

Genome skimming reveals the origin of the Jerusalem Artichoke tuber crop species: neither from Jerusalem nor an artichoke

Dan G. Bock¹, Nolan C. Kane², Daniel P. Ebert¹ and Loren H. Rieseberg^{1,3}

¹Biodiversity Research Centre and Department of Botany, University of British Columbia, Vancouver BC V6T 1Z4, Canada; ²Department of Ecology and Evolutionary Biology, University of Colorado at Boulder, Boulder, CO 80309, USA; ³Biology Department, Indiana University, Bloomington, IN 47405, USA

Summary

Author for correspondence:

Dan G. Bock

Tel: +1 604 827 3535

Email: dan.g.bock@gmail.com

Received: 12 July 2013

Accepted: 14 September 2013

New Phytologist (2014) 201: 1021–1030

doi: 10.1111/nph.12560

Key words: allopolyploid speciation, phylogenomics, polyploid parentage, polyploidy, ultra-barcoding.

- The perennial sunflower *Helianthus tuberosus*, known as Jerusalem Artichoke or Sunchoke, was cultivated in eastern North America before European contact. As such, it represents one of the few taxa that can support an independent origin of domestication in this region. Its tubers were adopted as a source of food and forage when the species was transferred to the Old World in the early 1600s, and are still used today.
- Despite the cultural and economic importance of this tuber crop species, its origin is debated. Competing hypotheses implicate the occurrence of polyploidization with or without hybridization, and list the annual sunflower *H. annuus* and five distantly related perennial sunflower species as potential parents.
- Here, we test these scenarios by skimming the genomes of diverse populations of Jerusalem Artichoke and its putative progenitors. We identify relationships among *Helianthus* taxa using complete plastomes (151 551 bp), partial mitochondrial genomes (196 853 bp) and 35S (8196 bp) and 5S (514 bp) ribosomal DNA.
- Our results refute the possibility that Jerusalem Artichoke is of *H. annuus* ancestry. We provide the first genetic evidence that this species originated recursively from perennial sunflowers of central-eastern North America via hybridization between tetraploid Hairy Sunflower and diploid Sawtooth Sunflower.

Introduction

The perennial sunflower *Helianthus tuberosus* is a taxon with a rich human-connected history. The Cree and Huron Indians of eastern North America, who referred to this plant as ‘askipaw’ and ‘skibwan’ (‘raw thing’), respectively, grew it for its large tubers before the first European contact (Heiser, 1976; Kosaric *et al.*, 1984; Kays & Nottingham, 2008). As such, although tuber archeological remains are yet to be recovered for this species, *H. tuberosus* represents one of the few domesticates that can support eastern North America as one of the world’s cradles of domestication. After being transferred to the Old World in the early 1600s, it was readily adopted as a food plant (Heiser, 1976; Kosaric *et al.*, 1984; Kays & Nottingham, 2008). In the process, it acquired an impressive assortment of common names that vary in botanical accuracy (Heiser, 1976; Kosaric *et al.*, 1984; Kays & Nottingham, 2008), such as ‘Jerusalem Artichoke’ or ‘Sunchoke’. Among these, ‘Jerusalem Artichoke’, thought to be a corruption of the Italian ‘*girasole articiocco*’ (‘sunflower artichoke’; Smith, 1807), is its most widely used appellation. By the mid-18th century, as farming of potato became widespread, the relative importance of Jerusalem Artichoke as a food plant decreased (Kays & Nottingham, 2008). Even so, it remains a globally cultivated multifunctional crop, well adapted to diverse geoclimatic regions

(Kosaric *et al.*, 1984), including dry climates with nutrient-poor soils (Kays & Nottingham, 2008). Recent surges in its production have been prompted by the health benefits associated with the consumption of inulin (Kleessen *et al.*, 2007; Roberfroid, 2007), the reserve carbohydrate stored in Jerusalem Artichoke tubers, and the utility of its below-ground and above-ground parts for biofuel production and livestock feed (Bajpai & Bajpai, 1991; Cheng *et al.*, 2009).

Despite its cultural and economic significance, important aspects of the origin of the Jerusalem Artichoke, with implications for germplasm preservation and cultivar improvement, remain unanswered. Specifically, although it is currently agreed that the Jerusalem Artichoke species originated in central-eastern North America, where its wild populations abound (Rogers *et al.*, 1982; Kays & Nottingham, 2008), other details of its evolution remain a mystery. For instance, it is uncertain whether this hexaploid species ($2n = 6x = 102$) is monophyletic (i.e. autopolyploid) or polyphyletic (i.e. allopolyploid or auto-allopolyploid; Kostoff, 1934, 1939; Darlington, 1956; Heiser & Smith, 1964). Among these, the polyphyletic auto-allopolyploid scenario appears to be the most likely, as it is supported by the cytogenetic observation that two of the three chromosome sets of Jerusalem Artichoke are homologous (Kostoff, 1939). Aside from the mechanism of formation, also unknown is the identity of the progenitor species.

Two competing hypotheses have been proposed, each based on different lines of evidence. The first hypothesis, drawing on the fact that the Jerusalem Artichoke can be crossed readily with the annual sunflower *H. annuus* (Kostoff, 1939), and shows similarity to this species based on immunochemistry data (Anisimova, 1982), posits that one parent of Jerusalem Artichoke is the annual Common Sunflower *H. annuus*. The alternative hypothesis is that the Jerusalem Artichoke originated strictly from perennial sunflowers, most likely via hybridization between tetraploid ($2n=4x=68$) and diploid ($2n=2x=34$; Heiser & Smith, 1964; Heiser, 1976) species. This hypothesis implicates as potential progenitors a group of five perennial sunflower taxa whose morphology and North American ranges overlap with that of Jerusalem Artichoke (Heiser *et al.*, 1969; Heiser, 1976; Kays & Nottingham, 2008). Of these, the Hairy Sunflower (*H. hirsutus*), a species whose rhizomes are often thickened terminally (Heiser *et al.*, 1969), which has been proposed as an autopolyploid of *H. divaricatus* (Heiser *et al.*, 1969), is seen as the most likely tetraploid progenitor (Heiser, 1976). The Sawtooth Sunflower (*H. grosseserratus*) and the Giant Sunflower (*H. giganteus*) are similarly considered the most probable diploid progenitors (Heiser, 1976).

Molecular phylogenetics has so far remained inconclusive in establishing the origin of the Jerusalem Artichoke (Gentzbittel *et al.*, 1992; Schilling, 1997; Schilling *et al.*, 1998; Timme *et al.*, 2007). This is because the diversification of perennial *Helianthus* species is characterized by several processes known to confound phylogenetic inference. These include their recent, rapid radiation (Schilling, 1997; Timme *et al.*, 2007), the formation of diploid hybrids (Long, 1955; Timme *et al.*, 2007) and of polyploids via whole-genome duplication with or without hybridization (Timme *et al.*, 2007), and the prevalence of post-speciation gene flow facilitated by high levels of interspecies fertility (Heiser & Smith, 1964). In addition, taxonomic ambiguity is common among perennial sunflowers, given their frequently overlapping morphologies (Heiser *et al.*, 1969).

