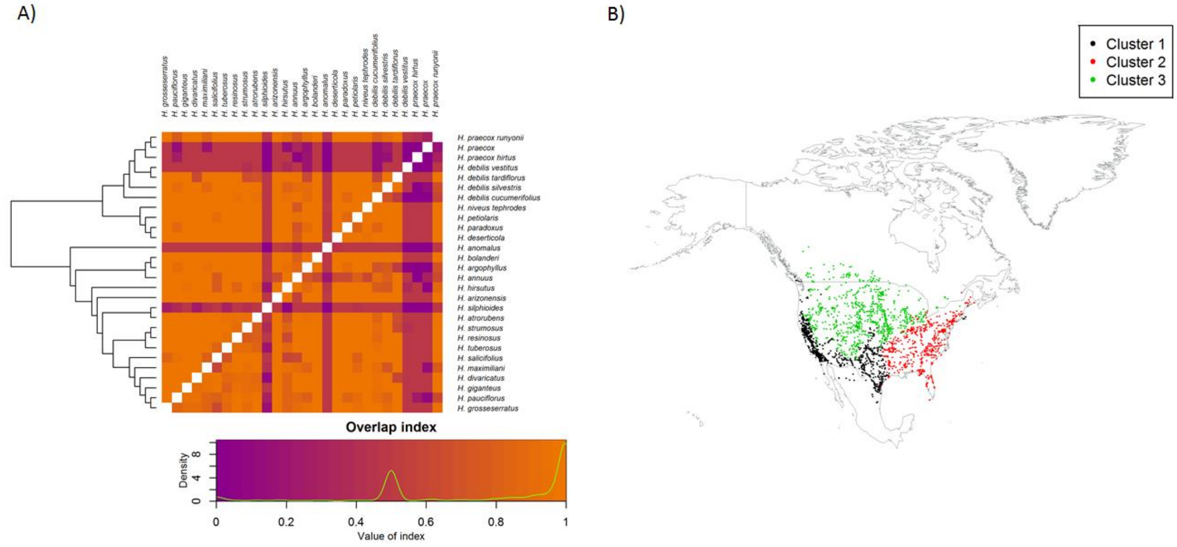

A)
H. grosseserratus
H. pauciflorus
H. giganteus
H. divaricatus
H. maximiliani
H. salicifolius
H. tuberosus
H. resinosus
H. strumosus
H. atrorubens
H. silphioides
H. arizonensis
H. hirsutus
H. annuus
H. argophyllus
H. bolanderi
H. anomalus
H. deserticola
H. paradoxus
H. petiolaris
H. niveus tephrodes
H. debilis cucumerifolius
H. debilis silvestris
H. debilis tardiflorus
H. debilis vestitus
H. praecox hirtus
H. praecox
H. praecox runyonii
H. praecox runyonii
H. praecox
H. praecox hirtus
H. debilis vestitus
H. debilis tardiflorus
H. debilis silvestris
H. debilis cucumerifolius
H. niveus tephrodes
H. petiolaris
H. paradoxus
H. deserticola
H. anomalus
H. bolanderi
H. argophyllus
H. annuus
H. hirsutus
H. arizonensis
H. silphioides
H. atrorubens
H. strumosus
H. resinosus
H. tuberosus
H. salicifolius
H. maximiliani
H. divaricatus
H. giganteus
H. pauciflorus
H. grosseserratus
Overlap index
Density
0
4
8
0
0.2
0.4
0.6
0.8
1
Value of index
B)
● Cluster 1
● Cluster 2
● Cluster 3

