Molecular Phylogenetics and Evolution 127 (2018) 834–842

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Editor's Choice Article

Phylogenetics of *Camelina* Crantz. (Brassicaceae) and insights on the origin of gold-of-pleasure (*Camelina sativa*)

Jordan R. Brock[a,*], Ali A. Dönmez[b], Mark A. Beilstein[c,*], Kenneth M. Olsen[a]

[a] Department of Biology, Washington University in St. Louis, St. Louis, MO 63130, USA
[b] Department of Biology, Faculty of Science, Hacettepe University, Ankara, Turkey
[c] School of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA

ARTICLE INFO

Keywords:
Camelina
Domestication
Genome size
Phylogeny
Polyploidy
RADseq

ABSTRACT

Camelina sativa (false flax or gold-of-pleasure) is an Old World oilseed crop that fell out of use in the mid 20th Century but has recently gained renewed interest as a biofuel source. The crop is hexaploid, and its relationship to its diploid and polyploid congeners has remained unresolved. Using 54 accessions representing five species sampled across *Camelina*'s center of diversity in Turkey and the Caucasus, we performed phylogenetic and genetic diversity analyses using RADseq genotyping and ITS sequencing. Flow cytometry was performed to assess relationships between genome size and phylogenetic groupings. Accessions fell into distinct, highly-supported clades that accord with named species, indicating that morphological characters can reliably distinguish members of the genus. A phylogenetically distinct lineage from Turkey may represent a currently unrecognized diploid species. In most analyses, *C. sativa* accessions nest within those of *C. microcarpa*, suggesting that the crop is descended from this wild hexaploid species. This inference is further supported by their similar genome size, and by lower genetic diversity in *C. sativa*, which is consistent with a domestication bottleneck. These analyses provide the first definitive phylogeny of *C. sativa* and its wild relatives, and they point to *C. microcarpa* as the crop's wild ancestor.

1. Introduction

Camelina sativa (L.) Crantz., also known as gold-of-pleasure or false flax, is an Old World oilseed crop with newfound and growing importance as a biofuel (Moser, 2010; Iskandarov et al., 2014). Described by Nikolai Vavilov as a 'secondary crop' (Vavilov, 1987), it is believed to have first been present in crop fields as an agricultural weed before it eventually came to be intentionally cultivated in its own right (Zohary et al., 2012). Seed remains of *Camelina* sp. have been found amongst flax seeds in several Neolithic and Chalcolithic archaeological sites (Nesbitt, 1996; Bouby, 1998), supporting its initial presence in flax and cereal fields as a crop weed. Although the timing of the transition from weed to domesticate is unresolved, a long history of *C. sativa* presence in Europe and Asia Minor is supported by many archaeobotanical studies (Van Zeist, 1981; Miller, 1991; Kroll, 2000; Dönmez and Belli, 2007; Hovsepyan and Willcox, 2008). *Camelina sativa* remained an important and widespread oilseed crop in much of Europe until the mid-20th Century, after which it was largely abandoned in favor of rapeseed (*Brassica napus* L.) and other oilseed crops. The evolutionary origins of *C. sativa* are unknown. The crop may be descended from conspecific wild populations; alternatively, *C. sativa* may not exist as a wild species and may instead represent a domesticated form of a different *Camelina* species.

In recent decades, the emergence of plant-based biofuel production has spurred a flurry of renewed interest in *C. sativa*. With fatty acid (FA) composition that is high in long-chain hydrocarbons, *C. sativa* seed oil is particularly well suited for aviation biofuel production and has been shown to achieve a 75% reduction in CO_2 emissions relative to traditional petroleum jet fuel (Shonnard et al., 2010). *Camelina sativa* also is resistant to many pathogens of cruciferous crops (Séguin-Swartz et al., 2009), tolerant to drought and cold, and capable of being grown in marginal and saline soils while requiring less fertilizer, water, and pesticides than other oilseed crops (Moser, 2010). With a seed oil content between 36% and 47% by weight, yields of *C. sativa* oil range between 540 and 1410 kg/ha, comparable to rapeseed (Moser, 2012). Naturally high omega-3 FA content in the seeds of *C. sativa* provides an added bonus for feedstock industries, specifically for poultry and salmon production which benefit from supplemented omega fatty acids. *Camelina sativa* breeding has been further facilitated by the development of an efficient transformation protocol (Lu and Kang, 2008, 2011;

* Corresponding authors.
E-mail addresses: brock@wustl.edu (J.R. Brock), mbeilstein@email.arizona.edu (M.A. Beilstein).

https://doi.org/10.1016/j.ympev.2018.06.031
Received 18 April 2018; Received in revised form 18 June 2018; Accepted 18 June 2018
Available online 19 June 2018
1055-7903/

Ruiz-Lopez et al., 2014), its close phylogenetic relationship to *Arabidopsis thaliana* (L.) Heynh. (Beilstein et al., 2006, 2008), and the recent publication of a reference genome (Kagale et al., 2014).

Despite the availability of a reference genome and a well-developed molecular toolkit, efforts to enhance this emerging biofuel crop are limited by low genetic and phenotypic diversity in available *C. sativa* lines. Much of the crop's varietal diversity was lost in the 20th Century, as European oilseed production shifted to rapeseed, sunflower and other species (European Commission, 2017). This loss of crop germplasm likely compounded what was already low diversity in this self-fertilizing species (Vollmann et al., 2005). One potential way to enhance natural variation in *C. sativa* would be through "wide crosses" – i.e., hybridization with reproductively compatible wild germplasm. At least some wild *Camelina* species are known to be sexually compatible with *C. sativa* (Séguin-Swartz, 2013), and genetic introgression could prove an effective method to increase genetic diversity and introduce agronomically valuable traits (e.g., drought tolerance and shorter generation time). However, little is known about the diversity in the rest of the *Camelina* genus, and even basic questions about species numbers and their evolutionary relationships have remained largely unaddressed. *Camelina* species are notoriously difficult to distinguish on the basis of morphology (Davis, 1988), and no genome-wide, genus-wide, molecular systematic studies have been undertaken. Knowledge of phylogenetic relationships is further complicated by chromosome number variation within the genus and the likelihood that some extant species evolved through past allo- and/or autopolyploidization events. For *C. sativa*, existing publicly available germplasm collections are composed almost entirely of cultivated varieties, and from these collections it has remained unclear whether there are any true wild populations of *C. sativa* that exist outside of human-mediated habitats.

The genus *Camelina* Crantz. comprises an estimated 5–10 species, including annuals and biennials, all of which are native to the Irano-Turanian floristic region (Al-Shehbaz & Beilstein, 2010). Some species, including *C. hispida* Boiss. and *C. laxa* C. A. Mey, have restricted geographical distributions within the range of highest species diversity in Turkey and the Caucasus. Others have become established worldwide as cosmopolitan weeds; these include *C. sativa*, *C. microcarpa* Andrz. ex DC. and *C. rumelica* Velen. Variation in genome size between species, as evidenced by flow cytometry (Hutcheon et al., 2010) and chromosome counts (BrassiBase, Koch et al., 2012), are consistent with variation in ploidy. Two species, *C. laxa* and *C. hispida* are known to have chromosome counts consistent with diploidy, with basal chromosome counts of n = 6 and n = 7, respectively (Maassoumi, 1980; Galasso et al., 2015). Other species appear to be polyploid; these include *C. rumelica*, putatively tetraploid (n = 12, 13) (Maassoumi, 1980; Galasso et al., 2015), *C. microcarpa*, putatively hexaploid (n = 20), and *C. sativa*, hexaploid (n = 20) (Gehringer et al., 2006; Francis & Warwick, 2009). Much within-species variation in chromosome counts has also been reported in the older literature; this could represent true intraspecific ploidy variation or simply inaccurate records reflecting past taxonomic misidentifications (BrassiBase, Koch et al., 2012).