Phylogenomics is an effective means of addressing complex phylogenetic questions. Although traditionally used to resolve deep splits in the tree of life, this approach is now being applied to shallow phylogenetic divisions (Emerson *et al.*, 2010; Wagner *et al.*, 2013). For recently diverged plant species in particular, and for those with large genomes, a genome skimming approach has been advocated (Straub *et al.*, 2012). Also known as ultra-barcoding, or UBC (Kane & Cronk, 2008; Kane *et al.*, 2012), genome skimming consists of the assembly and analysis of the high-copy genomic fraction, consisting of plastid and mitochondrial genomes as well as nuclear ribosomal DNA (rDNA). Aside from the large amount of data generated, the value of this approach stems from the complementary utility of the two marker categories. The non-recombining and uniparentally inherited organellar genomes allow the matrilineal genealogy to be recovered. In cases of reticulate speciation, organellar DNA can be used to discern between single vs multiple origin scenarios (Soltis & Soltis, 1989; Schwarzbach & Rieseberg, 2002; Guggisberg *et al.*, 2006; Slotte *et al.*, 2006), and to clarify whether maternal parentage was reciprocal or unidirectional (Soltis & Soltis, 1989). The

biparentally inherited rDNA is ideally suited for inferring species-level phylogenies. Provided that concerted evolution has not homogenized divergent parental genotypes, rDNA can readily reveal evidence of hybridization (Malinska *et al.*, 2010, 2011). In perennial sunflowers, in particular, rDNA has proven to be the most phylogenetically informative region studied so far (Timme *et al.*, 2007).

Here, we used a genome skimming approach to investigate the origin of the Jerusalem Artichoke. We collected the largest dataset used to date in *Helianthus* phylogeny, consisting of complete plastid genomes as well as partial sequences for the mitochondrial genome and nuclear-encoded 35S and 5S rDNA. We screened 38 accessions, representing geographically diverse populations of eight species (Fig. 1; Supporting Information Table S1), including the Jerusalem Artichoke and all diploid and tetraploid perennial sunflowers which have been proposed as its progenitors. We supplemented these data with corresponding sequences from *H. annuus*, such that all proposed parents of the Jerusalem Artichoke are represented in our dataset.

Materials and Methods

Molecular techniques

The accessions used in this study were obtained from the US Department of Agriculture (USDA) collections held at Ames, IA, and were chosen to maximize geographical representation for each species within central-eastern North America (Fig. 1; Table S1). The ploidy of each accession (Fig. 1; Table S1) was determined using flow cytometry, with the internal standards *Zea mays* ($2C=5.43$ pg), *Secale cereale* ($2C=16.19$ pg) and *Vicia faba* ($2C=26.90$ pg; Doležel *et al.*, 2007). DNA was extracted from leaf tissue of single individuals using established procedures (Doyle & Doyle, 1987). Illumina paired-end libraries (100 bp read length) were prepared from fragmented genomic DNA (fragment size *c.* 400 bp) following standard protocols. With the exception of the four *H. maximiliani* accessions, which were sequenced with samples from a related project, all libraries were run on one lane on an Illumina HiSeq 2000 machine, with pooling designed to achieve comparable total coverage for each species and ploidy level (Table S2).

Assembly of plastid and mitochondrial genomes

Before *de novo* assembly, we reduced the complexity of each library by aligning quality-filtered reads to the *H. annuus* plastid (GenBank accession NC007977) and mitochondrial (GenBank accession KF815390) genomes using Bowtie2 (Langmead & Salzberg, 2012). Apart from simplifying the assembly task, this step was used to gauge the average coverage across each genome (Table S3), and to calibrate the fragment length of each library for *de novo* assembly. Reads corresponding to organellar genomes were assembled using the *de novo* de Bruijn graph-based tool VELVET (version 1.2.06; Zerbino & Birney, 2008). We used a hash length of 21, and a minimum contig length of 100 bp. For the plastid assembly, for which the average coverage depth was

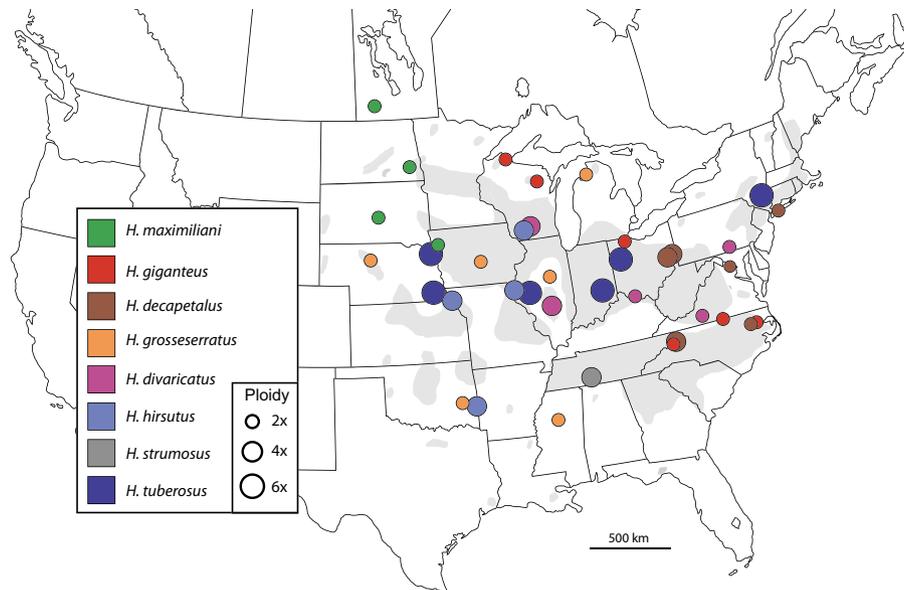


Fig. 1 Geographical distribution of the perennial *Helianthus* accessions sequenced. Gray shading is used to illustrate the range of Jerusalem Artichoke in the USA (as per Rogers *et al.*, 1982).

95× (Table S3), we set the coverage cut-off to 15. For the mitochondrial genome, for which the average coverage depth was 9× (Table S3), we allowed VELVET to automate the coverage cut-off. The resulting contigs were aligned to the corresponding organellar genome of *H. annuus*, ordered and merged (when overlapped) using CodonCode Aligner (version 2.0.4; CodonCode Corporation, Dedham, MA, USA). For the plastid genome, small gaps were filled using trimmed Illumina reads. Mononucleotide repeats that could not be bridged in all samples were collapsed, for all samples, to the smallest repeat size present in the dataset. For the mitochondrial genome, regions not covered by Illumina reads were coded as gaps. Draft assemblies for the plastid and mitochondrial genomes of each accession were validated by mapping quality-filtered Illumina reads and visually inspecting the coverage distribution using Tablet (version 1.12.03.26; Milne *et al.*, 2010). The full-length plastid genome of each accession was annotated using DOGMA (Wyman *et al.*, 2004).