Fig. 4

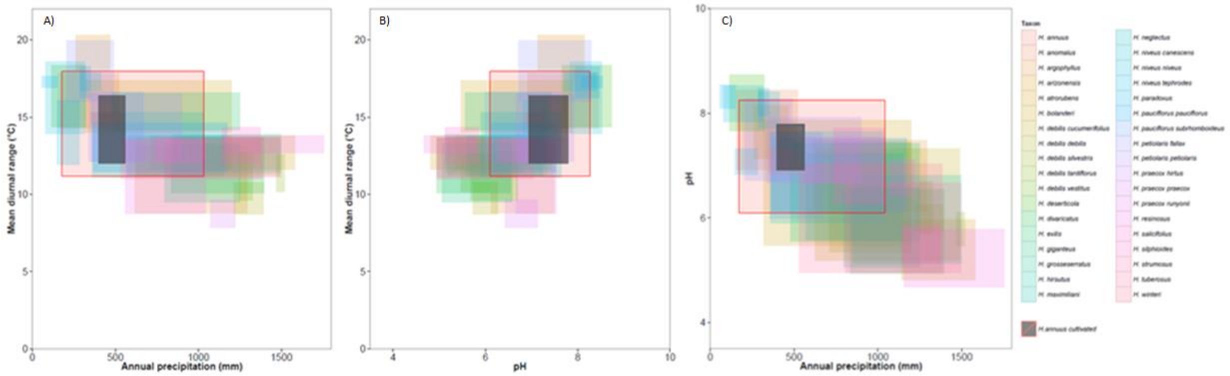
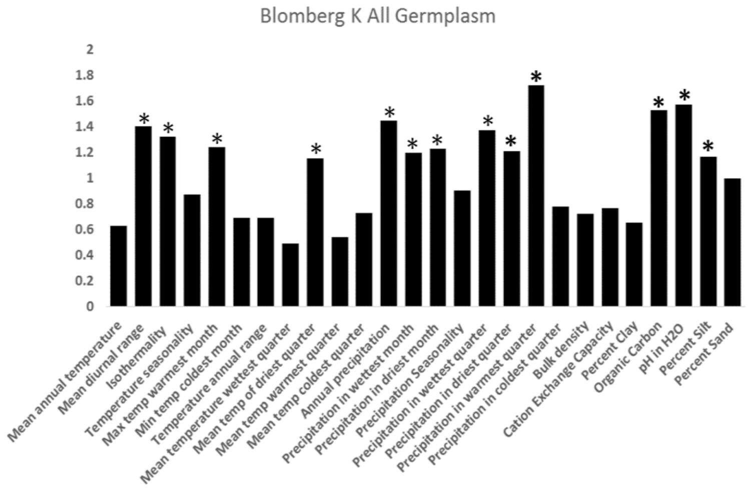

A)
B)
C)
Mean diurnal range (°C)
Annual precipitation (mm)
pH
0
5
10
15
20
500
1000
1500
4
6
8
10
Taxon
H. annuus
H. anomalus
H. argophyllus
H. arizonensis
H. atrorubens
H. bolanderi
H. debilis cucumerifolius
H. debilis debilis
H. debilis silvestris
H. debilis tardiflorus
H. debilis vestitus
H. deserticola
H. divaricatus
H. exilis
H. giganteus
H. grosseserratus
H. hirsutus
H. maximiliani
H. neglectus
H. niveus canescens
H. niveus niveus
H. niveus tephrodes
H. paradoxus
H. pauciflorus pauciflorus
H. pauciflorus subrhomboideus
H. petiolaris fallax
H. petiolaris petiolaris
H. praecox hirtus
H. praecox praecox
H. praecox runyonii
H. resinosus
H. salicifolius
H. silphioides
H. strumosus
H. tuberosus
H. winteri
H.annuus cultivated

Fig. 5


Blomberg K All Germplasm
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Mean annual temperature
Mean diurnal range
Isothermality
Temperature seasonality
Max temp warmest month
Min temp coldest month
Temperature annual range
Mean temperature wettest quarter
Mean temp of driest quarter
Mean temp warmest quarter
Mean temp coldest quarter
Annual precipitation
Precipitation in wettest month
Precipitation in driest month
Precipitation Seasonality
Precipitation in wettest quarter
Precipitation in driest quarter
Precipitation in warmest quarter
Precipitation in coldest quarter
Bulk density
Cation Exchange Capacity
Percent Clay
Organic Carbon
pH in H2O
Percent Silt
Percent Sand

