Ultra-barcoding of 154 Heliantheae complete chloroplast genomes gives insight into the phylogeny and evolution of the endangered sunflower, *Helianthus schweinitzii*

By
Eve Kruger
Ecology and Evolutionary Biology, University of Colorado at Boulder

Defense Date:
March 28, 2022

Thesis Advisor:
Dr. Nolan Kane,
Department of Ecology and Evolutionary Biology

Defense Committee:
Dr. Nolan Kane, Department of Ecology and Evolutionary Biology
Dr. Barbara Demmig-Adams, Department of Ecology and Evolutionary Biology
Dr. Gifford Miller, Department of Geological Sciences

Abstract

Asteraceae is considered the largest plant family, in terms of named species, with more than 1,620 genera and 23,600 described species of plants distributed worldwide. The genus *Helianthus* is native to North America and gives the name to the tribe Heliantheae, which includes 190 closely related genera. In recent years, numerous studies have been conducted to better understand the evolutionary relationships within the genus *Helianthus*. However, uncertainties remain within this family surrounding the connections between genomics and conservation. The investigation into *Helianthus* has the potential to provide more information on future conservation efforts and guide future studies, such as the adaptive capacity a species and constrictions to species distributions. This study presents the assembly of a new chloroplast genome for the endangered sunflower species *Helianthus verticillatus* and improves upon the assembly and annotation of the chloroplast genomes of several related species, including the endangered species *H. schweinitzii* and *H. paradoxus.* Using phylogenetic methods, comparative genomics, and related approaches, the comparison of the chloroplast genomes from these and 151 other members of the Heliantheae tribe clarifies the relationships among species and identifies major evolutionary shifts in the sequence or context of their chloroplast genomes. Additionally, this study provides further analysis into the endangered species *Helianthus schweinitzii* and its phylogenetic relationships.

Thesis: Established phylogenetic relationships will be present, and this analysis will further clarify relationships within Heliantheae species, specifically relationships of the rare and endangered species in this dataset.

1.0 Introduction

1.1 Study System & Research Focus

Within the largest plant family, Asteraceae, the tribe Heliantheae includes approximately 190 genera and 2,500 species (Karis, 1993). One of many common names for the Heliantheae family is the sunflower tribe, from the Greek word "helios", meaning "sun", and "anthos", meaning "flower". Many of these species inhabit North America and South America, with few from tropical areas (Karis, 1993). Commonly herbs or shrubs, other typical characteristics within Heliantheae include hairy leaves arranged in opposite pairs and black anthers (Karis, 1993). The two key genera that are the focus of this thesis are *Silphium* and *Helianthus.*

I compiled a dataset that comprises the chloroplast genomes of over 150 species mostly from the *Helianthus* genus, at least three of which are endangered, with other related species included. Although 142 of these genomes are already published in GenBank and NCBI (National Center for Biotechnology Information), 12 of them are newly assembled chloroplast genomes completed by students at the University of Colorado, Boulder, and members of the Kane Lab, including myself; these genomes are analyzed here for the first time. We present the largest phylogenetic study of Heliantheae, and maybe even Asteraceae, that utilizes complete chloroplast genomes, to improve the understanding of origins and evolution of this group of species. This thesis presents a description of the genomes, their submission to NCBI, and a comparative analysis of the *Helianthus* tribe.

In this study, newly assembled complete chloroplast genomes of 12 *Helianthus* and *Silphium* specimens from 12 different species were compared against 142 genomes from the Heliantheae tribe with comparative genetics and phylogenetic analysis. These additional genomes increase the total number of publicly available completed genomes by 12 Heliantheae species. The majority of genomes were from *Helianthus*, a genus that consists of around 50 species of annual,

plants with life cycles lasting one growing season or year, and perennial, plants with life cycles that span multiple years, herbs. A variety of bioinformatics tools were used to assemble and annotate these genomes, including Unix, BLAST, Chlorobox, SRA toolkit, *bwa*, and *samtools* (Kent, 2001; Li et al., 2009; Li & Durbin, 2009; Madden, 2003).

Assembly and analysis of chloroplast genomes of the *Helianthus* is necessary for the ability to address questions of sunflower genetics, evolutionary relationships, conservation methods, and restoration strategies for these families. The examination of these relationships can clarify how these related species evolved and adapted to a wide range of environments. Specifically, the aim of my analysis was to confirm known and hypothesized hybridizations in Heliantheae and *Helianthus.* Further research into the Heliantheae tribe and the *Helianthus* genus of sunflowers, and an improved understanding of the genetics of endangered taxa within the genus will help identify avenues for species recovery and research. More specifically, this research has potential to aid in the development of successful sampling design for the management of endangered plants, informing extinction risk evaluations, the adaptive capacity of a species, and the constrictions to distribution (Primmer, 2009; Rossetto et al., 2021). This analysis will result in a resource paper for future research into the Heliantheae tribe, providing 12 new chloroplast genomes resources.

Specifically, this thesis explored the relationship of the rare and endangered *Helianthus schweinitzii* to the rest of the Heliantheae tribe. This particular species is an endangered tetraploid sunflower of hybrid origin that is native to the eastern United States, specifically the Piedmont Plateau of North and South Carolina (Anderson et al., 2019; Grubbs & Wynes, 2015). This species was suggested to have arisen via hybridization between the diploids *Helianthus giganteus* and *Helianthus microcephalus* on the basis of a comparative analysis of 38 accessions of six species (Anderson et al., 2019), which is not an exhaustive sampling of potential parent species. This thesis paper builds on the latter analysis and compares *Helianthus schweinitzii* to 153 species within the

Heliantheae tribe to clarify its sister taxa within a larger dataset. This broader and more diverse dataset of Heliantheae is used to further resolve the phylogenetic relationships of *H. schweinitzii* and clarify that of other species within the Heliantheae tribe. Additionally, while *H. schweinitzii* may be related to the close relatives identified in the earlier report (Anderson et al., 2019), its placement can be refined in the context of a larger data set. This thesis aims to further resolve origin and relationships of *Helianthus schweinitzii* as a species crucial to conservation efforts and understand its adaptive capabilities alongside its very narrow distribution range.

1.2 Ultra-barcoding Using Chloroplast Genomes

Chloroplasts are unique photosynthetic organelles of plant cells that play a crucial role in sustaining life on earth. Chloroplast genomes (cpDNA) contain approximately 120 genes on average that encode some of the key proteins involved in photosynthesis and associated metabolic processes (Daniell et al., 2016; Palmer, 1985). Chloroplast genomes tend to occur at much higher copy numbers than the nuclear genome (Possingham & Saurer, 1969), which makes them an inexpensive genomic resource to sequence relative to the cost of sequencing the nuclear genome. In recent years, the use of ultra-barcoding analysis, which uses whole plastomes and nuclear ribosomal DNA (nrDNA) sequences for plant species identification, to analyze whole chloroplast genomes has grown significantly (Kane et al., 2012). Ultra-barcoding of whole chloroplast genomes allows for the exact identification of a sample due to the chloroplast's unique sequence in each species (Kane et al., 2012). This expansion of sequencing and ultra-barcoding analysis of chloroplast genomes has added to the understanding of plant evolutionary relationships and informs the expanding field of conservation genomics (Yuan, 2021). Additionally, study of these genomes has made significant contributions to the phylogenetic understanding of the relationships within several plant families, including that of Heliantheae, and resolving evolutionary relationships within phylogenetic clades.

Chloroplast genomes are circular and relatively conserved among land plants in terms of structure, and gene content (Prehistory of the Angiosperms: Characterization of the Ancient Genomes - ScienceDirect, n.d.). They tend to vary between 120,000 and 170,000 base pairs in length (Shaw et al., 2007). The chloroplast can be divided into three functional categories: protein-coding genes (PCGs), transfer RNAs (tRNAs), and ribosomal RNAs (rRNAs). Protein-coding genes provide the information for protein synthesis via transcription into RNA and translation into proteins. Transfer RNAs (tRNAs) are transcribed into functional RNAs that facilitate translation of messenger RNA transcripts into proteins via the ribosomal ribonuclear protein complex (Giegé et al., 2012). Finally, ribosomal RNAs (rRNAs) form a complex with ribosomal proteins and help catalyze translation of messenger RNAs into proteins (Giegé et al., 2012).

Additional components of the chloroplast genome are introns that are transcribed along with non-coding regions of their associated proteins and then removed or spliced out as well as intergenic DNA regions between functional components of the genome (*Frontiers | The Function of Introns | Genetics*, n.d.; *Introns: Evolution and Function - ScienceDirect*, n.d.). Moreover, the chloroplast genomes of land plants contain two inverted repeats that divide the genome into four parts: a large, long single copy (LSC), the first inverted repeat, a short single copy (SSC) region, and the second inverted repeat (Clegg et al., 1994; Daniell et al., 2016; Palmer, 1985). All of these components of the genomic sequence can be used as molecular characters to infer plant phylogenies, molecular evolution, and population genetics (Clegg et al., 1994; Daniell et al., 2016; Palmer, 1985).

2.0 Material and Methods

2.1 Plant material

A variety of annual and perennial *Helianthus* species and closely related *Silphium* and other Heliantheae species were used. In this dataset, there are a total of 154 Heliantheae species, 84 of which are *Helianthus*.

2.2 Genome Assembly & Annotation of Each Species

For genomes that were assembled as a part of this study, publicly available short-read sequence archives were downloaded from the NCBI Sequence Read Archive (SRA; see *Table S1*) and split into fastq-formatted files using *fastq-dump –split-3* from the SRA-Toolkit. The fastq files were then aligned to a reference genome using *bwa-mem (Li & Durbin, 2009)*, for the entirety of this research *Helianthus annuus* (NC_007977.1) was the reference chloroplast genome for *Helianthus* species and *Silphium perfoliatum* (MN445187) for *Silphium* species. This alignment resulted in the output of SAM files, which were then converted to BAM files using *samtools* for indexing and sorting (Li et al., 2009). The sorted files were then revised for errors and consistency. The sorted files were made into a VCF file using *bcftools* and provided a summary of the coverage of mapped reads on a reference sequence at a single base-pair resolution of each species (Li, 2011).

These mapped reads were trimmed using Trimmomatic-0.39 with set parameters to remove poor quality reads. Once trimmed, the reads were assembled using SPAdes v.3.9 with standard parameters "*-k 35,55,85*" (Bankevich et al., 2012). This program generates contigs using a deBruijin graph algorithm from k-mers sizes specified by the user. Assembled chloroplast contigs were identified using a command-line BLAST against existing the appropriate reference genome

from NCBI. The contigs were elongated and connected to other contigs using the sequence in the trimmed fastq libraries until the full-length genome was in a single piece. Each genome was then circularized, ensuring the end of the circular genome connected directly to the sequence at the beginning. The genomes were then error-corrected by aligning the short-reads to the newly assembled sequence and visualized in *samtools tview* (Li et al., 2009) to ensure there were no consensus disagreements between the assembled sequence and the reads. Furthermore, assembly errors were identified as abrupt cutoffs in aligned reads, and sequences were corrected using the same process used to connect contigs, described above.

Annotations of the genomic features were started in the GeSeq program (Tillich et al., 2017) within the bioinformatics toolkit of Chlorobox with *H. annuus* (NC_007977.1) and *H. giganteus* (KF746346) for any *Helianthus* genomes, and *Silphium perfoliatum* (NC_060408.1) for *Silphium* as the reference genomes. Once GeSeq analysis had finished, a comparison of the annotation was conducted using SPAdes v.3.9, Chloe v0. 2.0, and BLAT (Bankevich et al., 2012; Kent, 2001; Tillich et al., 2017), on gene identity length against the reference genomes in order to determine if any errors were present in the annotations. Corrections of the annotation of each genome were completed by manually editing the 5-column table formatted annotation file, focusing on multi-exon genes. Manual corrections for the *Helianthus* and *Silphium* taxa included tRNA-UGC for alanine, rpoc1, rpoc2, and psbC. Once accomplished, the assembled chloroplast genomes were able to be visualized using Chlorobox (Tillich et al., 2017), as seen in the *Helianthus verticillatus* chloroplast genome map in *Figure 1*.