Assembly of 35S and 5S rDNA regions

Quality-filtered reads for each accession were assembled using the *de novo* de Bruijn graph-based tool Trinity (version R2012-06-08; Grabherr *et al.*, 2011) at default parameters. Contigs for 35S and 5S rDNA were identified based on alignments to the corresponding *H. annuus* references for 35S (GenBank accession KF767534) and 5S (GenBank accession HM638217). Preliminary inspection of rDNA contigs revealed three regions that could be aligned unambiguously across all samples: a 7457-bp stretch of 35S rDNA (consisting of partial ETS, 18S, ITS1, 5.8S, ITS2, 26S and partial NTS), an additional 739-bp stretch of the NTS associated with 35S, and a 514-bp stretch of 5S rDNA (consisting of 5S and its corresponding NTS region). To incorporate intra-individual polymorphism between rDNA

repeats, we aligned quality-filtered reads to each of the three regions using Bowtie2, and called single nucleotide polymorphisms (SNPs) using Unified Genotyper from the Genome Analysis Toolkit (GATK; version 2.1–13; DePristo *et al.*, 2011). Because of the repetitive nature of rDNA, we treated all samples as polyploids at the SNP-calling step. To determine the ploidy setting, we surveyed 23 distinct values in Unified Genotyper (range 2×–200×; Figs S1, S2) for each accession and rDNA region, and recorded the number of SNPs called given the filtering criteria (i.e. GATK confidence score > 10; mapping quality > 15). For the final analysis, we used 100×, the ploidy setting for which the maximum number of SNPs was called, and beyond which the number of SNPs remained relatively constant (Figs S1, S2). Polymorphisms scored under these conditions and filtering criteria were incorporated in the *de novo* assemblies using IUPAC ambiguity codes.

Alignment and phylogenetic analyses

For each region, we retained full sequence data, consisting of both variable and invariable sites. Alignments performed in MAFFT (version 6.814b; Katoh & Toh, 2008) with default settings were inspected and edited in CodonCode Aligner. For the alignment of draft mitochondrial genomes, we removed sites with missing data in more than five samples, and excluded singleton SNPs, because of the low coverage obtained for this region (Table S3). We also excluded 15 segments that were classified, according to BLAST searches against the *H. annuus* plastid genome, as likely integrants of plastid DNA in the mitochondrial genome. For the 35S and 5S rDNA alignments, we excluded singleton SNPs identified (Fig. S3), to address the possibility that false positive calls may have been incorporated in the assemblies at the SNP-calling step. Maximum likelihood (ML) phylogenies were inferred using PhyML Best AIC Tree (version 1.02b),

implemented in Phylemon (version 2.0; Sánchez *et al.*, 2011). PhyML Best AIC Tree uses PhyML (version 3.0; Guindon & Gascuel, 2003) to select the best model of sequence evolution under the Akaike information criterion (AIC) and to build ML phylogenies. ML branch support was estimated using the Shimodaira–Hasegawa-like (SH-like) procedure implemented in PhyML (Guindon & Gascuel, 2003). The SH-like procedure assesses whether the branch being studied provides a significant likelihood gain compared with the null hypothesis that involves collapsing that branch (Guindon & Gascuel, 2003). It is a fast method for branch support estimation suitable for large datasets, which provides similar results to bootstrap (Anisimova & Gascuel, 2006). Bayesian inference analyses were conducted with MrBayes (version 3.2.1; Ronquist & Huelsenbeck, 2003), with parameters of sequence substitution set to follow as closely as possible the model inferred by PhyML. We used four runs, each with four Markov chains initiated from a random tree and run until the average standard deviation of split frequencies remained below 0.01 (range 1000 000–4000 000 generations). Trees were sampled every 500 generations. The first 25% of all trees sampled before convergence were discarded as burn-in. The mean level of sequence divergence between organellar haplotypes within each species was calculated in MEGA 5 (Tamura *et al.*, 2011), using the Tamura–Nei (TrN) model (Tamura & Nei, 1993).

Survey of diagnostic polymorphism in rDNA data

Diagnostic sites, defined here as sites that are fixed in a given species at its lowest ploidy level, were identified by scanning the rDNA alignments in CodonCode Aligner. When such diagnostic sites showed intra-individual polymorphism (i.e. were coded in IUPAC degenerate bases), we obtained the frequency of each underlying allele from individual quality-filtered VCF files.

Results and Discussion

Unrooted phylogenies of organellar genomes and rDNA revealed extensive sequence divergence between *H. annuus* and perennial sunflowers, including the Jerusalem Artichoke (Figs S4–S6). Two scenarios are compatible with this observation. The first is that *H. annuus* was involved in the parentage of Jerusalem Artichoke as the pollen donor, but concerted evolutionary forces acting since the polyploidization event have overwritten the *H. annuus*-derived rDNA to the maternal type. Homogenization of rDNA arrays has been documented in other polyploids (Wendel *et al.*, 1995), and can occur over the course of only a few generations (Malinska *et al.*, 2010, 2011). Nevertheless, in the case of Jerusalem Artichoke, frequent vegetative reproduction should have resulted in the retention of both parental sequences for prolonged