For *C. sativa*, whole genome sequencing indicates the existence of three minimally diverged subgenomes (n = 6 + 7 + 7), but with one subgenome appearing somewhat distinct, thereby suggesting at least one allopolyploid hybridization event in its evolution (Kagale et al., 2014). However, little is known about which taxa might have contributed to such hybridization events, or whether these taxa are extant. Gene trees of *fatty acid desaturase 2* (*FAD2*) and *fatty acid elongase 1* (*FAE1*) both show two out of three of the paralogous gene copies as sister between *C. sativa* and *C. microcarpa* with high support (Hutcheon et al., 2010), suggesting the possible origin of *C. sativa* via domestication of *C. microcarpa*. Alternatively, this pattern could reflect common ancestry of two subgenomes but with no direct ancestry of the crop species from *C. microcarpa*. Morphologically, *C. sativa* is very similar to *C. microcarpa*, differing primarily by larger fruit and seed size in the domesticate (Al-Shehbaz & Beilstein, 2010). This phenotypic similarity is potentially consistent with *C. sativa* representing a domesticated form of *C. microcarpa*.

With the goals of elucidating evolutionary relationships in the genus *Camelina* and determining the closest relatives of the crop *C. sativa*, we extensively collected *Camelina* germplasm from throughout Turkey, Georgia and Armenia, regions with the highest *Camelina* species diversity. Using phylogenetic analyses combined with examinations of genome size variation and genetic diversity, we addressed the following questions: (1) What are the phylogenetic relationships within *Camelina*? Do morphologically-based taxonomic designations correspond to evolutionarily distinct lineages within the genus? (2) How does genome size vary among species, and is there evidence of genome size or ploidy variation within species? (3) Is cultivated *C. sativa* descended from conspecific wild populations, or does it represent the domesticated form of a wild congener? If the latter, which species is the most likely wild progenitor of *C. sativa*?

2. Materials and methods

2.1. Collections

Camelina specimens were collected in Turkey in June 2012, 2013 and July 2014. Additional collections were undertaken in Georgia and Armenia in June and July 2013. Specimens were provisionally placed into morphologically determined species groups using the Flora of Turkey treatment of the genus (Davis, 1988). In some cases, accession identity was subsequently changed to reflect growth characters that could be observed in greenhouse-grown seed-borne offspring, particularly in the cases of senescent *C. microcarpa* and *C. rumelica* accessions. Key morphological characters used in *Camelina* species delimitation include size and shape of fruits and seeds, trichome presence and branching pattern, and flower color and size. All wild collected voucher specimens used herein for phylogenetic purposes have been deposited in the University of Arizona Herbarium (ARIZ) or the Missouri Botanical Garden Herbarium (MO), along with duplicates deposited in Hacettepe University (HUB), the National Academy of Sciences of Armenia (ERE), and the Georgian Academy of Sciences (TBI). In addition to vouchers, mature seeds were collected whenever possible and imported to the United States under USDA permit #P37-13-00443. Locality information, GPS coordinates, and elevation were recorded for all collection sites, and GPS coordinates were mapped in QGIS v. 2.14.3 (QGIS Development Team, 2016). Individuals which were taxonomically ambiguous based on morphological characters were given the provisionary designation of *Camelina* sp. The sampling of accessions for molecular work also included plants grown from seed obtained from the USDA GRIN collection (4 PI accessions), as well as one accession from the Brassicaceae Seed Bank at the Technical University of Madrid (UPM, http://www.upm.es/internacional). Information on all accessions used in analyses is presented in the Supporting Information (Table S1).

2.2. ITS sequencing and phylogenetic analysis

Phylogenetic analyses employed two complementary datasets: nuclear ribosomal Internal Transcribed Spacer (ITS) sequences, and genome-wide SNPs generated by ddRADseq. DNA for ITS sequencing was extracted from field-collected and dried leaf tissue or from fresh leaf tissue obtained from greenhouse-grown plants that were germinated from wild-collected seeds, using either the DNeasy Plant Mini Kit (Qiagen, Valencia, California, USA) or NucleoSpin Plant II kit (Macherey-Nagel, Düren, Germany). The Internal Transcribed Spacer (ITS) nuclear ribosomal region was amplified using the ITS1 and ITS4 primer pair (White et al., 1990). To ensure that alternative copies of ITS were recovered when present, PCR fragments were gel-excised and ligated into the pGEM-T Easy Vector (Promega, Madison, Wisconsin, USA), transformed into XL-1 Blu chemically competent cells

(Stratagene, La Jolla, California, USA), and plated on LB plates (X-GAL, IPTG, carbenicillin 50 mg/L). At least two colonies per accession were grown overnight in LB and carbenicillin (0.5 mg/ml) before harvesting plasmids via Plasmid Miniprep-Classic kits (Zymo Research, Irvine, California, USA) and sequencing. A *Capsella rubella* Reut. ITS sequence (AJ232912.1) was used as the outgroup sequence, and BLAST searches from the *C. sativa* reference genome DH55 (Kagale et al., 2014) revealed two unique ITS copies which were included in the analysis. Sequences that were identical or nearly identical were removed, resulting in some accessions being represented by a single sequence and others by two distinct ITS copies. The sequence alignment was generated in Geneious v7.1.4 (Biomatters, Auckland, New Zealand) using MUSCLE (Edgar, 2004). A Maximum Likelihood tree was inferred using RAxML v8.2.4 (Stamatakis, 2014) under the GTR GAMMA model; support for nodes was assessed using 1000 rapid bootstrap replicates. A consensus tree was prepared in Mesquite v3.02 (Maddison & Maddison, 2015). All new ITS sequences have been deposited in the NCBI nucleotide database under GenBank accession numbers MH492943-MH492978.

2.3. ddRADseq library prep and sequencing

A set of 48 accessions representing all five known extant species in the region of highest diversity were chosen for ddRADseq analyses. Of these, 44 accessions were sequenced to sufficient depth to be included in this study. To obtain DNA concentrations sufficient for generating ddRADseq libraries, we used a CTAB extraction protocol modified from Webb and Knapp (1990). Approximately 100 mg of freshly collected leaf material was ground in liquid nitrogen and combined with 800 µl warmed extraction buffer (100 mM Tris, 1.4 M NaCl, 20 mM EDTA, 2% CTAB, 1% PVP, with 0.3% added β-mercaptoethanol). The mixture was incubated for 90 min at 65 °C before the addition of 800 µl 24:1 chloroform:isoamyl alcohol with vortexing. The solution was spun at maximum speed for 10 min and the supernatant removed. Then, 1 ml PE Buffer (50 mM Tris, 10 mM EDTA, 1% CTAB) was mixed with the supernatant and incubated at room temperature for 30 min. The solution was then spun at 10 °C for 10 min at max speed before being incubated at 50 °C for 60 min. A total of 800 µl ice cold 100% EtOH was added to precipitate the DNA. The DNA was then washed twice with 700 µl 70% EtOH, spun at maximum speed, and the resulting pellets dried prior to resuspension in 100 µl of 65 °C TE buffer (10 mM Tris, 1 mM EDTA). DNA was quantified by Qubit fluorometry (Life Technologies, Carlsbad, California, USA).