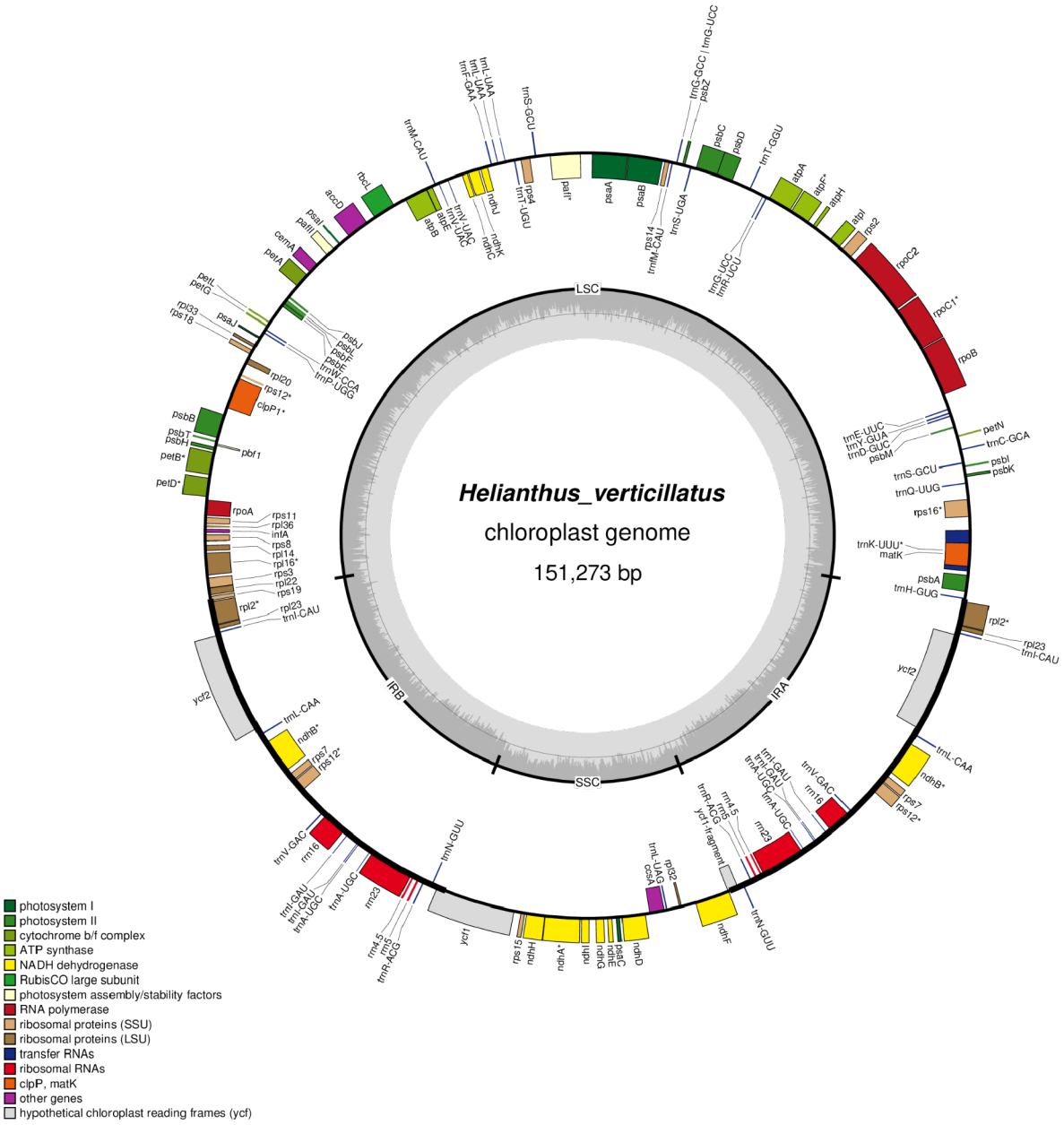


Figure 1: Chloroplast genome map for *Helianthus verticillatus*. The inverted repeat regions (IRa and IRb) are identified by the thicker lines on the outer circle of the map. On the outer region of the map, genes are inscribed in the clockwise direction on the outside and counterclockwise direction on the inside. The innermost region of the chloroplast genome represents G + C content (Walker et al., 2014).

After correcting common errors in the annotations, genome sequences and their annotations were submitted using NCBI's BankIt submission portal. This increased the total number of publicly available Heliantheae genomes by 12 on the NCBI database. See *Table S1*.

2.3 Comparative Genomics

To compare the content encoded in the chloroplast genomes, the novel sequence assembled for this study was aligned alongside 146 publicly available genomes (*see Table S1*). These genomes were compiled into a multi-fasta file and reoriented such that each sequence started at the canonical start of the chloroplast genome, at the junction between the second inverted repeat and the long single copy (IRb-LSC junction). Sequences were aligned in MAFFT 7.0 (Katoh & Standley, 2013) using default parameters for a nucleotide-based alignment. This program created a visual alignment of each species and a preliminary neighbor-joining tree. Alignments were visualized and curated using PhyDE v0.9971 to correct any obvious misalignments (Müller, 2016). The short single copy of five of the genomes, *Eclipta alba*, *Eclipta prostrata*, *Silphium perfoliatum* (MN445187), *Sphagneticola calendulacea*, and *Sphagneticola trilobata*, needed to be inverted to correctly align with the other genomes.

2.3.1 Size Comparisons

Genome size differences can reveal the variation between species in different regions, evolutionary rate differences, and real biological differences between genomes. To compare the sizes of the genomes, the total genome length was recorded for each species and compared to one another. These sizes are reported in *Table S1*.

Command-line functions were used to isolate the headers and determine genome size for each of the chloroplast genomes in the fasta file. After these commands were implemented, a completed file, *Genome_Lengths*, had been created that included the total size of each chloroplast genome.

2.3.2 GC Content

To review the GC content, the number of G and C base pairs in comparison to the whole genome, in chloroplast genomes, the percentage was calculated using command line functions to count the number of instances of G or C base pairs in each sequence and normalized by total sequence length. Once this was accomplished, the GC content percentage for each species was added to *Table S1*.

2.3.3 Statistical Analysis

Correlations between genome length and GC content percentage were tested with RStudio, using a linear model to compare genome length and GC content percentage and determine significance (RStudio Team, 2020).

2.3 Phylogenetic Analysis

In the phylogenetic analysis of these species of the *Heliantheae* tribe, a comparison of established phylogenetic relationships and clades was conducted. For this analysis, the genomes of the 154 Heliantheae species were unwrapped and aligned using MAFFT (Katoh & Standley, 2013).

Once this preliminary analysis of the genomes was completed, the multi-fasta file was then run through PhyDE, Phylogenetic Data Editor (Müller, 2016). The genomes were hand-annotated to put gaps together due to aligner error and improve bootstrap values. During this process, the

species chloroplast genomes were examined for the existence of clade-specific mutations. This allowed for clade-specific genes and base-pair differences to be clarified, which may have been missed in initial annotations and alignments until this point.

Using this final alignment of the genomes, preliminary maximum likelihood trees were constructed using Archaeopteryx and MEGA11 (Stamatakis, 2014; Tamura et al., 2021; Zmasek & Eddy, 2001). The outgroups of *Lactuca sativa* (lettuce)*, Lactuca tatarica* (blue lettuce), *Guizotia abyssinica* (Niger seed), *Cynara cardunculus var. scolymus* (artichoke)*, Cynara cornigera* (white artichoke), *Cynara baetica* (Moroccan artichoke), *Cynara humilis* (wild thistle), *Stevia Oliveira* (sweet leaf), *Ambrosia trifida* (giant ragweed), *Ambrosia artemisiifolia* (common ragweed), *Artemisia capillaris* (wormwood), *Artemisia lactiflora* (white mugwort), and *Artemisia argyi* (silvery wormwood) were used to root the final phylogenetic tree. The sequence data of these outgroups were obtained from the NCBI database; their accession numbers can be found in *Table S1*. Downloading the produced RAxML file, the final phylogeny was constructed using Interactive Tree of Life (iTOL), where clades are annotated to be color-coded for perennials, annuals, and species of interest and strong bootstrap values are present *(Figure S1; Letunic & Bork, 2007; Stamatakis, 2014)*.

2.3.1 Helianthus Schweinitzii Phylogenetic Analysis

The phylogenetic relationships of *Helianthus schweinitzii (*MW381301.1) were investigated after the completion of the finished phylogenetic tree. To do so, the genome of *H. schweinitzii* was compared against its sister taxa, *H. tuberosus Loeuville 793* (MT700541.1) and *H. tuberosus* (MT302562.1), using BLAST, to determine genome similarity (Kent, 2001; Madden, 2003). For further analysis and to generate figures, the genomes were then compared using zPicture (*Figures 4 & 5*; Ovcharenko et al., 2004).

3.0 Results

3.1 Comparative Genomics

The genomes of Heliantheae appeared to have the same structure and content, with few clade-specific differences *(Table S1).* There was high conservation among the 154 chloroplast genomes in the dataset, showing little variation across genome length and GC content percentage (*Table S1*). There was a significant negative relationship between genome length and GC content percentage in the *Helianthus* genus (*Figure 2*).

3.1.1 Size

An analysis of the genome size revealed that in this large number of genomes there is high conservation of size in these species of the *Helianthus* genus. The size of the chloroplast genomes of the 84 *Helianthus* species varies from 150,015 bp to 151,494 bp, with *H. cusickii* (MW366799.1) having the smallest, and *H. tuberosus* (MG696658.1) having the largest (*Table S1*). Typical *Helianthus* chloroplast genomes have a typical quadripartite structure, consisting of four parts, and contain an LSC region, SSC region, and two copies of the inverted repeat (IR). The average size among the chloroplast genomes in the 84 *Helianthus* species was found to be 151836.64 bp.

3.3.2 GC Content

The highest GC content percentage was 37.71% in *Helianthus cusickii* (MW366799.1) and the lowest was 37.55% in *Helianthus tuberosus* (MG696658.1). The average across these *Helianthus* genus genomes is 37.59%. GC content of *Helianthus schweinitzii* (MW381301.1) was 37.6%, which is 0.1% higher than the average within the investigated *Helianthus* genomes.

3.3.3 Relationship of Genome Length and GC Content Percentage

Genome length and GC content percentage of *Helianthus* genomes exhibited a significant negative relationship (*Figure 2*) (p-value of $2.2e^{-16}$).