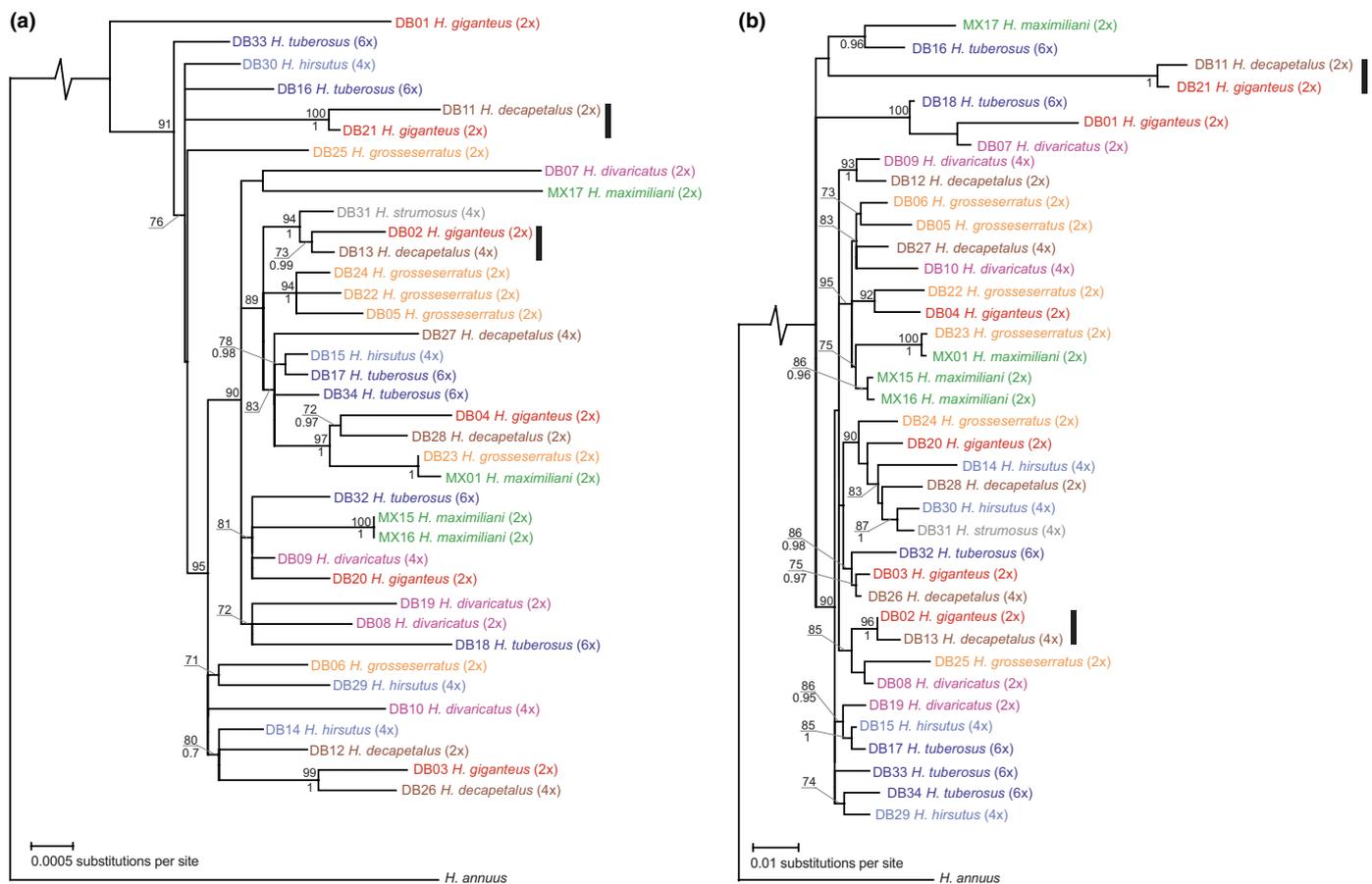


Fig. 2 Maximum likelihood phylogenetic reconstruction of (a) complete plastid genomes (151 551 bp) and (b) partial mitochondrial genomes (196 853 bp) for the perennial *Helianthus* accessions sequenced. Support is shown for nodes with Shimodaira–Hasegawa-like (SH-like) values > 70% (above) and Bayesian posterior probabilities > 0.7 (below). Groupings supported by geography are indicated by black vertical bars.

periods of time. The alternative scenario is that *H. annuus* did not contribute any of the three genomes in Jerusalem Artichoke. The two species are cross-fertile, and this possibility has been exploited in the past to transfer resistance to pathogens from Jerusalem Artichoke into cultivated sunflower (Atlagić *et al.*, 1993; Atlagić & Škorić, 2006). However, the resulting hybrids often show greatly reduced fertility (Heiser & Smith, 1964; Atlagić *et al.*, 1993; Atlagić & Škorić, 2006). Cytogenetic observations of Jerusalem Artichoke \times *H. annuus* progeny have also documented a high frequency of meiotic abnormalities linked to faulty homologue recognition, including univalent and multivalent formation (Atlagić *et al.*, 1993; Atlagić & Škorić, 2006). By contrast, hybrids between Jerusalem Artichoke and diploid perennial sunflowers, such as *H. divaricatus*, show more regular meiosis, with reduced univalent formation (Chandler, 1991). These studies suggest that differences in chromosomal structure between Jerusalem Artichoke and *H. annuus* are more pronounced than those between Jerusalem Artichoke and perennial sunflowers, and, as such, lend further weight to the view that the formation of Jerusalem Artichoke entailed the exclusive contribution of perennial sunflowers, and not *H. annuus*.

The organellar phylogenies rooted with *H. annuus* did not recover any perennial sunflower species as reciprocally monophyletic (Fig. 2). Incomplete lineage sorting (ILS) and reticulation, two alternative but not mutually exclusive processes, can be invoked to explain this pattern. Caused by the retention of ancestral polymorphism, ILS should be common in perennial sunflowers, given their recent, rapid radiation (Schilling, 1997; Timme *et al.*, 2007). Because ILS is stochastic in nature, it should result in discordant associations between accessions of different species, within and between the two organellar phylogenies. As expected, such discordant associations are a pervasive occurrence across both plastid and mitochondrial phylogenies (Fig. 2). A similar trend was found for the diploid-only subset (Fig. S7). Given that diploid-only phylogenies exclude the contribution of polyploid taxa of possible reticulate ancestry, this finding lends further support to the view that ILS is a major contributor to the discordances observed.

In contrast with ILS, reticulation should result in systematic associations between species, reflecting the prevalence of post-speciation organelle capture among pairs of taxa that are interfertile and/or the maternal ancestry of hybrid species with extant progenitors. Such consistent associations include those between *H. giganteus* and *H. decapetalus* accessions (Fig. 2). Two of these groupings were corroborated by geography (Fig. 2), indicating that they are likely instances of recent organelle capture, a phenomenon that is widespread in the genus (Rieseberg & Soltis, 1991). The only groupings of Jerusalem Artichoke accessions recovered repeatedly across analytical methods and organellar phylogenies are those with *H. hirsutus* and *H. divaricatus* (Fig. 2).

The survey of cytoplasmic genomes further revealed high levels of organellar genetic variation in Jerusalem Artichoke. Each accession had a unique plastid and mitochondrial haplotype (Fig. 2). The level of sequence divergence between Jerusalem Artichoke organellar haplotypes was also within the range of those recovered between the geographically diverse accessions of

other perennial sunflowers (Fig. 3). This is in stark contrast with the nearly complete lack of plastid variation reported in polyploids thought to have single origins (Guggisberg *et al.*, 2006; Slotte *et al.*, 2006). Under the assumption of Jerusalem Artichoke formation through a single genetic event, post-speciation organelle capture from other perennial sunflowers could be invoked as the source of this variation. However, given that most perennial sunflower species are diploid or tetraploid (Heiser *et al.*, 1969), and considering that strong pre- and post-zygotic barriers are typically associated with inter-cytype gene exchange (Husband & Sabara, 2003), organelle capture from other species is expected to be limited in Jerusalem Artichoke. An alternative, more plausible explanation is that, similar to many other polyploid taxa (reviewed in Soltis & Soltis, 1999), the Jerusalem Artichoke experienced multiple independent origins, each time sequestering different organellar haplotype combinations from its maternal parent.