The sequencing library was prepared via a modified protocol of Peterson et al. (2012). A double digest was used to improve efficiency of obtaining specific fragments across many individuals. For each sample, 500 ng total DNA template was used in a 50 µl ligation reaction consisting of 5 units each of NEB EcoRI-HF and MseI enzymes using the CutSmart buffer (New England Biolabs, Ipswich, Massachusetts, USA). Reactions were incubated for 3 h at 37 °C, and enzymes were then inactivated for 20 min at 80 °C. An annealed adaptor stock (16 µM) was made by combining top and bottom single stranded P1 and P2 adaptors with a 10× annealing buffer (100 mM Tris HCL pH 8, 500 mM NaCl, 10 mM EDTA) in a thermocycler (95 °C for 2 min, ramp cooling to 21 °C over 45 min). Upon optimization of adaptor ligation to DNA fragments, we discovered that a 20-fold excess of adaptors to estimated sticky ends was ideal. For the ligation, 6 µl 10× T4 ligase buffer and 1 µl T4 ligase (400 units/µl) was added with 1.5 µl of each working stock adaptors for a 60 µl total ligation reaction per sample. The ligation reaction was performed in a thermocycler (23 °C for 30 min, 65 °C for 10 min, ramp cool 2 °C every 90 s until 23 °C). Once the adaptors were successfully ligated to DNA fragments, all samples were pooled and cleaned with an AxyPrep Mag bead clean-up kit (Axygen Biosciences, Union City, California, USA) before size selection via Pippin Prep (Sage Science, Beverly, Massachusetts, USA) for fragments between 350 bp and 650 bp. To reduce PCR amplification bias, six low-cycle PCR amplifications were performed, each with 20 ng template, 4 µl 5× Phusion HF buffer, 0.2 µl Phusion HF polymerase (New England Biolabs, Ipswich, Massachusetts, USA), 0.4 µl dNTPs (10 µM) and 1 µl of each RAD primer (10 µM) in a thermocycler (98 °C for 2 min:: 98 °C 10 s, 65 °C 30 s, 72 °C 30 s; 12 cycles:: 72 °C 10 min). The six PCR reactions were pooled together and cleaned with the AxyPrep Mag bead clean-up kit to remove primer dimer. The final library was sequenced for 100 bp paired-end reads on a HiSeq2000 (Illumina, San Diego, California, USA) at the DNASU sequencing facility of Arizona State University. An initial HiSeq2000 run had failed due to low sequence diversity; this issue was remedied by spiking the run with 50% PhiX. Raw sequence reads have been deposited in NCBI under PRJNA477644.

2.4. ddRADseq phylogenetics and diversity measures

The RADseq analysis pipeline *pyRAD* v3.0.66 (Eaton, 2013) was utilized because of its advantages for phylogenetics and in dealing with intrageneric and polyploid data. Within *pyRAD* all combinations of the following parameters were tested: minimum coverage for a cluster (6); maximum number of sites with quality score below 20 (4); clustering threshold 0.86, 0.90, and 0.94; maximum individuals with a shared heterozygous site 1, 2, and 3; ploidy parameter 6 (to account for hexaploids in the dataset); minimum samples in a final locus 8, 12, 16, 20, and 24. Additional tests were conducted using the ploidy parameter 2, to explore its potential effect on the results. The morphologically and phylogenetically distinct species *C. laxa* (137JRB1 and 63JRB3) was designated as the outgroup taxon. Statistics for each parameter set are reported in the Supporting Information (Table S2). To further assess phylogenetic relationships between *C. sativa* and *C. microcarpa*, a reduced set of 16 samples representing all *C. microcarpa* and *C. sativa* and a single *C. hispida* outgroup individual was used in a second analysis. Parameters tested in this reduced comparison were the same as those previously listed except minimum samples in a final locus, which were reduced to 4, 6, 8, and 10. Error rate and heterozygosity statistics were given as an output from *pyRAD*. The Phylip format of the output from *pyRAD* was used to infer trees via RAxML v8.2.4 (Stamatakis, 2014) under the GTR GAMMA model. Support for nodes was assessed for preliminary trees with 100 bootstrap replicates and the final tree using 1000 rapid bootstrap replicates. The consensus of bootstrap replicate trees was generated in Mesquite v3.02 (Maddison & Maddison, 2015) with values projected on the best tree.

Nucleotide diversity, calculated as Watterson's estimator (θ_W), and pairwise F_{ST} measures were calculated for each species using the loci output file from *pyRAD* in DnaSP v. 6.10.04 (Rozas et al., 2017). SNPs recovered from RADseq loci in *pyRAD* were used to generate a PCA with the R package *adegenet* v.2.1.1 (Jombart, 2008).

2.5. Quantifications of genome size

Genome size estimates for 36 accessions from across five species and one putatively new species were produced using an Accuri C6 flow cytometer (BD Biosciences, San Jose, CA, USA). Approximately 100 mg fresh young leaf tissue per accession was collected and chopped along with leaves of an internal standard (either *Arabidopsis thaliana* line Col-0 or *C. sativa* reference genome line DH55; see below) in a 60 mm petri dish on ice with ice-cold modified Galbraith's buffer (45 mM $MgCl_2$, 30 mM sodium citrate, 20 mM MOPS, 0.1% (v/v) Triton X-100, 1% (w/v) PVP-40, pH 7.0), using a fresh razor each time (Galbraith et al., 1983). Large debris was removed via a 30 µm nylon filter. The homogenate was treated with 2.5 µl RNase A (10 mg/ml) for 10 min on ice and stained with 25 µl propidium iodide (1 mg/ml). The stained homogenate was incubated in the dark for 30 min prior to loading on the instrument. All samples examined were measured in biological triplicate. Estimates of 1C genome size were obtained by comparison to the Arabidopsis or *C. sativa* internal standards, which have known genome sizes of 157 Mb (Bennett et al., 2003) and 750 Mb (Hutcheon et al., 2010) respectively. We also independently confirmed the 750 Mb