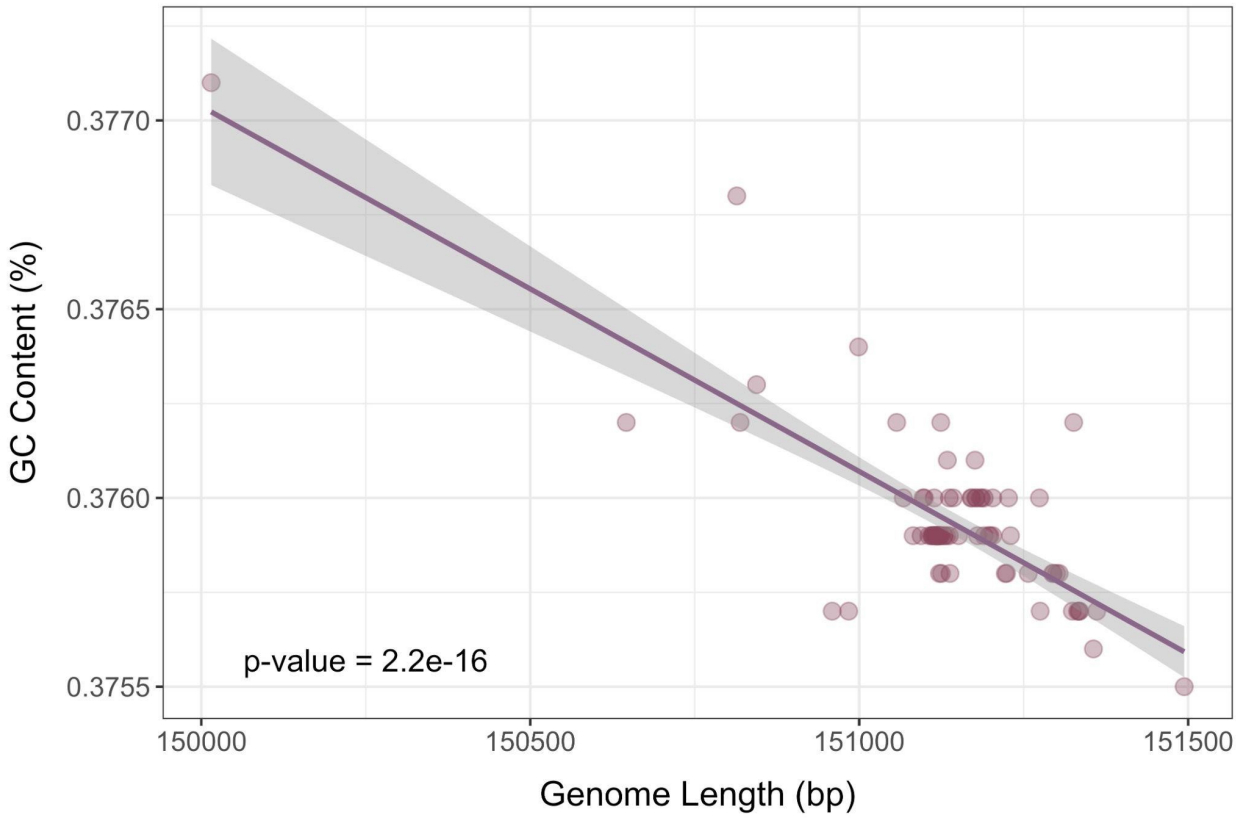


Figure 2: Scatterplot of the inverse relationship between genome length and GC content percentage in *Helianthus*.

3.2 Phylogenetic Analysis

This research confirmed many established phylogenetic relationships for the Heliantheae tribe and refined specific relationships of important groups and species of *Helianthus* and *Silphium*.

Additionally, the analysis of these species clarified the parental species of *Helianthus schweinitzii*.

3.2.1 Helianthus schweinitzii *Phylogenetic Analysis*

A phylogenetic tree was assembled to analyze the relationship of the rare and endangered *Helianthus schweinitzii*. In this phylogenetic tree (*Figure 3*), its sister taxa were identified as *Helianthus tuberosus* (MT302562.1) and *Helianthus tuberosus voucher Loeuille 793* (MT700541.1), with a bootstrap value of 76%. When compared against one another using NCBI's *blast*, *H. schweinitzii* showed 99.64% percent identity with *H. tuberosus voucher Loeuille 793* (*Figure 4*) and 99.73% percent identity with *H. tuberosus* (*Figure 5*).

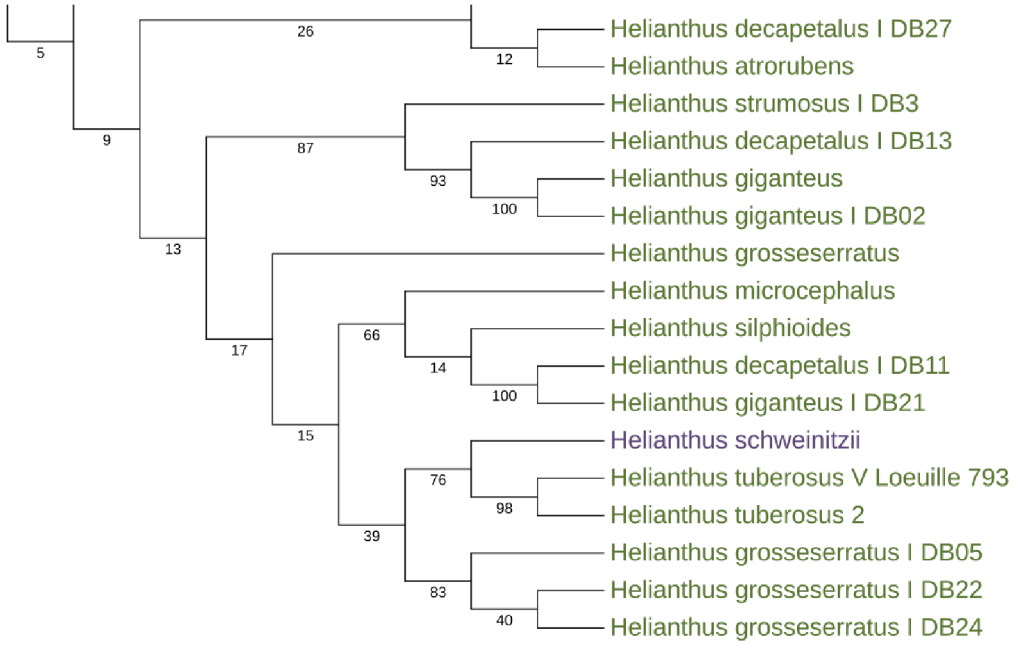
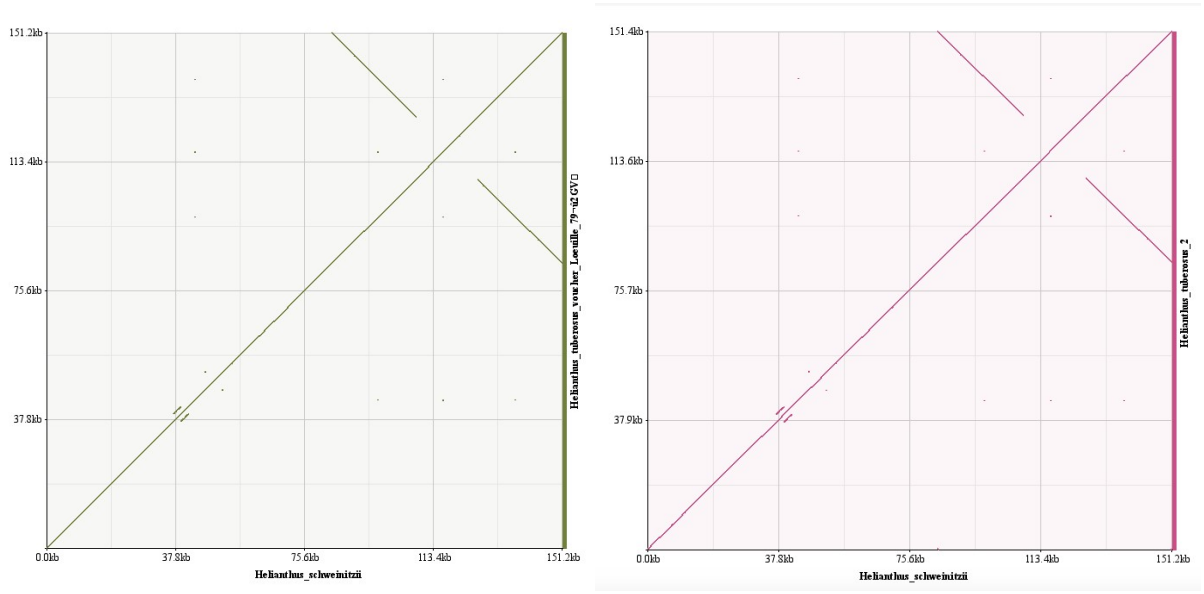


Figure 3: Portion of the phylogenetic tree (*Figure S1*) showing the relationship between *Helianthus schweinitzii* and its closest relatives, *Helianthus tuberosus* and *Helianthus grosseserratus*. The bootstrap value for the entire clade is 39.



Figures 4 (left) and 5 (right): zPicture visualizations of percent identity, a description of the similarity of query versus target sequence.

4.0 Discussion

4.1 Comparative Genomics and Phylogenetic Analysis

Members of Heliantheae tribe occur in diverse environments and habitats, with wide to narrow ranges (Karis, 1993). However, the genes that remain in the chloroplast appear to be relatively conserved in these Heliantheae genomes. Extreme conditions can result in clear adaptations to the environment, but despite adaptations to different geology and ecology, there is not much variation in these chloroplast genomes of Heliantheae. There is little variation in gene order or content, which provides insight into the differences between species and clades of Heliantheae regarding genome size and GC content. Although beyond the scope of this thesis, it would be of interest to investigate whether the same trend is true for (i) the many genes associated with photosynthesis and related processes that were transferred to the nuclear genome over the course of evolution and

(ii) the nuclear genes that exert strong control over chloroplast genes (*Gene Transfer from Organelles to the Nucleus: Frequent and in Big Chunks - PMC*, n.d.).

There was very little size variation of the chloroplast genomes in the genus of *Helianthus*, with *H. cusickii* (MW366799.1) exhibiting the smallest and *H. tuberosus* (MT302562.1) having the largest genome length. Similar to the findings of other studies on *Helianthus,* Heliantheae, and Asteraceae, the differences among chloroplast genome sizes are due to changes in length in non-coding regions and differences in boundary regions (Azarin et al., 2021; Loeuille et al., 2021; Walker et al., 2014). Even though there is little variation in the genome size of these *Helianthus* genomes, it concurs that the chloroplast genomes are conserved in size.

Analysis of GC content of the *Helianthus* genomes further confirmed the high level of conservation of the chloroplast genome with very little variation across the genomes. By comparison, the *Cymbidium* genus of the Orchid family exhibited a larger range in GC content of the chloroplast genome of 2.6% (*Gene Transfer from Organelles to the Nucleus: Frequent and in Big Chunks - PMC*, n.d.), which is fifteen times larger than the range found in *Helianthus* by as well as the present study. GC content, an important compositional feature of the genome, normally variation between chloroplast genomes exists due to genomic evolution (Singh et al., 2016; Šmarda et al., 2014; Talat & Wang, 2015). However, in *Helianthus* species, GC content percentage is extremely conserved, indicating that there is little evolutionary change across the chloroplast genomes over time.

The finding of an inverse relationship between genome length and GC content for the two *Helianthus* species with the largest relative difference in these parameters in this study may be due to the length of microsatellites, repetitive sequences of noncoding regions in the genome that can evolve quickly (Rubinsztein et al., 1995). The longer genome length may be associated with longer microsatellites, where the genome length increases due to the addition of As and Ts in

microsatellite regions (Rubinsztein et al., 1995). This correlation warrants further examination to identify and compare microsatellites across genomes for other datasets of *Helianthus*, Heliantheae, and Asteraceae.

Overall, phylogenetic analysis of the Heliantheae species in this dataset confirmed many established phylogenetic relationships for the Heliantheae tribe and *Helianthus* genus (Karis, 1993). It is noteworthy that this chloroplast-based phylogenetic tree places an annual species, *Helianthus annuus* cultivar HA89 MAX1, has been placed next to a perennial species, *H. maximiliani* (*Figure S2; Whelan, 2011)*. This placement is due to the hybridization between *H. maximiliani* and *H. annuus*, which produced *Helianthus annuus* cultivar HA89 MAX1, an *H. annuus* plant with an *H. maximiliani* chloroplast. These two species being grouped together on the phylogenetic tree of chloroplast genomes validates the results of this phylogenetic tree and the following analysis.