The rDNA phylogenies showed, in agreement with previous studies (Schilling *et al.*, 1998; Timme *et al.*, 2007), that no single rDNA region can resolve relationships among all perennial sunflowers (Fig. S8). The concatenated rDNA phylogeny was nevertheless highly informative, providing unprecedented resolution for this group. Notably, most taxa that have long been recognized as distinct based on morphology formed monophyletic groups, some with high support (Fig. 4). Two major clades, A and B, were recovered (Fig. 4). Clade A comprised all *H. giganteus* and *H. decapetalus* accessions, in line with organellar phylogenies which repeatedly group these species. Within clade B, the morphologically distinct *H. maximiliani* was recovered as highly

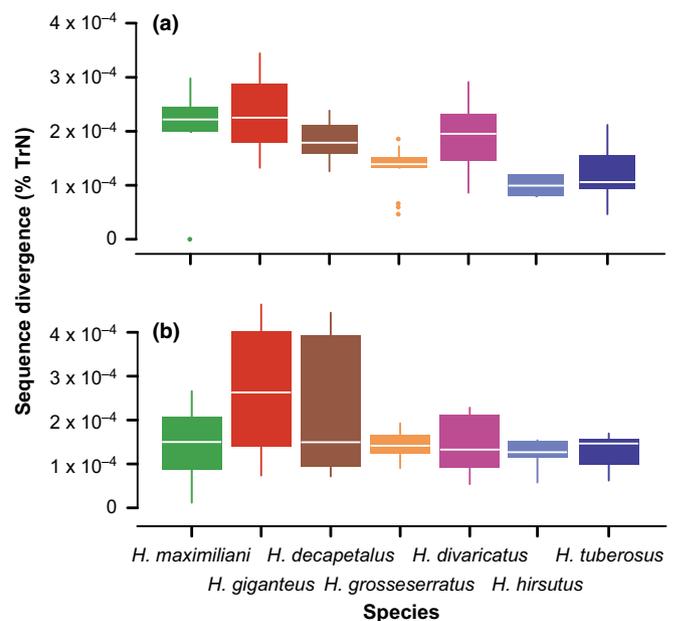


Fig. 3 Box plots of sequence divergences calculated between all pairs of accessions within each species for (a) whole-plastome haplotypes (151 551 bp) and (b) partial mitochondrial haplotypes (196 853 bp). The line within each box is the median. The box spans the interquartile range, and whiskers extend to extreme values. Filled circles indicate outliers at $> 1.5 \times$ the interquartile range. *Helianthus strumosus* was excluded from this analysis as only one accession was available for this species.

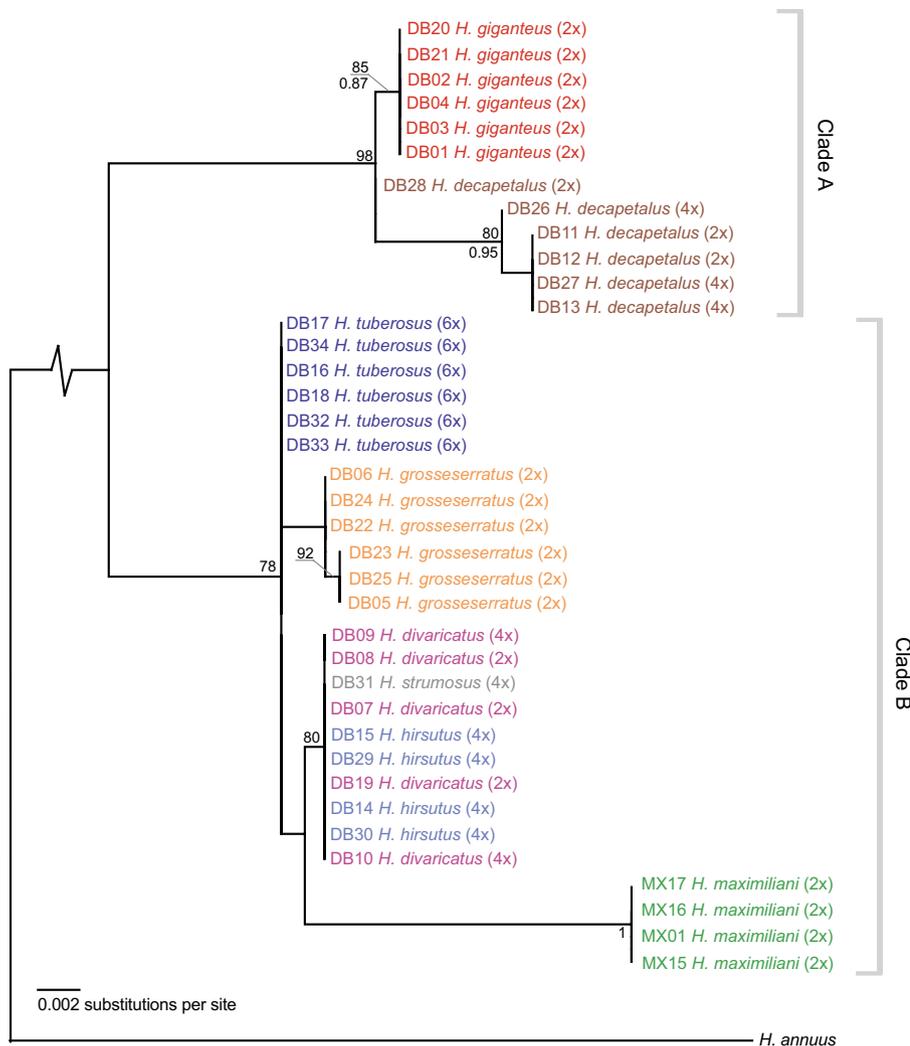


Fig. 4 Maximum likelihood phylogenetic reconstruction of concatenated rDNA sequences (8710 bp) for the perennial *Helianthus* accessions sequenced. Support is shown for nodes with Shimodaira–Hasegawa-like (SH-like) values > 70% (above) and Bayesian posterior probabilities > 0.7 (below).