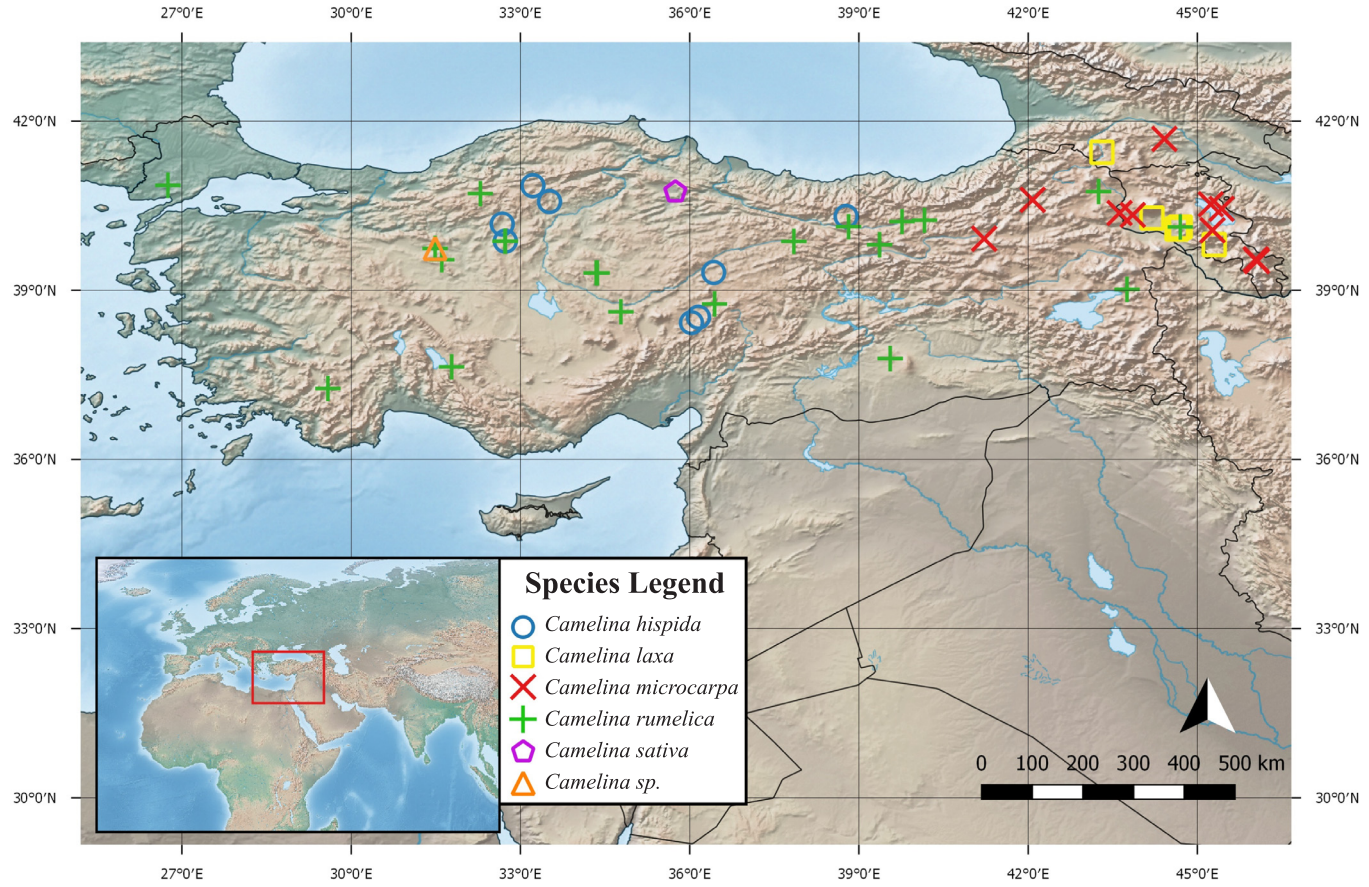

Fig. 1. Distribution of *Camelina* species collected during the summers of 2012, 2013, and 2014.

genome size of the *C. sativa* reference genome line DH55. Genome size estimates for individual accessions were calculated as an average across three biological replicates, and these data were used to assess within-species genome size variation and to calculate species-wide average values.

3. Results

3.1. Collections

Field collections during the summers of 2012, 2013, and 2014 yielded 163 collections representing five *Camelina* species and one putatively new species from a variety of habitats that span Turkey, Georgia, and Armenia (Fig. 1). A subset of these wild collections (N = 45) were used together with previously collected germplasm (N = 8) for the phylogenetic and genome size analyses (Table S1). Although they are listed in the Flora of Turkey, we were unable to find *C. anomala* Boiss. & Hausskn. and the putative species *C. stiefelhagenii* Bornm.; these species may be extinct as they are only represented by a handful of old herbarium specimens. Among those species collected, the two most morphologically distinct are *C. laxa* and *C. hispida*, which have unique raceme morphologies and smaller fruits that distinguish them from other members of the genus. Interestingly, despite extensive searching across a wide range of wild habitats over three years, we were able to find only one *C. sativa* plant growing outside of an agricultural context (accession JRB 153; Table S1); this single plant was found in a roadside Turkish woodland understory habitat, where *C. sativa* is not known to normally occur, and we suspect that it was there as a result of a chance seed introduction. The apparent absence of *C. sativa* populations in the wild is consistent with this species existing only in a domesticated or crop-weed form.

3.2. ITS phylogeny

ITS sequencing yielded low levels of variation within and among *Camelina* species and correspondingly poor phylogenetic resolution in the resulting maximum likelihood tree (Supporting data, Fig. S1). A 712 bp alignment for 38 ITS sequences representing 5 species yielded a total of only 38 phylogenetically informative sites. This lack of variation potentially suggests a relatively recent diversification of the genus. On the ITS tree, a clade comprising all *C. laxa* accessions is highly supported (99% bootstrap support), consistent with its morphological distinctness from other members of the genus. In addition, a clade composed almost exclusively of *C. microcarpa* and *C. sativa* sequences is moderately supported (66% bootstrap support). Beyond these two clades, however, there is no clear resolution for other members of the genus.

3.3. ddRADseq phylogeny

The pyRAD analysis pipeline yielded low estimates of error rate (E = 0.00043) and heterozygosity (H = 0.01099) for the final ddRADseq dataset. A total of 403,660 sequences were recovered, and 80,737 clusters were generated for the 44 accessions. The maximum cluster size was 46 sequences, with an average of 5.0 sequences/cluster. Filtering removed 29,128 singleton sequences (representing 7.2% of sequences and 36.1% of clusters). After final filtering, 17,191 loci were recovered in which at least 8 of the total 44 accessions were represented, yielding a total of 114,605 parsimony informative sites and 12,648 unlinked SNPs.

Regardless of parameter settings across pyRAD runs, maximum likelihood analyses of the resulting data consistently yielded trees composed of four well supported clades: *C. laxa*, *C. rumelica*, *C. hispida*