4.2 Origins of Helianthus schweinitzii

The fact that evolutionary origins and relationships of the rare and endangered species of *Helianthus schweinitzii* are somewhat unclear may be due to the taxonomic phenomenon of paraphyly, whereby a group consisting of the last common ancestor and most of its descendants excludes a few monophyletic subgroups (Farris, 1974). This explanation is supported by previous results (Anderson et al., 2019) as well as the present study of Heliantheae. This occurrence of paraphyly in Heliantheae is likely due to chloroplast incomplete lineage sorting that prevents descendants of a common ancestor grouped together until further speciation events occur and reticulation, origination of a new lineage due to hybridization of two ancestral lineages (Anderson et al., 2019).

Despite the difficulties in interpreting paraphyletic groupings, the present study builds on the past analysis of *H. schweinitzii* (Anderson et al., 2019) by further clarifying the relationship of the latter species to its parental and sister taxa. Prior to the availability of sequence data, *H. microcephalus* had been suggested to be a parental species based on morphological characterizations (Heiser, 1969). More recently, 38 accessions of six suspected parental species, including *H. microcephalus*, were compared to four accessions of *H. schweinitzii* using chloroplast genomic data (Anderson et al., 2019). Results from this past study had suggested *H. giganteus* as the most likely parental species for *H. schweinitzii* based on the large degree of shared cpDNA haplotypes with *H. schweinitzii* than any other sampled species (Anderson et al., 2019). However, inclusion of additional species in the present study suggests a different parentage and close relatives. When aligned with this larger sampling scheme, *H. schweinitzii* was more closely related to *Helianthus tuberosus* (MT302562.1) and *Helianthus tuberosus voucher Loeuille 793* (MT700541.1) than *H. giganteus* or *H. microcephalus*. The high bootstrap value of this relationship of 76% exceeds the BS=70 significance threshold used by many phylogeneticists (Soltis & Soltis, 2003). *H. schweinitzii* shared a 99.72% percent identity of genome relatedness with the most closely related *H. giganteus* (KF746366.1), compared to a 99.73% similarity with *H. tuberosus*. However, using an inference method that incorporates a nucleotide substitution model, i.e., the maximum-likelihood method used here, we found a greater evolutionary distance from *H. schweinitzii* to *H. giganteus* than to *H. tuberosus*. Furthermore, the closely related *H. giganteus* fell outside of the well-supported clade that *H. schweinitzii* shares with the two *H. tuberosus* accessions.

The finding that the genomes of three *H. grosseserratus* accessions (KF746350.1, KF746367.1, & KF746369.1) are sister to the clade forming *H. schweinitzii* and *H. tuberosus* may be because three species of this clade are all tuberous species, producing tuberous rhizomes or

tubers (Anderson et al., 2019; Grubbs & Wynes, 2015; *Helianthus Tuberosus - an Overview | ScienceDirect Topics*, n.d.). This result calls into question the previous finding that *H. giganteus* was the parental species

4.3 Broader Impacts

This thesis presents the largest study of Heliantheae whole chloroplast genomes, potentially even the Asteraceae family and thus contribute to the growing body of literature on ultra-barcoding resources used for endangered species and the growing field of conservation genomics. The use of ultra-barcoding of chloroplast genomes for conservation efforts allows for the investigation of inbreeding, the integration of genomic data across organismal, ecological, and landscape levels, inform alternative management, and support adaptive management approaches to species conservation (Ramakrishnan & Hohenlohe, n.d.).

For *H. schweinitzii*, this updated phylogenetic analysis can inform conservation efforts by providing a clear identification of the species and its evolutionary relationships. The results from this study have the potential to aid in the development and application of conservation genomics by providing unique insights into the dynamics of the endangered populations of *H. schweinitzii* and other threatened taxa. More specifically, this research and its contributions towards conservation genomics have potential to understand the very narrow habitat range of *H. schweinitzii* in the Piedmont Plateau and its adaptive capabilities to survive in a wider range of habitats (Primmer, 2009; Ramakrishnan & Hohenlohe, n.d.; Rossetto et al., 2021).

4.4 Limitations

Due to the highly reticulate nature of chloroplast lineages within *Helianthus,* and even Heliantheae*,* some relationships on the phylogenetic tree are not well supported or do not follow the assigned

taxonomy. This study included only one *H. schweinitzii* sample, in comparison to four in the previous study of Alexander et al. (2019). Even though the bootstrap value of its relationships is within a significant threshold, the identification of close relatives and clade groupings may change with the inclusion of additional *H. schweinitzii* samples.

Given that many of the *Helianthus* species included in this study were revealed to be paraphyletic when multiple samples are included in the phylogeny, we have reason to assume that *H. schweinitzii* may not be paraphyletic if multiple samples were included. Inclusion of more samples of *H. schweinitzii*, namely the accessions from Alexander et al. (2019), may possibly recover the same relationships as in the latter study. At this time, it is only possible to conclude that the sample incorporated in the present study is most closely related to the species of *H. grosseserratus, H. tuberosus*, and other lineages of *H. schweinitzii* of different origins.

5.0 Conclusions

The analysis comparing 154 Heliantheae chloroplast genomes provided a significant understanding of the Heliantheae chloroplast evolution and structure. Among the taxa included, the chloroplast genomes showed very similar lengths, GC content percentage, and solidified existing phylogenetic relationships. The results of this study show that the genomes of these species have remained highly conserved, indicating that regardless of their environment, their chloroplast genomes are well-suited for extreme environments and allow survival of the species.

This research applies specifically to the chloroplast genes that have remained in the chloroplast genome after many genes had been transferred to the nuclear genome over the course of plant evolution (*Gene Transfer from Organelles to the Nucleus: Frequent and in Big Chunks - PMC*, n.d.). It would appear that the gene functions in chloroplasts do not vary much across sunflower species adapted to diverse habitats. It would be of interest for future research to assess

whether photosynthetic genes in the nuclear genome, as well as nuclear genes that regulate chloroplast processes, exhibit similar trends of conservation.

By further analyzing the relationship among the rare and endangered species of *Helianthus schweinitzii, Helianthus verticillatus*, and *Helianthus paradoxus* and other species, *H. schweinitzii* has been established to be most closely related to *Helianthus tuberosus* species, *Helianthus tuberosus voucher Loeuille 793* (MT700541.1) and *Helianthus tuberosus* (MT302562.1). This clarifies and expands the findings from past research on the parental relationships and sister taxa of *Helianthus schweinitzii*. The results of the present study thus contribute to the growing body of knowledge based on ultra-barcoding of whole chloroplast genomes and its uses across scientific fields, including genetic analysis, phylogenetic relationships, and conservation genomics. The contributions of genome assembly and annotation from this study add 12 newly assembled genomes to the NCBI Short Read Archives database and allows for further analysis into the Asteraceae family, Heliantheae tribe, and *Helianthus* genus. The comparative genomics of 84 *Helianthus* species showed high conservation of the chloroplast genome, indicating that there has been little evolution of the genome across species. The results of phylogenetic analysis of these 154 species clarified the relationships within the Heliantheae tribe, specifically that of *Helianthus schweinitzii,* and identified its sister taxa in this dataset, *Helianthus tuberosus voucher Loeuille 793* and *Helianthus tuberosus*.

6.0 Acknowledgements

First and foremost, I would like to thank Dr. Nolan Kane, my Honors thesis advisor. From my initial interest in writing an Honors thesis, he encouraged and supported me throughout my entire thesis process and helped me find a love for genomics I never thought I could have. Additionally, I would like to thank other members of the Kane lab, Peter Innes and Kyle Keepers, who were fundamental in every step of this process and helped nurture this newfound love of genomics. Without Dr. Kane, Peter, and Kyle, I would not have been able to have accomplished this, so I am incredibly grateful. I would also like to thank the other members of my committee, Dr. Barbara Demmig-Adams and Dr. Gifford Miller. Finally, I would also like to acknowledge the CU Boulder Genomics students and past researchers whose genome sequences I was able to utilize in my thesis research.

References

Anderson, J., Kantar, M., Bock, D., Grubbs, K. C., Schilling, E., & Rieseberg, L. (2019). Skim-Sequencing Reveals the Likely Origin of the Enigmatic Endangered Sunflower *Helianthus schweinitzii*. *Genes*, *10*(12), 1040. https://doi.org/10.3390/genes10121040

Asteraceae | plant family | Britannica. (n.d.). Retrieved March 17, 2022, from https://www.britannica.com/plant/Asteraceae

Azarin, K., Usatov, A., Makarenko, M., Khachumov, V., & Gavrilova, V. (2021). Comparative analysis of chloroplast genomes of seven perennial *Helianthus* species. *Gene*, *774*, 145418. https://doi.org/10.1016/j.gene.2021.145418

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Bock, R. (2007). Structure, function, and inheritance of plastid genomes. In R. Bock (Ed.), *Cell and Molecular Biology of Plastids* (pp. 29–63). Springer. https://doi.org/10.1007/4735_2007_0223

Clegg, M. T., Gaut, B. S., Learn, G. H., & Morton, B. R. (1994). Rates and patterns of chloroplast DNA evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(15), 6795–6801.

CtDNAEN.pdf. (n.d.). Retrieved March 17, 2022, from https://bioscienceexplained.org/content/ctDNAEN.pdf

Daniell, H., Lin, C.-S., Yu, M., & Chang, W.-J. (2016). Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biology*, *17*, 134. https://doi.org/10.1186/s13059-016-1004-2

Du, Y.-P., Bi, Y., Yang, F.-P., Zhang, M.-F., Chen, X.-Q., Xue, J., & Zhang, X.-H. (2017). Complete chloroplast genome sequences of *Lilium*: Insights into evolutionary dynamics and phylogenetic analyses. *Scientific Reports*, *7*(1), 5751. https://doi.org/10.1038/s41598-017-06210-2

Edwards, T. P., Trigiano, R. N., Ownley, B. H., Windham, A. S., Wyman, C. R., Wadl, P. A., & Hadziabdic, D. (2020). Genetic Diversity and Conservation Status of *Helianthus verticillatus*, an Endangered Sunflower of the Southern United States. *Frontiers in Genetics*, *11*, 410. https://doi.org/10.3389/fgene.2020.00410

Farris, J. S. (1974). Formal Definitions of Paraphyly and Polyphyly. *Systematic Zoology*, *23*(4), 548–554. https://doi.org/10.2307/2412474

Frontiers | The Function of Introns | Genetics. (n.d.). Retrieved April 11, 2022, from https://www.frontiersin.org/articles/10.3389/fgene.2012.00055/full

Gene transfer from organelles to the nucleus: Frequent and in big chunks—PMC. (n.d.). Retrieved April 11, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC166356/

GeSeq—Versatile and accurate annotation of organelle genomes—PubMed. (n.d.). Retrieved March 18, 2022, from https://pubmed.ncbi.nlm.nih.gov/28486635/

Giegé, R., Jühling, F., Pütz, J., Stadler, P., Sauter, C., & Florentz, C. (2012). Structure of transfer RNAs: Similarity and variability. *WIREs RNA*, *3*(1), 37–61. https://doi.org/10.1002/wrna.103

Grubbs, K. C., & Wynes, A. (2015). Reproductive Biology of the Endangered Schweinitz's

Sunflower (*Helianthus schweinitzii*). *Castanea*, *80*(1), 20–28.