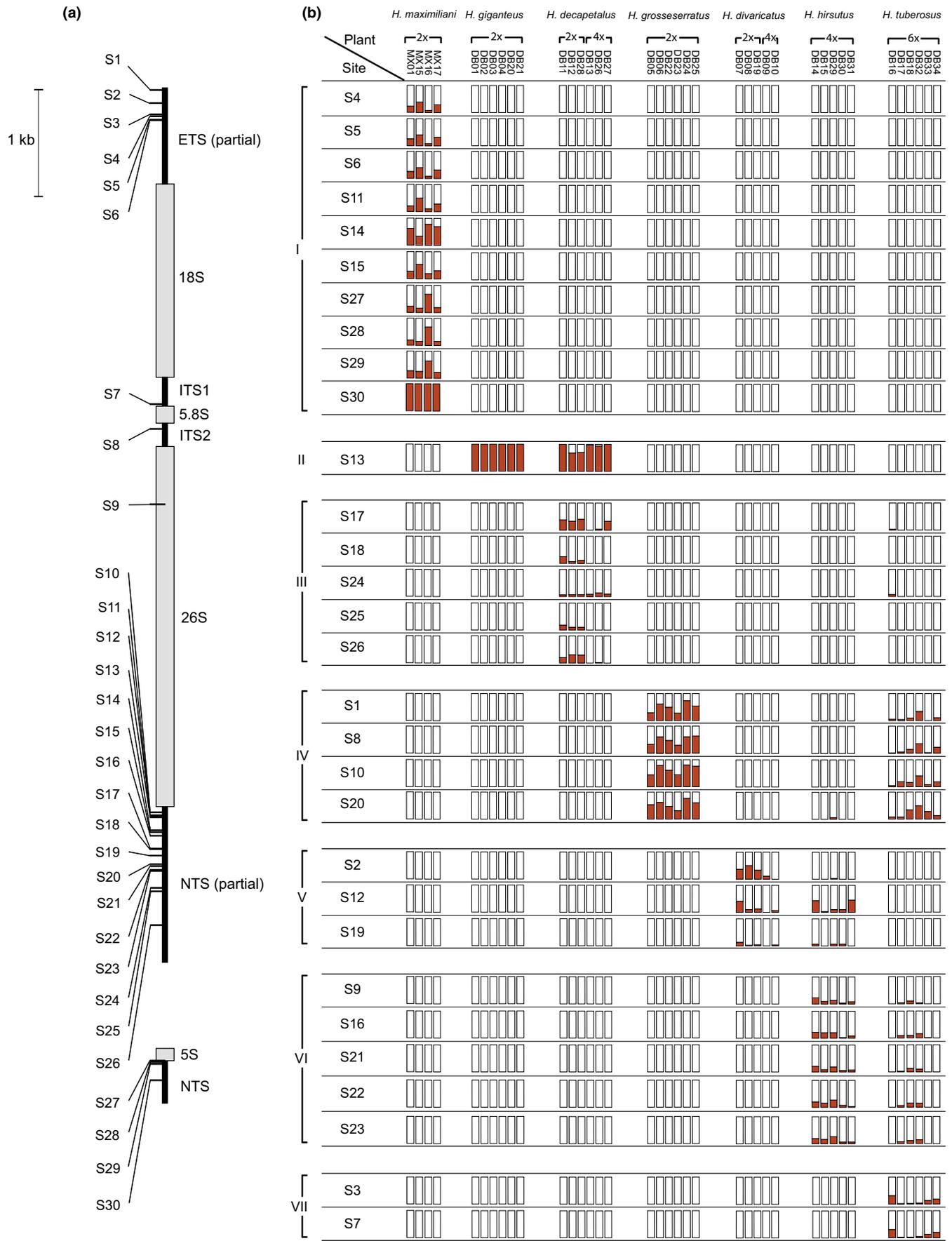
divergent and monophyletic. All *H. divaricatus* and *H. hirsutus* accessions, together with the tetraploid accession of *H. strumosus*, formed another group within clade B (Fig. 4). The grouping of *H. hirsutus* with *H. divaricatus* represents the first molecular phylogenetic support of the morphology-based assumption that *H. hirsutus* is an autotetraploid of *H. divaricatus* (Heiser *et al.*, 1969). However, in consideration of the fact that rDNA may underestimate the frequency of allopolyploid speciation events (Kim *et al.*, 2008), this possibility should be investigated further, to exclude the possibility that divergent rDNA arrays in *H. hirsutus* were homogenized to the *H. divaricatus* type. The placement of the *H. strumosus* accession with *H. divaricatus* and *H. hirsutus* is also in agreement with previous taxonomic work. Notably, tetraploid *H. strumosus* was proposed to be included with *H. hirsutus*, based on its high morphological resemblance to *H. hirsutus*, and the fact that its cross with *H. hirsutus* results in highly fertile progeny (Heiser *et al.*, 1969; Rogers *et al.*, 1982). In

line with this observation, for the remaining analyses, we treated the *H. strumosus* accession as *H. hirsutus*.

The Jerusalem Artichoke accessions were part of a polytomy within clade B, and were closely related to the monophyletic *H. divaricatus*/*H. hirsutus* and *H. grosseserratus* clades (Fig. 4). This indicates that few autapomorphies separate the Jerusalem Artichoke from *H. hirsutus* and *H. grosseserratus*, the two species considered as its most likely progenitors based on morphology and overlapping geographical ranges (Heiser, 1976). The unresolved phylogenetic placement of Jerusalem Artichoke accessions relative to *H. hirsutus* and *H. grosseserratus* further indicates the possibility that Jerusalem Artichoke rDNA contains alleles that are diagnostic for each of these putative progenitors, which have not been homogenized by concerted evolutionary forces.

To test this hypothesis, we defined diagnostic alleles as those that are present in all accessions of a species at the lowest ploidy level. Because these alleles must co-occur in populations sampled

Fig. 5 Survey of diagnostic rDNA polymorphism with position of diagnostic sites along the 35S and 5S rDNA regions (a) and allelic profiles for the diagnostic sites (S1–S30) identified (b). All sites were bi-allelic. Bar plots show the relative proportion of the common allele (white segments) and the species-diagnostic allele (red segments) in each accession and site. Diagnostic sites are grouped by species: *Helianthus maximiliani* (I), *H. giganteus* (II), *H. decapetalus* (III), *H. grosseserratus* (IV), *H. divaricatus* (V), *H. hirsutus* (VI) and *H. tuberosus* (VII).



from disparate geographical regions (Fig. 1), they probably originated early in the formation of each species and would represent a small fraction of the phylogenetic variation analyzed here. Nonetheless, they should be highly informative, particularly with regard to the ancestry of taxa that arose via hybridization. In all, we identified 30 diagnostic sites across the rDNA regions (Fig. 5). In agreement with our expectation formulated on the basis of the phylogenetic reconstruction, the Jerusalem Artichoke was revealed as containing diagnostic sites from both *H. hirsutus* and *H. grosseserratus*. All alleles diagnostic of *H. hirsutus* and *H. grosseserratus* are present in Jerusalem Artichoke (Fig. 5b). By contrast, with the exception of two alleles diagnostic of *H. decapetalus*, which were present at low frequency in one Jerusalem Artichoke accession, no alleles diagnostic of other putative progenitors segregated in Jerusalem Artichoke. Lastly, only two alleles present in all Jerusalem Artichoke accessions were not observed in any putative parental species.