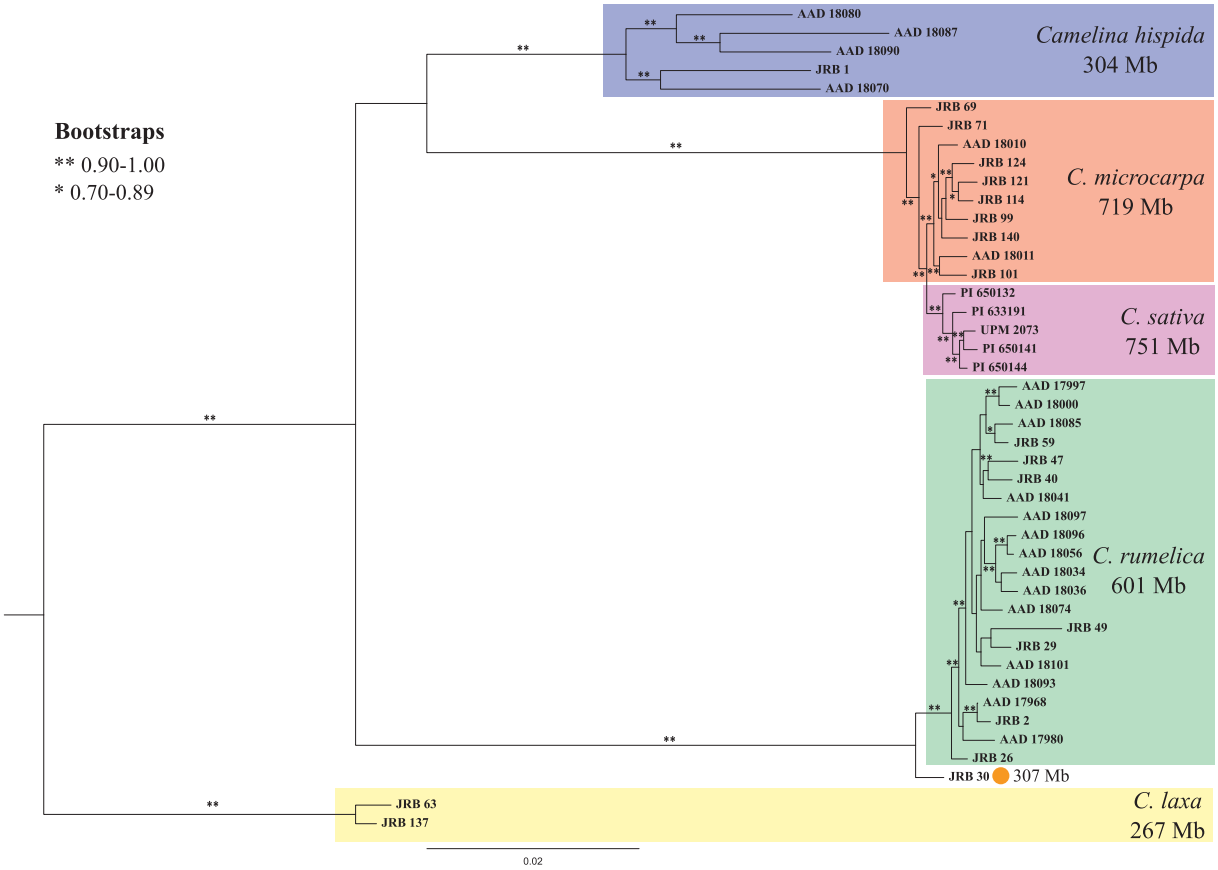

Fig. 2. Maximum Likelihood consensus tree based on ddRADseq dataset analyzed in *pyRAD*. Accessions are identified by collectors' initials and collection number (see Supplementary Table S1). Values under clade names represent average (1C) genome size values as determined by flow cytometry of at least five replicate accessions per species group. Single asterisks represent bootstrap support of 70% or more, while double asterisks represent bootstrap support of 90% or more. Orange circle corresponds to accession JRB 30, a putatively distinct diploid. Branch lengths represented by number of substitutions per site. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(with the exception of a single accession discussed below), and *C. microcarpa* + *C. sativa*. The ML tree is presented in Fig. 2. The clade comprising *C. microcarpa* and *C. sativa* is very strongly supported (100% bootstrap support), indicating that *C. microcarpa* is the closest relative of the domesticated species. Within this clade, the cultivated varieties of *C. sativa* form a well-supported subclade (100% bootstrap support). This pattern is potentially consistent with the domesticated species being descended from wild *C. microcarpa* populations. We also tested several parameter sets, including our final parameter set, using the diploid filter. These all returned congruent topologies with high support (data not shown).

While *C. sativa* was nested within *C. microcarpa* in the majority of parameter sets we tested (Table S2), there were some cases in which the two clades were resolved as sister to one another. To explore their relationships further, we tested a range of parameters on a reduced sample set comprising only *C. sativa* and *C. microcarpa* accessions plus a single outgroup accession (see Methods). These analyses of the hexaploid species almost always resulted in *C. sativa* being nested within *C. microcarpa* with strong support (Table S4). Thus, the majority of phylogenetic outcomes indicate that *C. sativa* is derived from within *C. microcarpa*. We found that a relaxed clustering threshold (0.86) is more appropriate for analyses comparing all species, whereas a more stringent clustering threshold (0.94) was better for the reduced set of closely related hexaploid accessions (Table S4, Fig. S3).

The sister clade to the *C. microcarpa* + *C. sativa* clade in the consensus ML tree is a strongly supported clade comprising all but one of the *C. hispida* accessions; the exception (JRB30) is placed as sister to the *C. rumelica* clade (Fig. 2). This accession, while showing a close morphological resemblance to *C. hispida* (e.g., in the presence of small fruits and large flowers), appears morphologically distinct due to the presence of pubescent racemes. It may thus represent a distinct, currently unrecognized species.

3.4. Genetic diversity and differentiation

Pairwise F_{ST} values between species and measures of within-species nucleotide diversity are presented in Table S3. The differentiation between *C. sativa* and *C. microcarpa* is less than half that of any other species pair (F_{ST} = 0.2091). This lower level of divergence is consistent with the close relationship that would be expected for a crop species and its wild progenitor. Similarly, a PCA generated from SNPs in the final dataset revealed that while most of the wild taxa are clearly separated into groups according to species-designation and putative ploidy, *C. sativa* and *C. microcarpa* are clustered together (Fig. S2). Measures of within-species nucleotide diversity further support the hypothesis that *C. sativa* may be derived from *C. microcarpa*. The diversity of *C. sativa* is lower than that of any other species and approximately 54% that of *C. microcarpa* (θ_W = 0.0032 and 0.0060 for the two species, respectively) (Table S2). This pattern is consistent with a domestication bottleneck if cultivated *C. sativa* were derived from the wild hexaploid species.

3.5. Flow cytometry

Genome size estimates were obtained for 35 accessions representing all five established species groups, and inferred values were highly

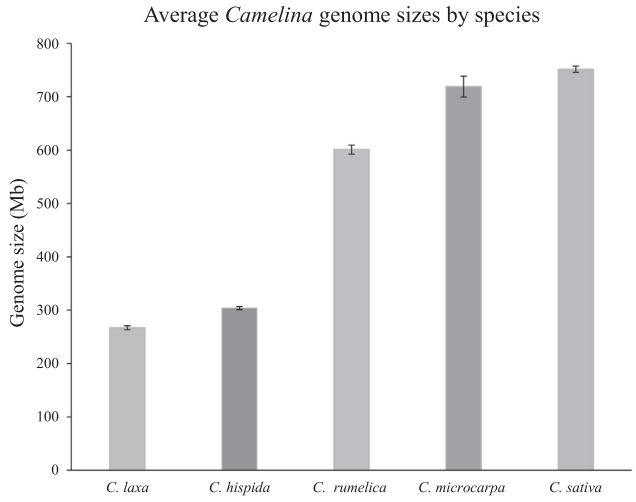

Fig. 3. Flow cytometry inferred genome estimates of *Camelina* species. Means from three biological replicates per accession were averaged per species and displayed as a histogram with standard deviations of species means. Numbers of representatives per species: *C. laxa* ($n = 5$), *C. hispida* ($n = 5$), *C. rumelica* ($n = 10$), *C. sativa* ($n = 5$), *C. microcarpa* ($n = 10$). *C. sp.* accession JRB 30 was found to have an average genome size of 307 Mb (not shown here).

similar for different accessions within each species (Fig. 3). The lowest average genome size was for the putative diploid species *C. laxa* (267 ± 1.6 Mb), while the highest was for the hexaploid *C. sativa* (751 ± 2.7 Mb). The average genome size for *C. microcarpa* (719 ± 6.2 Mb) was similar to that of *C. sativa* but nonetheless significantly different (*t*-test, $P < 0.005$). Chromosome counts for each species in BrassiBase (Kiefer et al., 2014) are concordant with our genome size calculations: genome sizes of 260–320 Mb are ostensible diploids, 580–630 Mb tetraploids, and 690–760 Mb hexaploids. Accordingly, we find all *C. laxa* and *C. hispida* to be in the diploid range, all *C. rumelica* to be in the tetraploid range, and all *C. sativa* and *C. microcarpa* to be hexaploid. These inferred ploidies are also completely concordant with the four major clades identified in the ML tree (Fig. 2). The accession JRB 30, has a genome size of 307 Mb, consistent with accessions found in the diploid range. Taken together, our genome size estimates are consistent with there being ploidy differences among *Camelina* species but little genome size or ploidy variation within species.

4. Discussion

Phylogenetic relationships within the genus *Camelina* have historically been difficult to resolve, owing to subtle morphological character differences among species and ploidy variation that might or might not transcend species boundaries. Our molecular phylogenetic analyses, together with direct assessments of genome size and quantifications of genetic diversity, have permitted key insights into the evolution of this genus. We find that the traditional morphological species designations correspond to distinct evolutionary lineages, and that with the benefit of genome-wide SNP data we can resolve their phylogenetic relationships (Fig. 2). Moreover, we find that ploidy differences (inferred from genome size estimates) also correspond to species boundaries (Fig. 3), with no evidence for within-species ploidy variation. Finally, we find several lines of evidence supporting the hypothesis that the oilseed crop *C. sativa* was domesticated from wild *C. microcarpa* populations (Figs. 2 and 3; see also Figs. S1–S3 and Table S3). Below we discuss these insights together with caveats for understanding the evolution of *Camelina* species.

4.1. Phylogenetic relationships

The phylogeny inferred from ddRADseq data suggests there are four major evolutionarily-diverged clades among the sampled species; these correspond to *C. laxa* (diploid), *C. hispida* (diploid), *C. rumelica* (tetraploid), and *C. microcarpa* + *C. sativa* (hexaploid) (Fig. 2, Fig. S1). Because the ddRADseq dataset did not include an outgroup taxon outside the genus, we designated *C. laxa* as the outgroup for the ddRADseq ML analyses. This choice appears justified given the strong support in the ITS tree for its placement as sister to the other *Camelina* species. This evolutionary relationship is also consistent with other aspects of *C. laxa's* biology. Morphologically, *C. laxa* is characterized by a unique raceme architecture and smaller seed size than its congeners (Davis, 1988), and it had the smallest genome size of any of the species tested (Fig. 3). The identification of *C. rumelica* accessions as a distinct clade is also noteworthy, as the morphological intermediacy of this species has sometimes led to it being considered a "catch-all" taxon designation for difficult-to-identify *Camelina* specimens.