Supplementary information

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

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

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

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

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

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

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

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

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

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

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

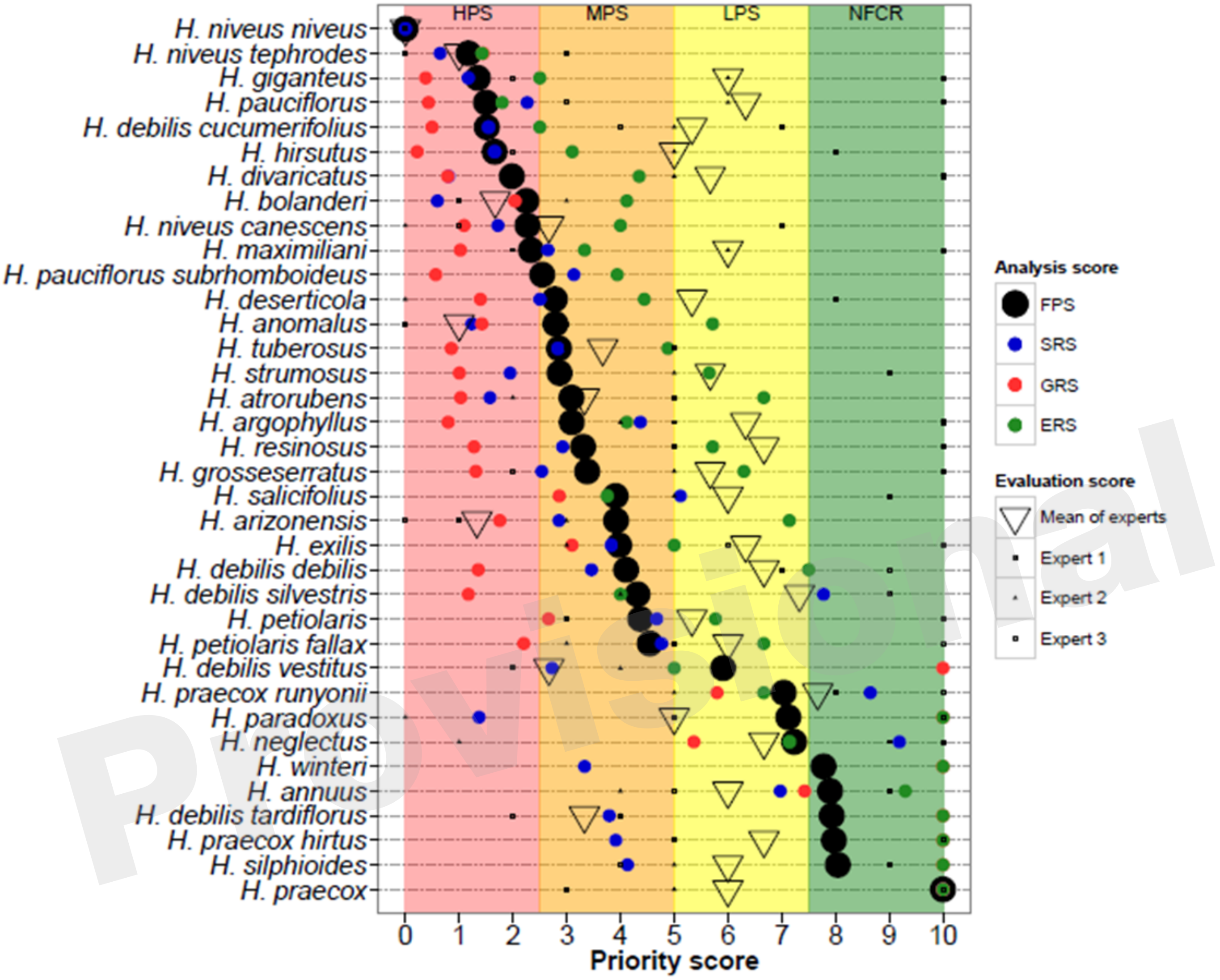


Figure 2.TIF

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.