The phylogenetic placement of Jerusalem Artichoke (Fig. 4), as well as the pattern of hybridity revealed for the geographically diverse accessions analyzed here (Fig. 5b), indicate that a monophyletic, autopolyploid origin of this species is highly unlikely. Instead, our results provide strong support that the origin of Jerusalem Artichoke was polyphyletic, involving hybridization between *H. hirsutus* and *H. grosseserratus*. Among the two polyploidization scenarios that can be characterized as polyphyletic, allopolyploidization and auto-allopolyploidization, the latter is the most plausible according to the results presented here, which indicate that the speciation event involved the merger of two duplicate genomes contributed by the *H. hirsutus* parent, and a third differentiated genome contributed by the *H. grosseserratus* parent. This auto-allopolyploidization scenario is also in agreement with previous cytogenetic observations pointing to a high degree of homology between two of the three chromosome complements of Jerusalem Artichoke (Kostoff, 1939).

Conclusions

The origin of Jerusalem Artichoke, a tuber-producing species that is widely grown as a cultivated plant, has long fascinated botanists. The dataset and analyses presented here provide strong genetic evidence that the origin of Jerusalem Artichoke is polyphyletic, from perennial sunflowers in central-eastern North America. The likely progenitors of this species, as indicated by additive patterns of rDNA variation, are the Hairy Sunflower (*H. hirsutus*), which is supported as a likely autotetraploid of *H. divaricatus*, and the diploid Sawtooth Sunflower (*H. grosseserratus*). Additional information was provided by organellar phylogenies. Notably, high levels of organellar genome variation indicate that Jerusalem Artichoke probably experienced recurrent formation. Furthermore, maternal origins of Jerusalem Artichoke appear to have been unidirectional, from *H. hirsutus*. This conclusion is supported by the fact that, although Jerusalem Artichoke–*H. hirsutus* groupings were recovered repeatedly across analytical methods and organellar phylogenies, there was no case in which organellar genomes of Jerusalem Artichoke were grouped with those of *H. grosseserratus*.

This information can be used to direct efforts of germplasm preservation for Jerusalem Artichoke and its wild species progenitors, and should form the foundation of future improvement programs aiming to add novel valuable diversity in Jerusalem Artichoke cultivars from closely related congeners. Our findings also provide a previously lacking evolutionary framework that allows us to investigate the evolution and genetic architecture of perennial life habit and tuber production in sunflowers. Beyond these considerations, the results presented here highlight the promise and applicability of next-generation sequencing technologies in general, and the genome skimming approach in particular, for the resolution of species boundaries, origins and relationships in previously intractable polyploid complexes.

Acknowledgements

We thank members of the Rieseberg laboratory for stimulating discussions, as well as Andrew A. Forbes, Michael B. Kantar, the subject editor Hongzhi Kong and three anonymous reviewers for providing valuable comments that improved earlier versions of the manuscript. We also thank Christopher J. Grassa for assistance with identifying likely integrants of plastid DNA in the mitochondrial genome. We are indebted to the US Department of Agriculture for providing the accessions used in this study, to Anastasia Kuzmin for assisting with sequencing, and to Jaroslav Doležel for providing the standards used for flow cytometry. This work was supported by a *Natural Sciences and Engineering Research Council* (NSERC) Vanier CGS and Killam Doctoral Fellowship to D.G.B. and NSERC grant (327475) to L.H.R. The data associated with this publication are available from Dryad (doi: 10.5061/dryad.138vs).

References

- Anisimova M, Gascuel O. 2006. Approximate likelihood ratio test for branches: a fast, accurate and powerful alternative. *Systematic Biology* 55: 539–552.
- Anisimova IN. 1982. [Nature of the genomes in polyploid sunflower species]. *Byulleten' Vsesoyuznogo Ordena Lenina i Ordena Druzhby Narodov Instituta Rastenievodstva Imeni N.I. Vavilov* 118: 27–29.
- Atlagić J, Dozet B, Škorić D. 1993. Meiosis and pollen viability in *Helianthus tuberosus* L. and its hybrids with cultivated sunflower. *Plant Breeding* 111: 318–324.
- Atlagić J, Škorić D. 2006. Cytogenetic study of hexaploid species *Helianthus tuberosus* and its F1 and BC1F1 hybrids with cultivated sunflower, *H. annuus*. *Genetika* 38: 203–213.
- Bajpai PK, Bajpai P. 1991. Cultivation and utilization of Jerusalem artichoke for ethanol, single cell protein, and high fructose syrup production. *Enzyme and Microbial Technology* 13: 359–362.
- Chandler JM. 1991. Chromosome evolution in sunflower. In: Tsuchiya P, Gupta PK, eds. *Chromosome engineering in plants: genetics, breeding, evolution, part B*. Amsterdam, the Netherlands: Elsevier, 229–249.
- Cheng Y, Zhou W, Gao C, Lan K, Gao Y, Wu Q. 2009. Biodiesel production from Jerusalem artichoke (*Helianthus tuberosus* L.) tuber by heterotrophic microalgae *Chlorella protothecoides*. *Journal of Chemical Technology and Biotechnology* 84: 777–781.
- Darlington CD. 1956. *Chromosome botany*. London, UK: Allen and Unwin.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M *et al.* 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 43: 491–498.