Han: PhyloXML: XML for evolutionary biology and comparati... - Google Scholar. (n.d.). Retrieved March 18, 2022, from https://scholar.google.com/scholar_lookup?title=phyloXML%3A%20XML%20for%20evolutionary%20biology%20and%20comparative%20genomics&author=MV%20Han&author=CM%20Zmasek&publication_year=2009&journal=BMC%20Bioinformatics&volume=10&pages=356

Helianthus tuberosus—An overview (pdf) | ScienceDirect Topics. (n.d.). Retrieved March 17, 2022, from https://www.sciencedirect.com/topics/agricultural-and-biologicalsciences/helianthus-tuberosus/pdf

How and why should we implement genomics into conservation? - McMahon—2014—Evolutionary Applications—Wiley Online Library. (n.d.). Retrieved April 11, 2022, from https://onlinelibrary.wiley.com/doi/full/10.1111/eva.12193

Huo, Y., Gao, L., Liu, B., Yang, Y., Kong, S., Sun, Y., Yang, Y., & Wu, X. (2019). Complete chloroplast genome sequences of four *Allium* species: Comparative and phylogenetic analyses. *Scientific Reports*, *9*, 12250. https://doi.org/10.1038/s41598-019-48708-x

Introns: Evolution and function—ScienceDirect. (n.d.). Retrieved April 11, 2022, from https://www.sciencedirect.com/science/article/pii/0959437X94900663?casa_token=BnGxwIVh1xYAAAAA:mVlxkmXS-HSzcMKCbuo7aqTIoyrNgN1u84J5TH2xjSTQwWxi5XoJI_elas5yZJ2-slrvp6rpjTg

Kane, N. C., Gill, N., King, M. G., Bowers, J. E., Berges, H., Gouzy, J., Bachlava, E., Langlade, N. B., Lai, Z., Stewart, M., Burke, J. M., Vincourt, P., Knapp, S. J., & Rieseberg, L. H.

(2011). Progress towards a reference genome for sunflower. *Botany*, *89*(7), 429–437. https://doi.org/10.1139/b11-032

Kane, N., Sveinsson, S., Dempewolf, H., Yang, J. Y., Zhang, D., Engels, J. M. M., & Cronk, Q. (2012). Ultra-barcoding in cacao (Theobroma spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *American Journal of Botany*, *99*(2), 320–329. https://doi.org/10.3732/ajb.1100570

Karis, P. O. (1993). Heliantheae sensu lato (Asteraceae), clades and classification. *Plant Systematics and Evolution*, *188*(3), 139–195. https://doi.org/10.1007/BF00937727

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kent, J. W. (2001, December 19). *BLAT—The BLAST-Like Alignment Tool*. https://genome.cshlp.org/content/12/4/656.short

Lee-Yaw, J. A., Grassa, C. J., Joly, S., Andrew, R. L., & Rieseberg, L. H. (2019). An evaluation of alternative explanations for widespread cytonuclear discordance in annual sunflowers (*Helianthus*). *New Phytologist*, *221*(1), 515–526. https://doi.org/10.1111/nph.15386

Letunic, I., & Bork, P. (2007). Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics*, *23*(1), 127–128. https://doi.org/10.1093/bioinformatics/btl529

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Loeuille, B., Thode, V., Siniscalchi, C., Andrade, S., Rossi, M., & Pirani, J. R. (2021). Extremely low nucleotide diversity among thirty-six new chloroplast genome sequences from *Aldama* (Heliantheae, Asteraceae) and comparative chloroplast genomics analyses with closely related genera. *PeerJ*, *9*, e10886. https://doi.org/10.7717/peerj.10886

Madden, T. (2003). The BLAST Sequence Analysis Tool. In *The NCBI Handbook [Internet]*. National Center for Biotechnology Information (US). https://www.ncbi.nlm.nih.gov/books/NBK21097/

Makarenko, M. S., Usatov, A. V., Tatarinova, T. V., Azarin, K. V., Logacheva, M. D., Gavrilova, V. A., & Horn, R. (2019). Characterization of the mitochondrial genome of the MAX1 type of cytoplasmic male-sterile sunflower. *BMC Plant Biology*, *19*(Suppl 1), 51. https://doi.org/10.1186/s12870-019-1637-x

McMahon, B. J., Teeling, E. C., & Höglund, J. (2014). How and why should we implement genomics into conservation? *Evolutionary Applications*, *7*(9), 999–1007. https://doi.org/10.1111/eva.12193

Ovcharenko, I., Loots, G. G., Hardison, R. C., Miller, W., & Stubbs, L. (2004). zPicture: Dynamic Alignment and Visualization Tool for Analyzing Conservation Profiles. *Genome Research*, *14*(3), 472–477. https://doi.org/10.1101/gr.2129504

Palmer, J. D. (1985). Comparative Organization of Chloroplast Genomes. *Annual Review of Genetics*, *19*(1), 325–354. https://doi.org/10.1146/annurev.ge.19.120185.001545

Phylogeny of the Coneflowers and Relatives (Heliantheae: Asteraceae) Based on Nuclear rDNA Internal Transcribed Spacer (ITS) Sequences and Chloroplast DNA Restriction Site Data on JSTOR. (n.d.). Retrieved March 17, 2022, from https://www.jstor.org/stable/2666695

Primmer, C. R. (2009). From Conservation Genetics to Conservation Genomics. *Annals of the New York Academy of Sciences*, *1162*(1), 357–368. https://doi.org/10.1111/j.1749-6632.2009.04444.x

PhyloXML: XML for evolutionary biology and comparative genomics | SpringerLink. (n.d.). Retrieved March 18, 2022, from https://link.springer.com/article/10.1186/1471-2105-10356

Possingham, J. V., & Saurer, W. (1969). Changes in chloroplast number per cell during leaf development in spinach. *Planta*, *86*(2), 186–194. https://doi.org/10.1007/BF00379826

Prehistory of the Angiosperms: Characterization of the Ancient Genomes—ScienceDirect. (n.d.). Retrieved March 18, 2022, from https://www.sciencedirect.com/science/article/pii/B9780124171633000093

Ramakrishnan, U., & Hohenlohe, P. A. (n.d.). *Conservation Genomics*. Retrieved March 18, 2022, from https://www.frontiersin.org/journal/conservationscience/section/conservation-genomics

Robinson, H. E. & Smithsonian Institution. (1981). *A revision of the tribal and subtribal limits of the Heliantheae (Asteraceae)*. Washington : Smithsonian Institution Press. http://archive.org/details/revisionoftriba511981robi

Rossetto, M., Yap, J.-Y. S., Lemmon, J., Bain, D., Bragg, J., Hogbin, P., Gallagher, R., Rutherford, S., Summerell, B., & Wilson, T. C. (2021). A conservation genomics workflow to guide practical management actions. *Global Ecology and Conservation*, *26*, e01492. https://doi.org/10.1016/j.gecco.2021.e01492

Rubinsztein, D. C., Amos, W., Leggo, J., Goodburn, S., Jain, S., Li, S.-H., Margolis, R. L., Ross, C. A., & Ferguson-Smith, M. A. (1995). Microsatellite evolution—Evidence for directionality and variation in rate between species. *Nature Genetics*, *10*(3), 337–343. https://doi.org/10.1038/ng0795-337

Serna-Sánchez, M. A., Alvarez-Yela, A. C., Arcila, J., Pérez-Escobar, O. A., Dodsworth, S., & Arias, T. (2019). *Plastid phylogenomics of the orchid family: Solving phylogenetic ambiguities within Cymbidieae and Orchidoideae* (p. 774018). bioRxiv. https://doi.org/10.1101/774018

Sharma, S. K., Dkhar, J., Kumaria, S., Tandon, P., & Rao, S. R. (2012). Assessment of phylogenetic inter-relationships in the genus *Cymbidium* (Orchidaceae) based on internal transcribed spacer region of rDNA. *Gene*, *495*(1), 10–15. https://doi.org/10.1016/j.gene.2011.12.052

Shaw, J., Lickey, E. B., Schilling, E. E., & Small, R. L. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. *American Journal of Botany*, *94*(3), 275–288. https://doi.org/10.3732/ajb.94.3.275

Singh, R., Ming, R., & Yu, Q. (2016). Comparative Analysis of GC Content Variations in Plant Genomes. *Tropical Plant Biology*, *9*(3), 136–149. https://doi.org/10.1007/s12042-0169165-4

Šmarda, P., Bureš, P., Horová, L., Leitch, I. J., Mucina, L., Pacini, E., Tichý, L., Grulich, V., & Rotreklová, O. (2014). Ecological and evolutionary significance of genomic GC content diversity in monocots. *Proceedings of the National Academy of Sciences*, *111*(39), E4096–E4102. https://doi.org/10.1073/pnas.1321152111

Sharma, S. K., Dkhar, J., Kumaria, S., Tandon, P., & Rao, S. R. (2012). Assessment of phylogenetic inter-relationships in the genus Cymbidium (Orchidaceae) based on internal transcribed spacer region of rDNA. *Gene*, *495*(1), 10–15. https://doi.org/10.1016/j.gene.2011.12.052

Soltis, P. S., & Soltis, D. E. (2003). Applying the Bootstrap in Phylogeny Reconstruction. *Statistical Science*, *18*(2), 256–267.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Structure of transfer RNAs: Similarity and variability—Giegé—2012—WIREs RNA - Wiley Online Library. (n.d.). Retrieved April 11, 2022, from https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wrna.103?casa_token=ChDi1c4j24AAAAAA%3Anqca9_8e9c5axJJE0DHc5FTUNrCGL67mYf6OcVxp4dClizmgX_R7ytZyy7kYeY3v0x-_yhVmeTKb_lup

Sunflower | Description, Uses, & Facts | Britannica. (n.d.). Retrieved March 17, 2022, from https://www.britannica.com/plant/sunflower-plant

Talat, F., & Wang, K. (2015). Comparative Bioinformatics Analysis of the Chloroplast Genomes of a Wild Diploid Gossypium and Two Cultivated Allotetraploid Species. *Iranian Journal of Biotechnology*, *13*(3), 47–56. https://doi.org/10.15171/ijb.1231

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, *38*(7), 3022–3027. https://doi.org/10.1093/molbev/msab120

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., & Greiner, S. (2017). GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Research*, *45*(W1), W6–W11. https://doi.org/10.1093/nar/gkx391

Timme, R. E., Kuehl, J. V., Boore, J. L., & Jansen, R. K. (2006). *A Comparison of the First Two Sequenced Chloroplast Genomes in Asteraceae: Lettuce and Sunflower* (LBNL-59386). Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States). https://doi.org/10.2172/960402

Walker, J. F., Zanis, M. J., & Emery, N. C. (2014). Comparative analysis of complete chloroplast genome sequence and inversion variation in *Lasthenia burkei* (Madieae, Asteraceae). *American Journal of Botany*, *101*(4), 722–729. https://doi.org/10.3732/ajb.1400049

Whelan, E. (2011). Hybridization Between Annual and Perennial Diploid Species of Helianthus. *Canadian Journal of Genetics and Cytology*, *20*, 523–530. https://doi.org/10.1139/g78-061

Xue, S., Shi, T., Luo, W., Ni, X., Iqbal, S., Ni, Z., Huang, X., Yao, D., Shen, Z., & Gao, Z. (2019). Comparative analysis of the complete chloroplast genome among *Prunus mume*, *P. armeniaca*, and *P. salicina*. *Horticulture Research*, *6*(1), 1–13. https://doi.org/10.1038/s41438-019-0171-1

Yang, J., Yue, M., Niu, C., Ma, X.-F., & Li, Z.-H. (2017). Comparative Analysis of the Complete Chloroplast Genome of Four Endangered Herbals of *Notopterygium*. *Genes*, *8*(4), 124. https://doi.org/10.3390/genes8040124

Yang, Z., & Rannala, B. (2012). Molecular phylogenetics: Principles and practice. *Nature*

Reviews Genetics, *13*(5), 303–314. https://doi.org/10.1038/nrg3186

Zhang, W., Wang, H., Dong, J., Zhang, T., & Xiao, H. (2021). Comparative chloroplast genomes and phylogenetic analysis of *Aquilegia*. *Applications in Plant Sciences*, *9*(3), e11412. https://doi.org/10.1002/aps3.11412

Zhong, Q., Yang, S., Sun, X., Wang, L., & Li, Y. (2019). The complete chloroplast genome of the Jerusalem artichoke (*Helianthus tuberosus* L.) and an adaptive evolutionary analysis of the ycf2 gene. *PeerJ*, *7*, e7596. https://doi.org/10.7717/peerj.7596

Zmasek, C. M., & Eddy, S. R. (2001). ATV: Display and manipulation of annotated phylogenetic trees. *Bioinformatics*, *17*(4), 383–384. https://doi.org/10.1093/bioinformatics/17.4.383

Appendix

Table S1: A table of the genomes analyzed in this study. This table includes the following information from left to right: Species Name, GenBank accession number, Life History (annual or perennial), Genome Length, and GC Content Percentage.