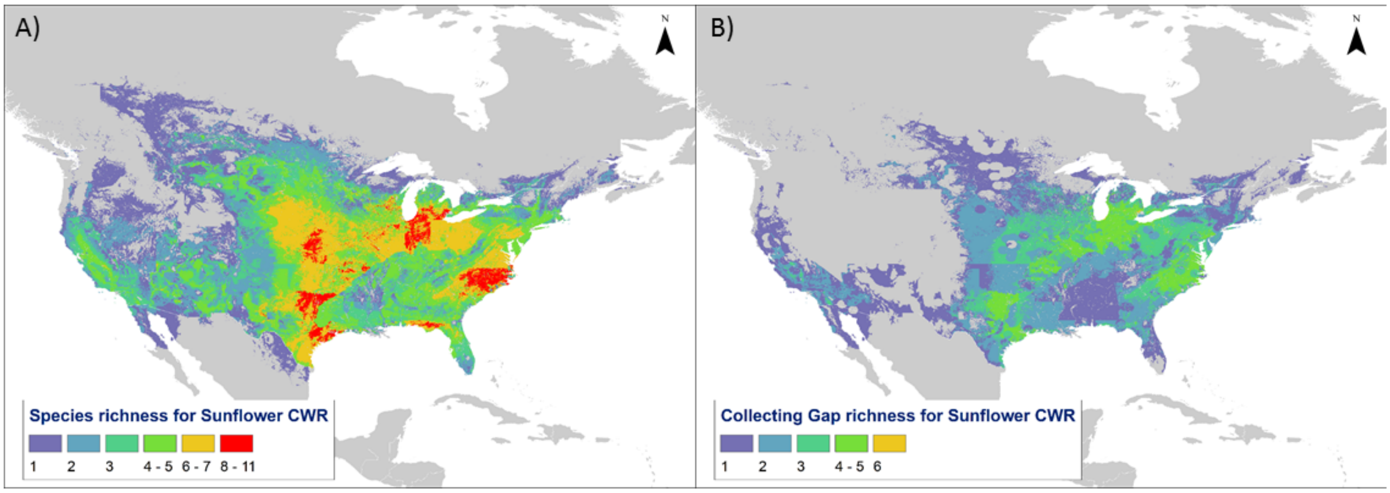

A)
N
Species richness for Sunflower CWR
1
2
3
4 - 5
6 - 7
8 - 11
B)
N
Collecting Gap richness for Sunflower CWR
1
2
3
4 - 5
6

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.