The application of RADseq in non-model organisms allows for genome-wide SNPs to be leveraged for phylogenetic analysis without the need for a reference genome. For the present study, we believe that the assessment of species-level clades is valid, as they predominately hold true throughout the rigorous parameterization we used in the *pyRAD* pipeline (Tables S2, S4). Further studies, including chloroplast phylogenomics and analysis of multiple low-copy nuclear gene trees, may prove useful for verifying these findings. Such follow-up studies could also be useful for identifying maternal sub-genome origins in the polyploid species, an important step in resolving allopolyploid species parentage.

4.2. Genome size and ploidy

Our genome size estimates indicate that ploidy variation in *Camelina* corresponds to species-specific differences, with no evidence for within-species ploidy variation (Fig. 3). This inference was based on sample sizes of 5–10 accessions per species tested. While our sampling was designed to represent the known native geographical distributions of the species (Fig. 1), it is possible that additional intensive sampling might reveal intraspecific ploidy variation that we did not capture with the current sampling. In addition, although chromosome counts have previously been recorded for nearly all extant *Camelina* species, it is important to determine the karyotypes of the accessions analyzed here for verification of the relationship between genome size and inferred chromosome number.

Given that most reliable chromosome count assessments of *C. microcarpa* and *C. sativa* have yielded n = 20, and given the very close relationship of these species in our analyses, it is interesting that the genome size of *C. sativa* (751 Mb) appears to be inflated relative to *C. microcarpa* (719 Mb). More replicate sampling of these two species should be performed to confirm this observation. If real, and if *C. microcarpa* is the wild progenitor of *C. sativa*, this pattern potentially suggests the occurrence of genome expansion during the course of domestication. Genome expansions frequently occur through mobile element proliferation (Kumar and Bennetzen, 1999). Both *C. microcarpa* and *C. sativa* are self-fertile species; their low effective recombination rates, together with bottlenecks during domestication, could potentially have hindered the effectiveness of selection in purging weakly deleterious transposable elements from the *C. sativa* genome. The genome sequence of *C. sativa* comprises 28% transposable elements (Kagale et al., 2014); however, transposable element content in *C. microcarpa* remains unknown. Comparative whole genome sequencing of *C. sativa* and *C. microcarpa* would be useful for testing the hypothesis that transposable element proliferation accounts for these genome size differences.

It is also noteworthy that ploidy variation in *Camelina* is correlated with mating system variation in a pattern consistent with theoretical

predictions for the breakdown of self-incompatibility through polyploidization. The putative diploid species examined here (*C. laxa*, *C. hispida*) are predominantly outcrossing, with little-to-no self-fertility observed in greenhouse experiments (JRB, unpublished data); in contrast, the polyploid species (*C. rumelica*, *C. microcarpa* and *C. sativa*) are all self-compatible. This pattern suggests that the diploid species possess a functioning *S*-locus whose functionality has been lost in their congeners following genome doubling events (Levin, 1983). For members of the Brassicaceae such as *Camelina*, this loss of self-incompatibility may be the result of dominance interactions among *S*-locus haplotypes (Tsuchimatsu et al., 2012). Polyploidization may also have had ecological consequences in this genus. *Camelina* diploids (*C. hispida* and *C. laxa*) have relatively narrow distributions in the Irano-Turanian region, whereas the tetraploid and hexaploid species (*C. sativa*, *C. microcarpa*, *C. rumelica*) have much broader ranges, and have been able to expand into temperate mesic climates around the world. A combination of gene dosage and neofunctionalization may contribute to the adaptive plasticity and ability to colonize new niches in polyploid species (Flagel and Wendel, 2009; te Beest et al., 2012).

Our discovery of the potentially novel diploid (JRB 30) may prove important to future studies aimed at elucidating subgenome origins of polyploid *Camelina* lineages, including the allohexaploid crop species, *C. sativa*. Additional phylogenetic and cytological work must be performed on this accession (and on others which share its morphological characteristics but were not included in this study), to fully understand both their status as a species and relationship to *C. sativa*. As it stands, the present data suggest that the accession JRB 30 is sister to, but distinct from *C. rumelica* (Fig. 2). Thus, it is possible that this accession represents a diploid progenitor species that contributed a subgenome to the tetraploid, *C. rumelica*. However, because this is a single sample, we leave open the possibility that artifacts in the data, or contamination, could have produced this result.

4.3. Origin of *C. sativa* and prospects for germplasm enhancement

The preponderance of data from our analyses support the hypothesis that *C. sativa* is a domesticated form of *C. microcarpa*. Lines of evidence include the following:

(1) Absence of wild *C. sativa* populations. Despite three years of extensive *Camelina* collecting across the center of diversity for the genus (and in a variety of habitats, including pastures, agricultural fields, and steppe), we found no wild *C. sativa* populations. We observed only a single *C. sativa* plant, which, given its atypical location in a woodland understory, we strongly suspect was an accidental introduction. The apparent absence of wild *C. sativa* populations is consistent with this species existing solely in agricultural habitats, where it can be found as a domesticate and agricultural weed. While the absence of wild *C. sativa* could in principle be due to the past extinction of 'true' wild populations of this species, the ability of contemporary *C. sativa* to persist and proliferate worldwide as an agricultural weed would seem to argue against a past extinction scenario for wild populations.

(2) Phylogenetic relationship to *C. microcarpa*. Our ddRADseq ML analysis indicates that, under most parameter settings, *C. sativa* is nested within *C. microcarpa* rather than forming a sister clade (Fig. 2, Tables S2, S4). This pattern is consistent with *C. sativa* having evolved from *C. microcarpa*. Of the 45 parameter combinations employed, those yielding the highest number of parsimony-informative sites all supported a nested relationship, and fewer than one-sixth of all parameter combinations strongly supported a non-nested topology (Table S2). These findings were also upheld by our follow-up analysis, which focused only on the hexaploid species to eliminate any potential confounding effects of cross-ploidy comparisons (Table S4). The support for a nested relationship between *C. sativa* and *C. microcarpa* suggests at a minimum that these two species are far more closely related to each other than either is to any other *Camelina* species in our sampling.

(3) Low genetic differentiation and diversity. The very close relationship between *C. sativa* and *C. microcarpa* is further evidenced by their low pairwise F_{ST} relative to other species pairs (Table S3). In addition, the lower nucleotide diversity of the crop species in relation to *C. microcarpa* is consistent with a genetic bottleneck during domestication. However, it should be noted that this low diversity in *C. sativa* may be equally likely to reflect the recent loss of crop germplasm diversity following 20th Century shifts away from *C. sativa* oilseed production.

(4) Shared ploidy and morphology with *C. microcarpa*. *Camelina microcarpa* is the only other hexaploid (n = 20) species in the genus, and its morphological differences from *C. sativa* are primarily in traits that would have been targets of artificial selection during domestication – most notably, fruit and seed size. Consistent with this idea, archaeobotanical studies indicate the early presence of *C. microcarpa* seeds in Eurasian agricultural contexts (Hovsepyan and Willcox, 2008; Riehl et al., 2012). It is reasonable to propose that, as selection for seed use occurred, small-seeded *C. microcarpa* populations were transformed into larger-seeded plants that, by virtue of having larger seeds, came to be recognized as *C. sativa*. Taken together, these lines of evidence suggest a domestication scenario whereby large-fruited *C. sativa* emerged as a crop species from populations of *C. microcarpa* that were adapted to agricultural habitats, where they were most likely present initially as weeds and only later came to be actively cultivated as oilseed plants.