Species	GenBank (accession)	Life History	Genome Length (bp)	GC Content
Acmella paniculata		Perennial	152092	37.57%
Acmella paniculata	MZ292978.1	Perennial	152178	37.55%
Aldama anchusifolia voucher Filartiga 8	MN337902.1	Perennial	151412	37.66%
Aldama arenaria voucher Magenta 275	MN337903.1	Perennial	151423	37.70%
Aldama arenaria voucher Magenta 383	MN337904.1	Perennial	151424	37.69%
Aldama asoilioides voucher Filartiga 18	MN337905.1	Perennial	151444	37.69%
Aldama excelsa voucher Schilling H2481	MN356026.1	Perennial	151393	37.71%
Aldama bakeriana voucher Loeuille 867	MN337906.1	Perennial	151456	37.67%
Aldama vernonioides voucher Magenta 460	MN356036.1	Perennial	151303	37.65%
Aldama bracteata voucher Filartiga 15	MN337907.1	Perennial	151419	37.69%
Aldama canescens voucher Schilling 17	MN337908.1	Perennial	151466	37.70%
Aldama corumbensis voucher Loeuille 909	MN337909.1	Perennial	151409	37.70%
Aldama dentata	MN356024.1	Perennial	151392	37.79%

Aldama dentata voucher Schilling 331	MN337910.1	Perennial	151381	37.81%
Aldama discolor voucher Bombo 72	MN356025.1	Perennial	151435	37.70%
Aldama filifolia voucher Loeuille 849	MN337890.1	Perennial	151413	37.71%
Aldama fusiformis voucher Siniscalchi 398	MN337891.1	Perennial	151398	37.68%
Aldama gardneri voucher Filartiga 16	MN337892.1	Perennial	151391	37.68%
Aldama goyazii voucher Magenta 716	MN337893.1	Perennial	151356	37.69%
Aldama grandiflora voucher Loeuille 750	MN337894.1	Perennial	151435	37.72%
Aldama kunthiana voucher voucher Silva s.n. (ESA 122873)	MN337895.1	Perennial	151426	37.66%
Aldama linearis voucher Schilling 70	MN337896.1	Perennial	151393	37.73%
Aldama macrorhiza Magenta 476	MN337897.1	Perennial	151394	37.66%
Aldama megapotamica voucher Magenta 502	MN337898.1	Perennial	151445	37.64%
Aldama nudibasilaris voucher Filartiga 1	MN337899.1	Perennial	151399	37.71%
Aldama nudicaulis voucher Loeuille 734	MN337900.1	Perennial	151414	37.65%
Aldama pilosa voucher Filartiga 10	MN337901.1	Perennial	151431	37.67%
Aldama revoluta voucher Loeuille 799	MN356027.1	Perennial	151455	37.70%
Aldama robusta voucher Bringel 985	MN356028.1	Perennial	151420	37.68%
Aldama rubra voucher Magenta 388	MN356029.1	Perennial	151397	37.70%
Aldama santacatarinensis voucher Magenta 706	MN356030.1	Perennial	151449	37.66%

Aldama squalida voucher Loeuille 790	MN356031.1	Perennial	151478	37.70%
Aldama tenuifolia voucher Silva s.n. (ESA122870)	MN356032.1	Perennial	151464	37.69%
Aldama trichophylla voucher Magenta 390	MN337911.1	Perennial	151435	37.67%
Aldama trichophylla voucher Magenta 561	MN311247.1	Perennial	151381	37.64%
Aldama tuberosa voucher Loeuille 719	MN356033.1	Perennial	151402	37.65%
Aldama tukamanensis voucher Heiden 1837	MN356034.1	Perennial	151418	37.67%
Aldama veredensis voucher Loueille 921	MN356035.1	Perennial	151412	37.69%
Ambrosia artemisiifolia	MF362689.1	Annual	152281	37.59%
Ambrosia artemisiifolia	MG019037.1	Annual	152289	37.59%
Ambrosia trifida	NC_036810.2	Annual	154211	37.09%
Ambrosia trifida	MG029118.2	Annual	152099	37.61%
Dimerostemma asperatum	MT700540.1	Annual	151926	38.19%
Echinacea angustifolia	KX548221.1	Perennial	151996	37.57%
Echinacea atrorubens	KX548220.1	Perennial	151971	37.56%
Echinacea laevigata	KX548219.1	Perennial	151944	37.58%
Echinacea pallida	KX548218.1	Perennial	151939	37.57%
Echinacea paradoxa	KX548217.1	Perennial	151894	37.58%
Echinacea purpurea	KX548224.1	Perennial	151970	37.59%
Echinacea purpurea voucher IB-003	MK055334.1	Perennial	152020	37.59%

Echinacea sanguinea	KX548225.1	Perennial	151984	37.54%
Echinacea speciosa	KX548222.1	Perennial	151917	37.59%
Echinacea tennesseensis	KX548223.1	Perennial	151939	37.60%
Eclipta alba	MF993496.1	Annual	151788	37.47%
Eclipta prostrata	KU361242.1	Annual	151817	37.47%
Helianthus grosseserratus	MT302568.1	Perennial	151222	37.58%
Helianthus angustifolius	MW366797.1	Perennial	151361	37.57%
Helianthus annuus		Annual	151124	37.62%
Helianthus annuus cultivar HA89	MK341449.1	Annual	151170	37.60%
Helianthus annuus cultivar HA89(ANN2)	MK341448.1	Annual	151230	37.59%
Helianthus annuus cultivar HA89(MAX1)	MK341450.1	Annual	151335	37.57%
Helianthus annuus cultivar HA89(PET1)	MK341451.1	Annual	151197	37.59%
Helianthus annuus cultivar HA89(PET2)	MK341452.1	Annual	151180	37.59%
Helianthus annuus cultivar line HA383	DQ383815.1	Annual	151184	37.60%
Helianthus annuus cultivar XRQ/B	CM007907.1	Annual	151190	37.59%
Helianthus annuus cultivar XRQ/B HanXRQCP	MNCJ02000333.1	Annual	151203	37.59%
Helianthus annuus isolate Var 1	MN596419.1	Annual	151224	37.58%
Helianthus annuus isolate Var33	MN602834.1	Annual	151186	37.60%

Helianthus annuus strain PI468494	KU315426.1	Annual	151137	37.60%
Helianthus strumosus isolate DB31 plastid	KF745376.1	Perennial	151116	37.59%

Helianthus annuus subpecies texanus	KU304406.1	Annual	150844	37.63%
Helianthus argophyllus	KU314500.1	Annual	151134	37.61%
Helianthus atrorubens	MW375413.1	Perennial	151304	37.58%
Helianthus carnosus		Perennial	151171	37.60%
Helianthus cusickii	MW366799.1	Perennial	150015	37.71%
Helianthus debilis	KU312928.1	Annual	151178	37.60%
Helianthus decapetalus isolate DB11 plastid	KF746356.1	Perennial	151122	37.58%
Helianthus decapetalus isolate DB12 plastid	KF746357.1	Perennial	151118	37.59%
Helianthus decapetalus isolate DB13 plastid	KF746358.1	Perennial	151113	37.59%
Helianthus decapetalus isolate DB26 plastid	KF746371.1	Perennial	151112	37.59%
Helianthus decapetalus isolate DB27 plastid	KF746372.1	Perennial	151121	37.59%
Helianthus decapetalus isolate DB28 plastid	KF746373.1	Perennial	151114	37.60%
Helianthus deserticola		Annual	151190	37.60%
Helianthus divaricatus isolate DB07 plastid	KF746352.1	Perennial	151119	37.59%
Helianthus divaricatus isolate DB08 plastid	KF746353.1	Perennial	151111	37.59%
Helianthus divaricatus isolate DB09 plastid	KF746354.1	Perennial	151123	37.59%
Helianthus divaricatus isolate DB10 plastid	KF746355.1	Perennial	151097	37.60%
Helianthus divaricatus isolate DB19 plastid	KF746364.1	Perennial	151121	37.59%
Helianthus floridanus	MW351833.1	Perennial	150819	37.62%