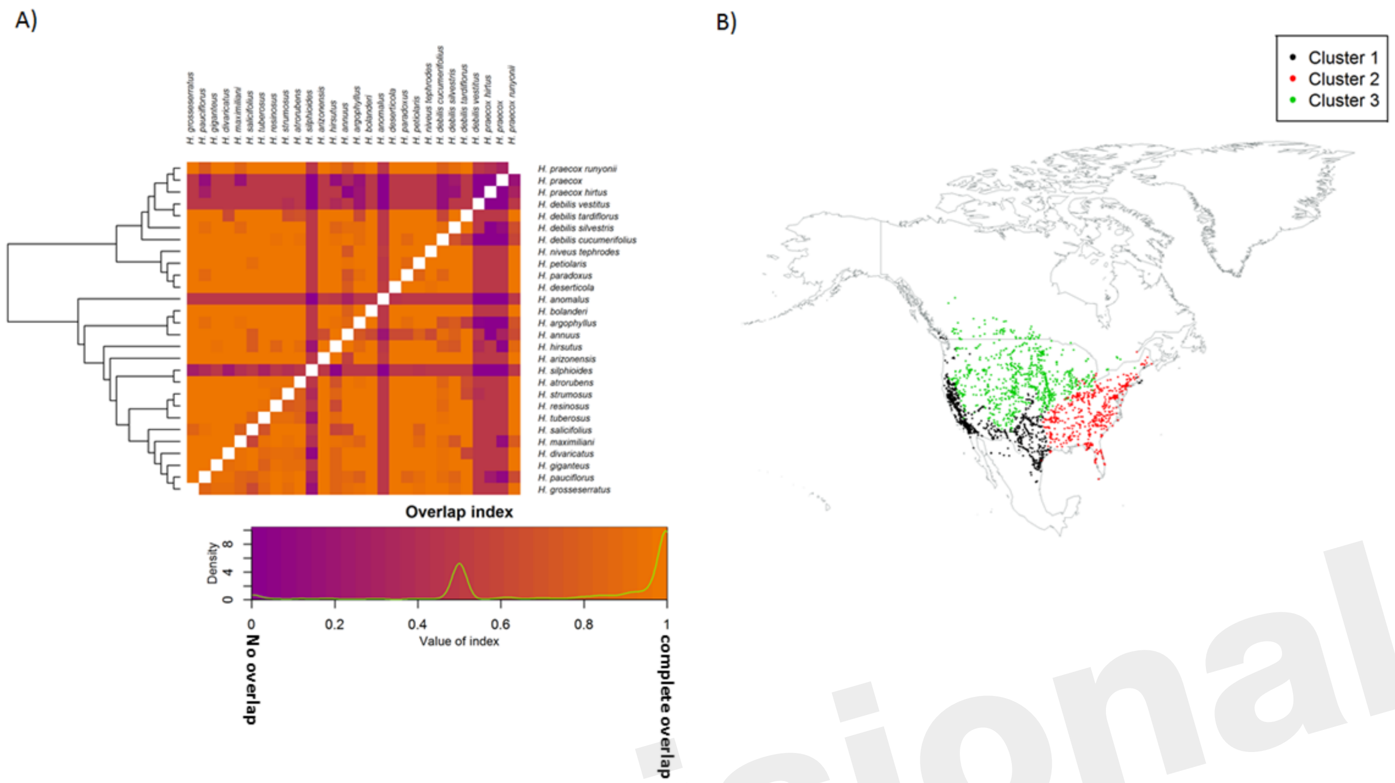


Figure 4.TIF

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.