Weedy origins for cultivated *C. sativa* are also consistent with the well-documented status of some *C. sativa* strains as classic examples of "Vavilovian mimicry," i.e., cryptic mimicry of a crop species by an unrelated agricultural weed (Barrett, 1983). In the case of *C. sativa*, a morphologically distinct weedy form (sometimes given the taxonomic designation *C. sativa* var. *linicola*) has evolved a suite of adaptations that allow it to persist undetected in fields of flax (*Linum usitatissimum* L.) (Barrett, 1983). The flax-mimic grows to the same height as flax, it reproduces synchronously, and its indehiscent fruits are harvested together with flaxseed. Perhaps most dramatically, its seeds have evolved an increased surface area that mimics flax's flat, thin seeds and enables it to escape winnowing after harvest (Harper et al., 1970; Barrett, 1983).

Given the longstanding presence of *C. sativa* in agricultural fields of Europe and Asia minor, it is striking how few plants of this species we encountered – either in cultivation or as a weed – over three successive years of collecting in 2012, 2013 and 2014. As noted above, extensive collection efforts yielded only a single *C. sativa* individual in an atypical habitat. A subsequent collecting trip in the summer of 2017 has confirmed this absence of *C. sativa* in areas of Eurasia where, based on herbarium collections, it was once common but has declined in occurrence in recent decades (JRB, personal observations). It seems likely that this dramatic decline is attributable to a combination of the widespread adoption of rapeseed over *C. sativa* for oilseed production (European Commission, 2017) and the simultaneous increase in widespread herbicide use for weed control. Given the paucity of germplasm available to breeders today, it would be to the benefit of the research community to make additional concerted efforts to collect and conserve remaining *C. sativa* varieties. Our sampling for the present study focused predominately on the region of highest species diversity for the genus, which would not necessarily have to correspond to the area in which *C. sativa* was domesticated. As such, we believe it could be valuable to expand collection efforts more broadly across Europe and elsewhere to acquire genetic material for future analyses.

Beyond *C. sativa*, the potential for introgression into the domesticate from wild *Camelina* species (e.g., Séguin-Swartz, 2013) may prove an additional valuable strategy for breeders who are looking for new traits or to increase genetic diversity. However, before the genetic potential

of wild *Camelina* species may be tapped, some questions must be answered: What is the origin of *C. sativa* subgenomes? Which agronomically valuable traits exist in wild *Camelina* species (e.g., flowering time, drought tolerance, favorable seed oil composition) that could offer improvements to the crop? Moving forward, whole genome sequences of the diploid *Camelina* species will allow for more accurate assembly of future *C. sativa* genomes, while also providing more robust data to elucidate the origin and evolutionary history of *C. sativa*.

5. Conclusions

With the recently published *C. sativa* reference genome (Kagale et al., 2014), the promise for the improvement of this oil seed crop is higher than ever. The phylogenetic insights of the present study provide a step towards elucidating the potential for other species in the genus to contribute valuable diversity to the *C. sativa* crop. Furthermore, our study indicates that *Camelina*, a genus with a muddled genomic ancestry, may prove to be a useful model for the evolution and genetics of polyploid plants. Previously studied as a model for Vavilovian mimicry, *C. sativa* may prove a tractable system for studying questions of crop mimicry and de-domestication. As high throughput sequencing technologies become more inexpensive, the possibility of comparing whole genomes of *Camelina* species is within reach, allowing for high confidence measures of introgression and selection.

Acknowledgements

We thank Dr. Ihsan Al-Shehbaz for assistance in confirming species determinations, Dr. Zübeyde Uğurlu for helping as a local guide in Turkey, Dr. Evan Forsythe for assistance with field work in 2013, and Dr. George Nakhutsrishvili and George Tedoradze of the Botanical Institute in Georgia. Dr. George Fayvush and Dr. Alla Aleksanyan of the Armenian Institute of Botany for being local guides and lending botanical expertise of the Caucasus. We thank Dr. David Galbraith for his expertise and mentorship in flow cytometry, Michael McConaughy for assisting with the plants, Dr. Mauricio Bonifacino for his mentorship in GIS mapping, Dr. Bruno Vilela for assistance in R, and Mike Dyer and the Washington University greenhouse staff for keeping a watchful eye on the plants. This work is supported by funding from the State of Arizona through a grant from the Water Environmental and Energy Solutions program to MAB and a Washington University in St. Louis International Center for Advanced Renewable Energy and Sustainability grant to JRB and KMO. A National Geographic Society Committee for Research and Exploration Grant (#9292-13) to MAB supported field collections to Armenia, Georgia, and Turkey in 2013 and collection of RADseq data. A University of Arizona undergraduate research abroad grant (BRAVO!) in part funded by the Howard Hughes Medical Institute (HHMI 52006942) supported the 2014 field collections of JRB in central Turkey. Additional student funding comes from the Undergraduate Biology Research Program (UBRP) through HHMI 52006942 and the Donald Danforth Plant Science Fellowship.

Author contributions

J.R.B and M.A.B. planned and designed the research. J.R.B, A.A.D., and M.A.B. conducted field work. J.R.B. performed all experiments. J.R.B. and K.M.O. analyzed the data and wrote the manuscript.

Declarations of interest

None.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version, at https://doi.org/10.1016/j.ympev.2018.06.031.