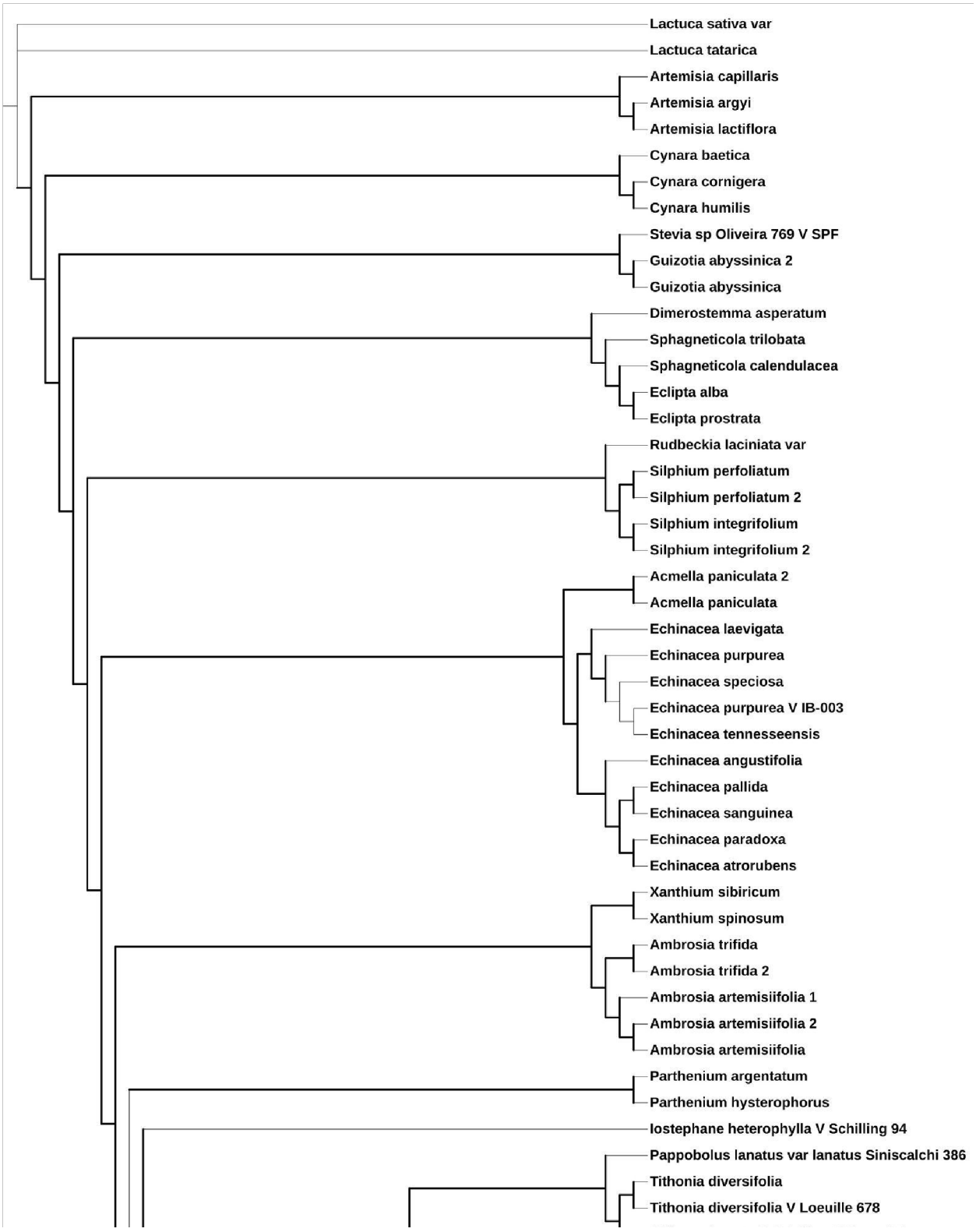
Helianthus giganteus		Perennial	151332	37.57%
Helianthus giganteus isolate DB01 plastid	KF746346.1	Perennial	151138	37.58%
Helianthus giganteus isolate DB02 plastid	KF746347.1	Perennial	151133	37.59%
Helianthus giganteus isolate DB03 plastid	KF746348.1	Perennial	151119	37.59%
Helianthus giganteus isolate DB04 plastid	KF746349.1	Perennial	151099	37.60%
Helianthus giganteus isolate DB20 plastid	KF746365.1	Perennial	151129	37.59%
Helianthus giganteus isolate DB21 plastid	KF746366.1	Perennial	151125	37.58%
Helianthus grosseserratus isolate DB05 plastid	KF746350.1	Perennial	151094	37.59%
Helianthus grosseserratus isolate DB06 plastid	KF746351.1	Perennial	151129	37.59%
Helianthus grosseserratus isolate DB22 plastid	KF746367.1	Perennial	151119	37.59%
Helianthus grosseserratus isolate DB23 plastid	KF746368.1	Perennial	151067	37.60%
Helianthus grosseserratus isolate DB24 plastid	KF746369.1	Perennial	151106	37.59%
Helianthus grosseserratus isolate DB25 plastid	KF746370.1	Perennial	151122	37.59%
Helianthus hirsutus	MT302566.1	Perennial	151334	37.57%
Helianthus hirsutus isolate DB14 plastid	KF746359.1	Perennial	151116	37.59%
Helianthus hirsutus isolate DB15 plastid	KF746360.1	Perennial	151117	37.59%
Helianthus hirsutus isolate DB29 plastid	KF746374.1	Perennial	151110	37.59%
Helianthus hirsutus isolate DB30 plastid	KF746375.1	Perennial	151120	37.59%
Helianthus laciniatus	MW353165.1	Perennial	151177	37.60%

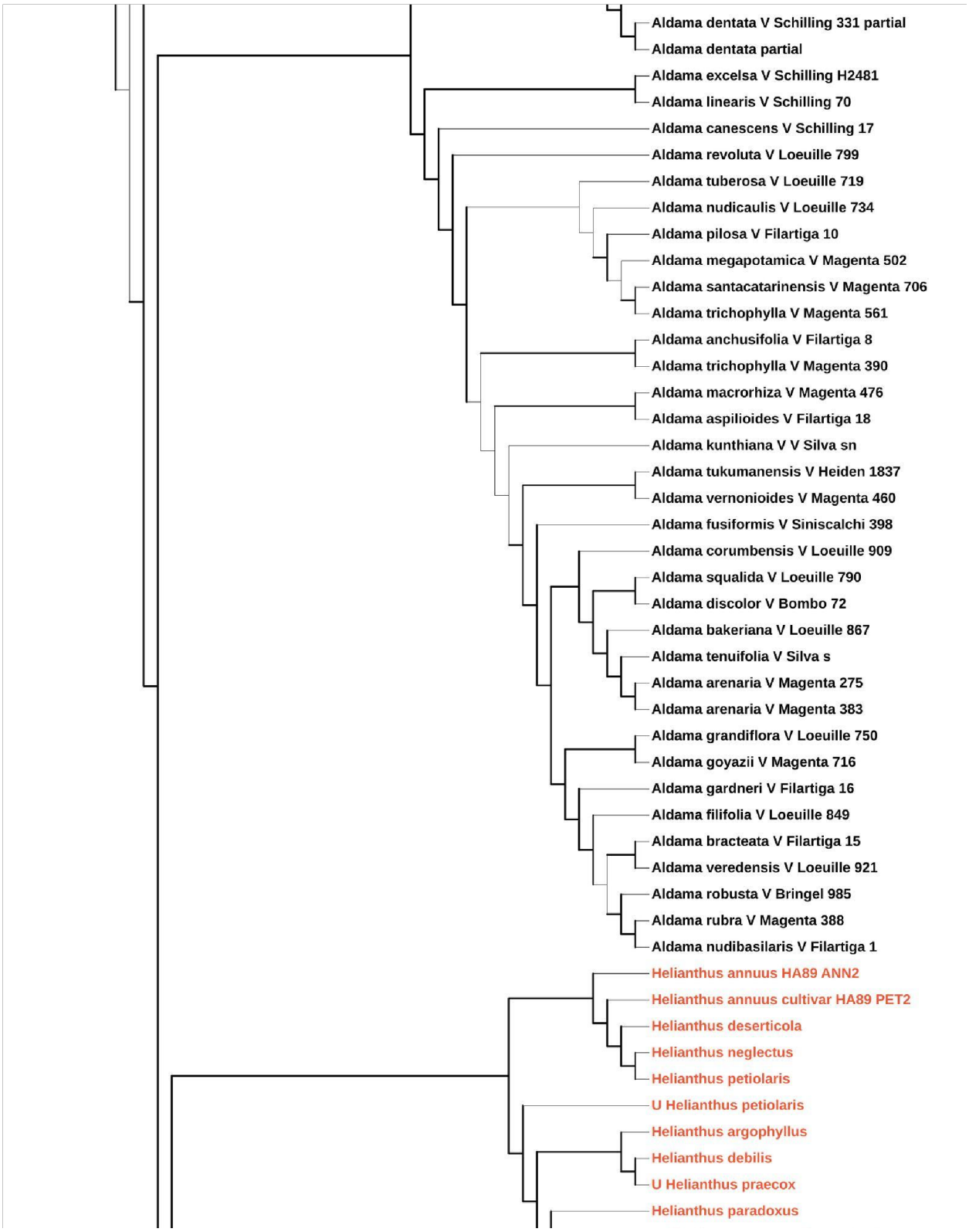
Helianthus maximiliani isolate MAX01 plastid	KF746380.1	Perennial	151082	37.59%
Helianthus maximiliani isolate MAX15 plastid	KF746381.1	Perennial	151122	37.59%
Helianthus maximiliani isolate MAX16 plastid	KF746382.1	Perennial	151124	37.59%
Helianthus maximiliani isolate MAX17 plastid	KF746383.1	Perennial	151151	37.59%
Helianthus microcephalus	MT302565.1	Perennial	151356	37.56%
Helianthus microcephalus	MW351834.1	Perennial	150959	37.57%
Helianthus neglectus		Annual	151203	37.60%
Helianthus nuttallii	MW366798.1	Perennial	151176	37.61%
Helianthus nuttallii subspecies nuttallii		Perennial	150646	37.62%
Helianthus paradoxus		Annual	151057	37.62%
Helianthus pauciflorus subspecies pauciflorus	MT302564.1	Perennial	151294	37.58%
Helianthus praecox	KU308401.1	Annual	151143	37.60%
Helianthus petiolaris	KU310904.1	Annual	150814	37.68%
Helianthus petiolaris subspecies fallax	KU295560.1	Annual	150984	37.57%
Helianthus porteri		Perennial	150999	37.64%
Helianthus salicifolius	MT302563.1	Perennial	151295	37.58%
Helianthus schweinitzii	MW381301.1	Perennial	151274	37.60%
Helianthus silphiodes	MW375415.1	Perennial	151199	37.59%
Helianthus simulans	MW366800.1	Perennial	151227	37.60%
Helianthus strumosus	MT302567.1	Perennial	151275	37.57%

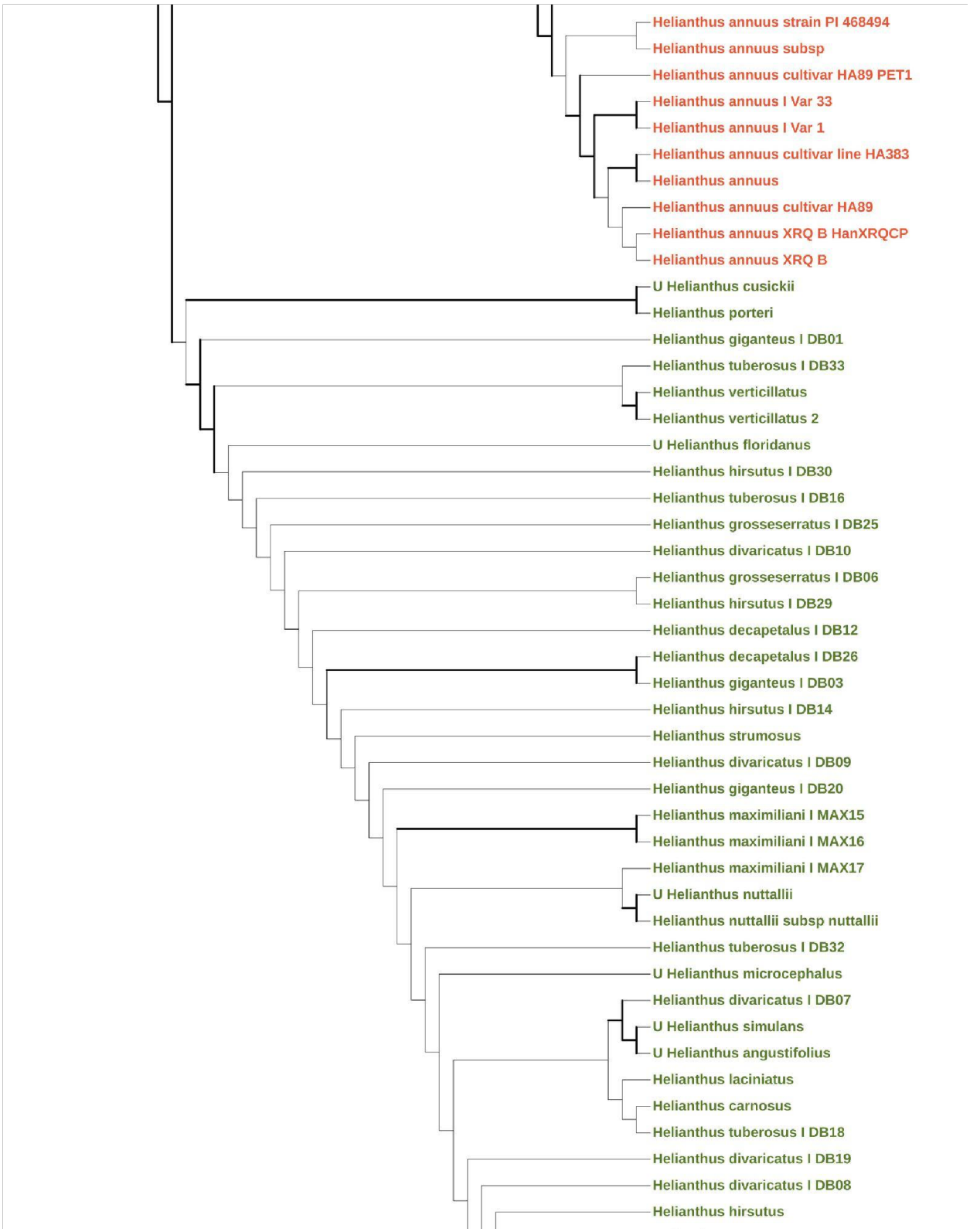
Helianthus tuberosus	MG696658.1	Perennial	151494	37.55%
Helianthus tuberosus	MT302562.1	Perennial	151324	37.57%
Helianthus tuberosus isolate DB16 plastid	KF746361.1	Perennial	151119	37.59%
Helianthus tuberosus isolate DB17 plastid	KF746362.1	Perennial	151118	37.59%
Helianthus tuberosus isolate DB18 plastid	KF746363.1	Perennial	151119	37.59%
Helianthus tuberosus isolate DB32 plastid	KF746377.1	Perennial	151137	37.59%
Helianthus tuberosus isolate DB33 plastid	KF746378.1	Perennial	151116	37.59%
Helianthus tuberosus isolate DB34 plastid	KF746379.1	Perennial	151117	37.59%
Helianthus tuberosus voucher Loueille 793	MT700541.1	Perennial	151326	37.62%
Helianthus verticillatus		Perennial	151300	37.58%
Helianthus verticillatus	MW375414.1	Perennial	151257	37.58%
Iostephane heterophylla voucher Schilling 94	MT700542.1	Perennial	151582	37.67%
Pappobolus lanatus variation lanatus voucher Siniscalchi 386	MT700543.1	Perennial	151456	37.70%
Parthenium argentatum	GU120098.1	Perennial	152859	37.59%
Parthenium hysterophorus	MT576959.1	Annual	151979	37.59%
Rudbeckia laciniata var. laciniata	MN518844.1	Perennial	151917	37.62%
Silphium integrifolium 1		Perennial	152242	37.47%
Silphium integrifolium 2		Perennial	152085	37.52%
Silphium perfoliatum		Perennial	151992	37.52%
Silphium perfoliatum	MN445187.1	Perennial	151987	37.51%
Sphagneticola calendulacea	KY828438.1	Perennial	151817	37.47%