- Doležel J, Greilhuber J, Suda J. 2007. Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols* 2: 2233–2244.
- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.
- Emerson KJ, Merz CR, Catchen JM, Hohenlohe PA, Cresko WA, Bradshaw WE, Holzapfel CM. 2010. Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences, USA* 107: 16196–16200.
- Gentzbittel L, Perrault A, Nicolas P. 1992. Molecular phylogeny of the *Helianthus* genus, based on nuclear restriction fragment-length-polymorphism (RFLP). *Molecular Biology and Evolution* 9: 872–892.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al.* 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644–652.
- Guggisberg A, Bretagnolle F, Mansion G. 2006. Allopolyploid origin of the Mediterranean endemic, *Centaurium bianoris* (Gentianaceae), inferred by molecular markers. *Systematic Botany* 31: 368–379.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* 52: 696–704.
- Heiser CB. 1976. *The sunflower*. Norman, OK, USA: University of Oklahoma Press.
- Heiser CB, Smith DM. 1964. Species crosses in *Helianthus*: II. Polyploid species. *Rhodora* 66: 344–358.
- Heiser CB, Smith DM, Clevenger S, Martin WC. 1969. The North American sunflowers. *Memoirs of the Torrey Botanical Club* 22: 1–218.
- Husband BC, Sabara HA. 2003. Reproductive isolation between autotetraploids and their diploid progenitors in fireweed, *Chamerion angustifolium* (Onagraceae). *New Phytologist* 161: 703–713.
- Kane NC, Cronk Q. 2008. Botany without borders, barcoding in focus. *Molecular Ecology* 17: 5175–5176.
- Kane NC, Sveinsson S, Dempewolf H, Yang JY, Zhang D, Engels JM, Cronk Q. 2012. Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *American Journal of Botany* 99: 320–329.
- Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics* 9: 286–298.
- Kays SJ, Nottingham SF. 2008. *Biology and chemistry of the Jerusalem Artichoke: Helianthus tuberosus L.* Boca Raton, FL, USA: CRC Press.
- Kim S-T, Sultan SE, Donoghue MJ. 2008. Allopolyploid speciation in *Persicaria* (Polygonaceae): insights from a low-copy nuclear region. *Proceedings of the National Academy of Sciences, USA* 105: 12370–12375.
- Kleessen B, Schwarz S, Boehm A, Fuhrmann H, Richter A, Henle T, Krueger M. 2007. Jerusalem artichoke and chicory inulin in bakery products affect faecal microbiota of healthy volunteers. *British Journal of Nutrition* 98: 540–549.
- Kosaric N, Cosentino GP, Wiczorek A, Duvnjak Z. 1984. The Jerusalem Artichoke as an agricultural crop. *Biomass* 5: 1–36.
- Kostoff D. 1934. A contribution to the meiosis of *Helianthus tuberosus* L. *Zeitschr für Pflanzenzüchtung* 19: 423–438.
- Kostoff D. 1939. Autosynthesis and structural hybridity in F1-hybrid *Helianthus tuberosus* L. × *Helianthus annuus* L. and their sequences. *Genetica* 21: 285–300.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357–359.
- Long RW. 1955. Hybridization between the perennial sunflowers *Helianthus salicifolius* A. Dietr. and *H. grosseserratus* Martens. *American Midland Naturalist* 54: 61–64.
- Malinska H, Tate JA, Matyasek R, Leitch AR, Soltis DE, Soltis PS, Kovarik A. 2010. Similar patterns of rDNA evolution in synthetic and recently formed natural populations of *Tragopogon* (Asteraceae) allotetraploids. *BMC Evolutionary Biology* 10: 291.
- Malinska H, Tate JA, Mavrodiev E, Matyasek R, Lim KY, Leitch AR, Soltis DE, Soltis PS, Kovarik A. 2011. Ribosomal RNA genes evolution in *Tragopogon*: a story of New and Old World allotetraploids and synthetic lines. *Taxon* 60: 348–354.
- Milne I, Bayer M, Cardle L, Shaw P, Stephen G, Wright F, Marshall D. 2010. Tablet-next generation sequence assembly visualization. *Bioinformatics* 26: 401–402.
- Rieseberg LH, Soltis DE. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* 5: 65–84.
- Roberfroid MB. 2007. Inulin-type fructans: functional food. *Journal of Nutrition* 137: 2493S–2502S.
- Rogers C, Thompson T, Seiler GJ. 1982. *Sunflower species of the United States*. Bismark, ND, USA: National Sunflower Association.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.
- Sánchez R, Serra F, Tárraga J, Medina I, Carbonell J, Pulido L, de María A, Capella-Gutiérrez S, Huerta-Cepas J, Gabaldón T *et al.* 2011. Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. *Nucleic Acids Research* 39: 1–5.
- Schilling EE. 1997. Phylogenetic analysis of *Helianthus* (Asteraceae) based on chloroplast DNA restriction site data. *Theoretical and Applied Genetics* 94: 925–933.
- Schilling EE, Linder CR, Noyes R, Rieseberg LH. 1998. Phylogenetic relationships in *Helianthus* (Asteraceae) based on nuclear ribosomal DNA internal transcribed spacer region sequence data. *Systematic Botany* 23: 177–187.
- Schwarzbach AE, Rieseberg LH. 2002. Likely multiple origins of a diploid hybrid sunflower species. *Molecular Ecology* 11: 1703–1715.
- Slotte T, Ceplitis A, Neuffer B, Hurka H, Lascoux M. 2006. Intrageneric phylogeny of *Capsella* (Brassicaceae) and the origin of the tetraploid *C. bursa-pastoris* based on chloroplast and nuclear DNA sequences. *American Journal of Botany* 93: 1714–1724.
- Smith JE. 1807. *An introduction to physiological and systematic botany*. London, UK: Longman, Hurst, Reese, Orme, Paternoster Row, White.
- Soltis DE, Soltis PS. 1989. Allopolyploid speciation in *Tragopogon*: insights from chloroplast DNA. *American Journal of Botany* 76: 1119–1124.
- Soltis DE, Soltis PS. 1999. Polyploidy: recurrent formation and genome evolution. *Trends in Ecology & Evolution* 14: 348–352.
- Straub SC, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. 2012. Navigating the tip of the genomic iceberg: next generation sequencing for plant systematics. *American Journal of Botany* 99: 349–364.
- Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* 10: 512–526.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28: 2731–2739.
- Timme RE, Simpson BB, Linder CR. 2007. High-resolution phylogeny for *Helianthus* (Asteraceae) using the 18S-26S ribosomal DNA external transcribed spacer. *American Journal of Botany* 94: 1837–1852.
- Wagner CE, Keller I, Wittwer S, Selz OM, Mwaiko S, Greuter L, Sivasundar A, Seehausen O. 2013. Genome-wide RAD sequence data provides unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology* 22: 787–798.
- Wendel JF, Schnabel A, Seelanan T. 1995. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proceedings of the National Academy of Sciences, USA* 92: 280–284.
- Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252–3255.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* 18: 821–829.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Number of single nucleotide polymorphisms (SNPs) called vs ploidy setting used in Unified Genotyper for the 35S segment in all species and accessions.

Fig. S2 Number of single nucleotide polymorphisms (SNPs) called vs ploidy setting used in Unified Genotyper for the 5S segment in all species and accessions.

Fig. S3 Number of singletons and polymorphism shared between two or more samples for each surveyed region in all species and accessions.

Fig. S4 Unrooted maximum likelihood phylogeny of complete plastid genomes.

Fig. S5 Unrooted maximum likelihood phylogeny of partial mitochondrial genomes.

Fig. S6 Unrooted maximum likelihood phylogeny of concatenated rDNA sequences.

Fig. S7 Maximum likelihood phylogenies of complete plastid genomes and partial mitochondrial genomes for diploid-only accessions.

Fig. S8 Maximum likelihood phylogenies of 35S rDNA and 5S rDNA regions.

Table S1 Details of samples used for sequencing

Table S2 Summary of sequencing data, including reads that passed quality control

Table S3 Total reads and mean coverage for the surveyed high-copy regions

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.



About *New Phytologist*

- *New Phytologist* is an electronic (online-only) journal owned by the New Phytologist Trust, a **not-for-profit organization** dedicated to the promotion of plant science, facilitating projects from symposia to free access for our Tansley reviews.
- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as ready' via *Early View* – our average time to decision is <25 days. There are **no page or colour charges** and a PDF version will be provided for each article.
- The journal is available online at Wiley Online Library. Visit **www.newphytologist.com** to search the articles and register for table of contents email alerts.
- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk) or, if it is more convenient, our USA Office (np-usaoffice@ornl.gov)
- For submission instructions, subscription and all the latest information visit **www.newphytologist.com**