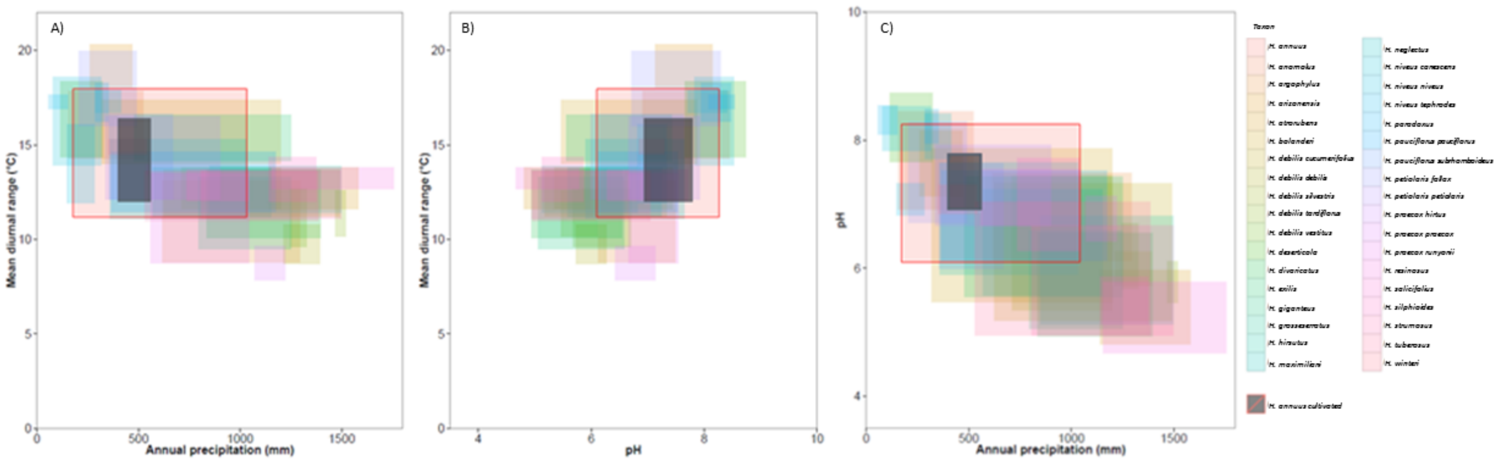
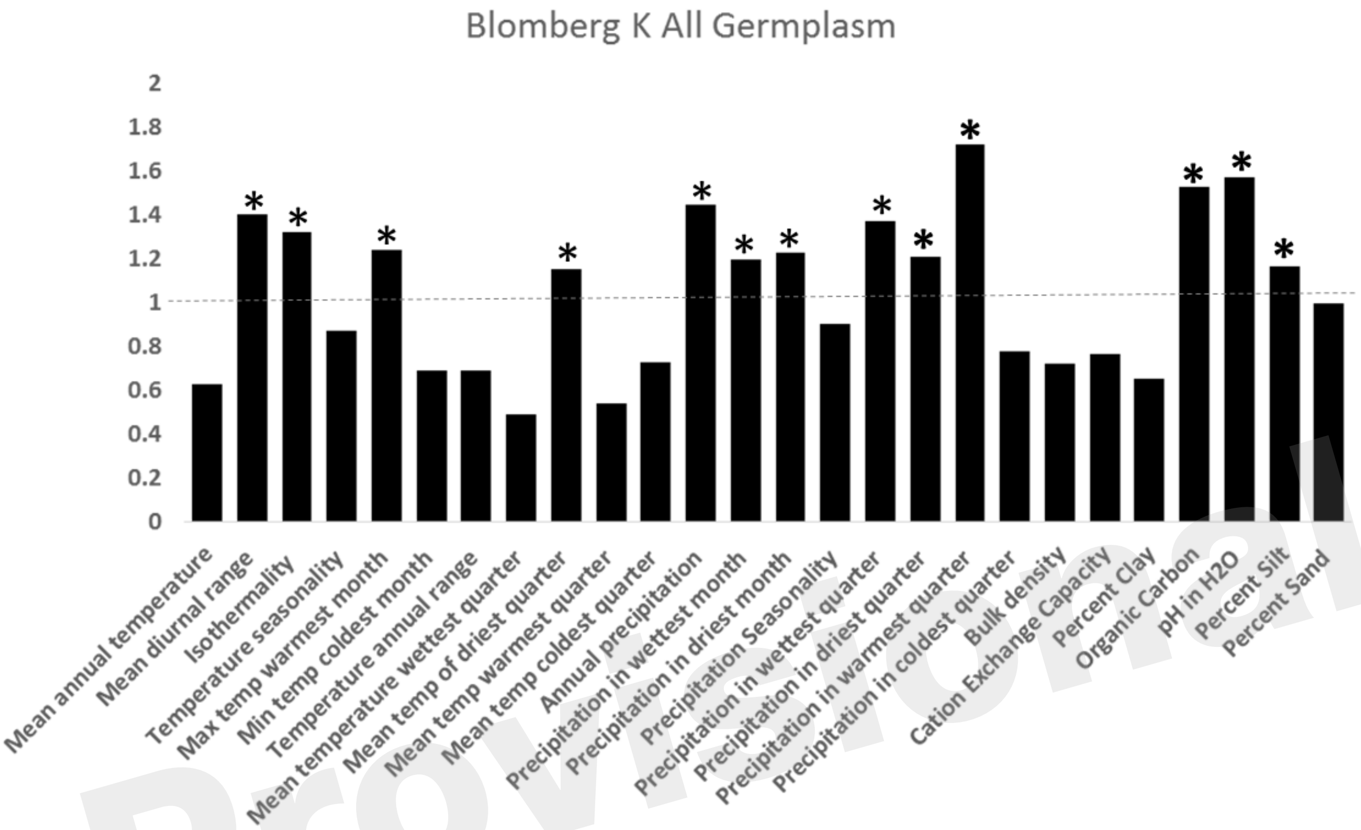


Figure 5.TIF

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).


Blomberg K All Germplasm
*
Mean annual temperature
Mean diurnal range
Isothermality
Temperature seasonality
Max temp warmest month
Min temp coldest month
Temperature annual range
Mean temperature wettest quarter
Mean temp of driest quarter
Mean temp warmest quarter
Mean temp coldest quarter
Annual precipitation
Precipitation in wettest month
Precipitation in driest month
Precipitation Seasonality
Precipitation in wettest quarter
Precipitation in driest quarter
Precipitation in warmest quarter
Precipitation in coldest quarter
Bulk density
Cation Exchange Capacity
Percent Clay
Organic Carbon
pH in H2O
Percent Silt
Percent Sand