Sphagneticola trilobata	KY940274.1	Perennial	151993	37.45%
Tithonia diversifolia	MT576958.1	Perennial	151225	37.64%
Tithonia diversifolia voucher Loeuille 678	MT700544.1	Perennial	151441	37.66%
Xanthium spinosum	MT668935.1	Annual	152482	37.43%
Xanthium sibiricum	MH473582.1	Annual	151958	37.50%
Outgroups				
Artemisia argyi	OK647842.1	Perennial	153342	36.94%
Artemisia capillaris	MK307819.1	Perennial	153232	36.93%
Artemisia lactiflora	MZ151340.1	Perennial	153318	36.94%
Cynara baetica	NC_028005.1	Perennial	154727	37.18%
Cynara cornigera	NC_028006.1	Perennial	154729	37.18%
Cynara humilis	NC_028006.1	Perennial	154764	37.17%
Guizotia abyssinica	NC_010601.1	Annual	153930	37.09%
Guizotia abyssinica	EU549769.1	Annual	153930	37.09%
Lactuca sativa var. angustana	MT215016.1	Annual	154927	37.02%
Lactuca tatarica	NC_058613.1	Perennial	154578	37.06%
Stevia sp. Oliveira 769 voucher SPF	MT793846.1	Perennial	153408	37.02%

Figure S1: Phylogenetic Tree of 154 *Heliantheae* species with the following outgroups to root the tree: *Artemisia argyi*, *Artemisia capillaris*, *Artemisia lactiflora*, *Cynara baetica*, *Cynara cornigera*, *Cynara humilis*, *Guizotia abyssinica*, *Lactuca sativa*, *Lactuca tatarica*, and *Stevia*. The tree is annotated with the following color labels: green = perennial, red = annual, purple = *Helianthus schweinitzii*. Bootstrap values over 70 are indicated by bolded tree branches.


Lactuca sativa var
Lactuca tatarica
Artemisia capillaris
Artemisia argyi
Artemisia lactiflora
Cynara baetica
Cynara cornigera
Cynara humilis
Stevia sp Oliveira 769 V SPF
Guizotia abyssinica 2
Guizotia abyssinica
Dimerostemma asperatum
Sphagneticola trilobata
Sphagneticola calendulacea
Eclipta alba
Eclipta prostrata
Rudbeckia laciniata var
Silphium perfoliatum
Silphium perfoliatum 2
Silphium integrifolium
Silphium integrifolium 2
Acmella paniculata 2
Acmella paniculata
Echinacea laevigata
Echinacea purpurea
Echinacea speciosa
Echinacea purpurea V IB-003
Echinacea tennesseensis
Echinacea angustifolia
Echinacea pallida
Echinacea sanguinea
Echinacea paradoxa
Echinacea atrorubens
Xanthium sibiricum
Xanthium spinosum
Ambrosia trifida
Ambrosia trifida 2
Ambrosia artemisiifolia 1
Ambrosia artemisiifolia 2
Ambrosia artemisiifolia
Parthenium argentatum
Parthenium hysterophorus
Iostephane heterophylla V Schilling 94
Pappobolus lanatus var lanatus Siniscalchi 386
Tithonia diversifolia
Tithonia diversifolia V Loeuille 678


Aldama dentata V Schilling 331 partial
Aldama dentata partial
Aldama excelsa V Schilling H2481
Aldama linearis V Schilling 70
Aldama canescens V Schilling 17
Aldama revoluta V Loeuille 799
Aldama tuberosa V Loeuille 719
Aldama nudicaulis V Loeuille 734
Aldama pilosa V Filartiga 10
Aldama megapotamica V Magenta 502
Aldama santacatarinensis V Magenta 706
Aldama trichophylla V Magenta 561
Aldama anchusifolia V Filartiga 8
Aldama trichophylla V Magenta 390
Aldama macrorhiza V Magenta 476
Aldama aspilioides V Filartiga 18
Aldama kunthiana V V Silva sn
Aldama tukumanensis V Heiden 1837
Aldama vernonioides V Magenta 460
Aldama fusiformis V Siniscalchi 398
Aldama corumbensis V Loeuille 909
Aldama squalida V Loeuille 790
Aldama discolor V Bombo 72
Aldama bakeriana V Loeuille 867
Aldama tenuifolia V Silva s
Aldama arenaria V Magenta 275
Aldama arenaria V Magenta 383
Aldama grandiflora V Loeuille 750
Aldama goyazii V Magenta 716
Aldama gardneri V Filartiga 16
Aldama filifolia V Loeuille 849
Aldama bracteata V Filartiga 15
Aldama veredensis V Loeuille 921
Aldama robusta V Bringel 985
Aldama rubra V Magenta 388
Aldama nudibasilaris V Filartiga 1
Helianthus annuus HA89 ANN2
Helianthus annuus cultivar HA89 PET2
Helianthus deserticola
Helianthus neglectus
Helianthus petiolaris
U Helianthus petiolaris
Helianthus argophyllus
Helianthus debilis
U Helianthus praecox
Helianthus paradoxus


Helianthus annuus strain PI 468494
Helianthus annuus subsp
Helianthus annuus cultivar HA89 PET1
Helianthus annuus I Var 33
Helianthus annuus I Var 1
Helianthus annuus cultivar line HA383
Helianthus annuus
Helianthus annuus cultivar HA89
Helianthus annuus XRQ B HanXRQCP
Helianthus annuus XRQ B
U Helianthus cusickii
Helianthus porteri
Helianthus giganteus I DB01
Helianthus tuberosus I DB33
Helianthus verticillatus
Helianthus verticillatus 2
U Helianthus floridanus
Helianthus hirsutus I DB30
Helianthus tuberosus I DB16
Helianthus grosseserratus I DB25
Helianthus divaricatus I DB10
Helianthus grosseserratus I DB06
Helianthus hirsutus I DB29
Helianthus decapetalus I DB12
Helianthus decapetalus I DB26
Helianthus giganteus I DB03
Helianthus hirsutus I DB14
Helianthus strumosus
Helianthus divaricatus I DB09
Helianthus giganteus I DB20
Helianthus maximiliani I MAX15
Helianthus maximiliani I MAX16
Helianthus maximiliani I MAX17
U Helianthus nuttallii
Helianthus nuttallii subsp nuttallii
Helianthus tuberosus I DB32
U Helianthus microcephalus
Helianthus divaricatus I DB07
U Helianthus simulans
U Helianthus angustifolius
Helianthus laciniatus
Helianthus carnosus
Helianthus tuberosus I DB18
Helianthus divaricatus I DB19
Helianthus divaricatus I DB08
Helianthus hirsutus

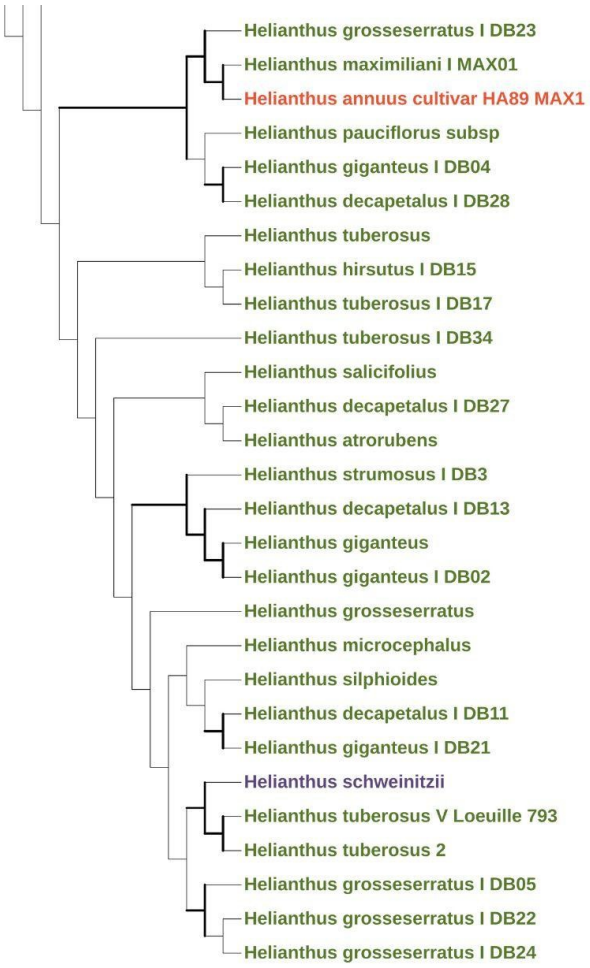

Helianthus grosseserratus I DB23
Helianthus maximiliani I MAX01
Helianthus annuus cultivar HA89 MAX1
Helianthus pauciflorus subsp
Helianthus giganteus I DB04
Helianthus decapetalus I DB28
Helianthus tuberosus
Helianthus hirsutus I DB15
Helianthus tuberosus I DB17
Helianthus tuberosus I DB34
Helianthus salicifolius
Helianthus decapetalus I DB27
Helianthus atrorubens
Helianthus strumosus I DB3
Helianthus decapetalus I DB13
Helianthus giganteus
Helianthus giganteus I DB02
Helianthus grosseserratus
Helianthus microcephalus
Helianthus silphioides
Helianthus decapetalus I DB11
Helianthus giganteus I DB21
Helianthus schweinitzii
Helianthus tuberosus V Loeuille 793
Helianthus tuberosus 2
Helianthus grosseserratus I DB05
Helianthus grosseserratus I DB22
Helianthus grosseserratus I DB24

Figure S2: Portion of the phylogenetic tree (*Figure S1*) showing the relationship between *Helianthus maximiliani* I MAX 01 and *Helianthus annuus* cultivar HA89 MAX1. The bootstrap value of this relationship is 81.