References

Al-Shehbaz, I.A., Beilstein, M.A., 2010. Flora of North America North of Mexico. New York and Oxford 7, 451–453.
Barrett, S.H., 1983. Crop mimicry in weeds. Econ. Bot. 37 (3), 255–282.
Bennett, M.D., Leitch, I.J., Price, H.J., Johnston, J.S., 2003. Comparisons with *Caenorhabditis* (~100 Mb) and *Drosophila* (~175 Mb) using flow cytometry show genome size in *Arabidopsis* to be ~157 Mb and thus ~25% larger than the Arabidopsis genome initiative estimate of ~125 Mb. Ann. Bot. 91 (5), 547–557.
Beilstein, M.A., Al-Shehbaz, I.A., Kellogg, E.A., 2006. Brassicaceae phylogeny and trichome evolution. Am. J. Bot. 93, 607–619.
Beilstein, M.A., Al-Shehbaz, I.A., Matthews, S., Kellogg, E.A., 2008. Brassicaceae phylogeny inferred from phytochrome A and ndhF sequence data: tribes and trichomes revisited. Am. J. Bot. 95, 1307–1327.
Bouby, L., 1998. Two early finds of gold-of-pleasure (Camelina sp.) in middle Neolithic and Chalcolithic sites in western France. Antiquity 72, 391–398.
Davis, P., 1988. Flora of Turkey and the East Aegean Islands. Edinburgh University Press, Edinburgh, United Kingdom, pp. 490–494.
Dönmez, E.O., Belli, O., 2007. Urartian plant cultivation at Yoncatepe (Van), eastern Turkey. Econ. Bot. 61, 290–298.
Eaton, D.A.R., Ree, R.H., 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). Syst. Biol. 62 (5), 689–706.
Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. 32 (5), 1792–1797.
European Commission, 2017. Oilseeds and Protein Crops Market Situation. Report by the Committee for the Common Organisation of Agricultural Markets (Agriculture and Rural Development), 25 October 2017. < https://ec.europa.eu/agriculture/sites/agriculture/files/cereals/presentations/cereals-oilseeds/market-situation-oilseeds_en.pdf > .
Flagel, L.E., Wendel, J.F., 2009. Gene duplication and evolutionary novelty in plants. New Phytol. 183, 557–564.
Francis, A., Warwick, S.I., 2009. The Biology of Canadian Weeds. 142. *Camelina alyssum* (Mill.) Thell.; *C. microcarpa* Andrz. ex DC.; *C. sativa* (L.) Crantz. Can. J. Plant Sci. 89, 791–810.
Galasso, I., Manca, A., Braglia, L., Ponzoni, E., Breviario, D., 2015. Genomic fingerprinting of *Camelina* species using cTBP as molecular marker. Am. J. Plant Sci. 6, 1184–1200.
Galbraith, D.W., Harkins, K.R., Maddox, J.M., Ayres, N.M., Sharma, D.P., Firoozabady, E., 1983. Rapid flow cytometric analysis of the cell cycle in intact plant tissues. Science 220, 1049–1051.
Gehringer, A., Friedt, W., Lühs, W., Snowdon, R.J., 2006. Genetic mapping of agronomic traits in false flax (*Camelina sativa* subsp. sativa). Genome 49, 1555–1563.
Harper, J.L., Lovell, P.H., Moore, K.G., 1970. The shapes and sizes of seeds. Annu. Rev. Ecol. Syst. 1, 327–356.
Hovsepyan, R., Willcox, G., 2008. The earliest finds of cultivated plants in Armenia: evidence from charred remains and crop processing residues in pisé from the Neolithic settlements of Aratashen and Aknashen. Veget. History Archaeobot. 17 (Suppl. 1), S63–S71.
Hutcheon, C., Ditt, R.F., Beilstein, M., Comai, L., Schroeder, J., Goldstein, E., Shewmaker, C.K., Nguyen, T., De Rocher, J., Kiser, J., 2010. Polyploid genome of *Camelina sativa* revealed by isolation of fatty acid synthesis genes. BMC Plant Biol. 10 (1), 233.
Iskandarov, U., Kim, H.J., Cahoon, E.B., 2014. *Camelina*: an emerging oilseed platform for advanced biofuels and bio-based materials. In: McCann, M.C., Buckeridge, M.S., Carpita, N.C. (Eds.), Plants and Bioenergy. Springer, Berlin, pp. 131–140.
Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405.
Kagale, S., Koh, C., Nixon, J., Bollina, V., Clarke, W.E., Tuteja, R., Spillane, C., Robinson, S.J., Links, M.G., Clarke, C., et al., 2014. The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. Nat. Commun. 5, 3706.
Kang, J., Snapp, A.R., Lu, C., 2011. Identification of three genes encoding microsomal oleate desaturases (FAD2) from the oilseed crop *Camelina sativa*. Plant Physiol. Biochem. 49, 223–229.
Koch, M.A., Kiefer, M., German, D.A., Al-Shehbaz, I.A., Franzke, A., Mummenhoff, K., Schmickl, R., 2012. BrassiBase: tools and biological resources to study characters and traits in the Brassicaceae version 1.1. TAXON 61 (5), 1001–1009.
Kiefer, M., Schmickl, R., German, D.A., Mandáková, T., Lysak, M.A., Al-Shehbaz, I.A., Franzke, A., Mummenhoff, K., Stamatakis, A., Koch, M.A., 2014. BrassiBase: introduction to a novel knowledge database on brassicaceae evolution. Plant Cell Physiol. 55 (1), 1–9.
Kroll, H., 2000. Agriculture and arboriculture in mainland greece at the beginning of the first millenium B.C. Pallas 52, 61–68.
Kumar, A., Bennetzen, J.L., 1999. Plant retrotransposons. Annu. Rev. Genet. 33, 479–532.
Levin, D.A., 1983. Polyploidy and novelty in flowering plants. Am. Natural. 122 (1), 1–25.
Lu, C., Kang, J., 2008. Generation of transgenic plants of a potential oilseed crop *Camelina sativa* by Agrobacterium - mediated transformation. Plant Cell Rep. 27, 273–278.
Maassoumi, A., 1980. Cruciferes de la flore d'Iran: etrude caryosystematique. Strasbourg, France.
Maddison, W.P., Maddison, D.R., 2015. Mesquite: A Modular System for Evolutionary Analysis. Version 3.04. < http://mesquiteproject.org > .
Miller, N.F., 1991. The near east. In: Van Zeist, W., Wasylikowa, K., Behre, K.E. (Eds.), Progress in Old World Palaeoethnobotany. Balkema, Rotterdam, Netherlands, pp. 133–160.
Moser, B.R., 2010. Camelina (*Camelina sativa* L.) oil as a biofuels feedstock: golden opportunity or false hope? Lipid Technol. 22 (12), 270–273.

Moser, B.R., 2012. Biodiesel from alternative oilseed feedstocks: camelina and field pennycress. Biofuels 3 (2), 193–209.
Nesbitt, M., 1996. Chalcolithic crops from Kuruçay Höyük: an interim report. In: Duru, P. (Ed.), Results of the excavations 1978-1988, the late Chalcolithic and early Bronze settlements. Türk Tarih Kurumu Basimevi, Ankara, pp. 90 134–7.
Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7 (5).
QGIS Development Team, 2016. QGIS 2.14.3 Geographic Information System. Open Source Geospatial Foundation Project. < http://qgis.osgeo.org > .
Riehl, S., Benz, M., Conard, N.J., Darabi, H., Deckers, K., Nashli, H.F., Zeidi-Kulehparcheh, M., 2012. Plant use in three Pre-Pottery Neolithic sites of the northern and eastern Fertile Crescent: a preliminary report. Veget. History Archaeobot. 21 (2), 95–106.
Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., Sánchez-Gracia, A., 2017. DnaSP v6: DNA sequence polymorphism analysis of large datasets. Mol. Biol. Evol. 34, 3299–3302.
Ruiz-Lopez, N., Haslam, R.P., Napier, J.A., Sayanova, O., 2014. Successful high-level accumulation of fish oil omega-3 long-chain polyunsaturated fatty acids in a transgenic oilseed crop. Plant J. 77 (2), 198–208.
Séguin-Swartz, G., Eynck, C., Gugel, R.K., Strelkov, S.E., Olivier, C.Y., Li, J.L., Klein-Gebbinck, H., Borhan, H., Caldwell, C.D., Falk, K.C., 2009. Diseases of *Camelina sativa* (false flax). Can. J. Plant Pathol. 31 (4), 375–386.
Séguin-Swartz, G., Nettleton, J.A., Sauder, C., Warwick, S.I., Gugel, R.K., 2013. Hybridization between *Camelina sativa* (L.) Crantz (false flax) and North American Camelina species. Plant Breed. 132 (4), 390–396.
Shonnard, D.R., Williams, L., Kalnes, T.M., 2010. Camelina-derived jet fuel and diesel: sustainable advanced biofuels. Environ. Prog. Sustain. Energy 29 (3), 382–392.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30 (9), 1312–1313.
te Beest, M., Le Roux, J.J., Richardson, D.M., Brysting, A.K., Suda, J., Kubesova, M., Pysek, P., 2012. The more the better? The role of polyploidy in facilitating plant invasions. Ann. Bot. 109 (1), 19–45.
Tsuchimatsu, T., Kaiser, P., Yew, C.L., Bachelier, J.B., Shimizu, K.K., 2012. Recent loss of self-incompatibility by degradation of the male component in allotetraploid *Arabidopsis kamchatica*. PLOS Genet. 8 (7).
Van Zeist, W.A., 1981. Plant remains from Iron Age Noordbarge, province of Drenthe, the Netherlands. Palaeohistoria 23, 169–193.
Vavilov, N.I., 1987. Origin and Geography of Cultivated Plants. Cambridge University Press, Cambridge, United Kingdom.
Vollmann, J., Grausgruber, H., Stift, G., Dryzhyruk, C., Lelley, T., 2005. Genetic diversity in camelina germplasm as revealed by seed quality characteristics and RAPD polymorphism. Plant Breed. 124 (5), 446–453.
Webb, D.M., Knapp, S.J., 1990. Protocol DNA extraction from a previously recalcitrant plant genus. Plant Mol. Biol. Rep. 8 (3), 180–185.
White, T.J., Bruns, T., Lee, S.B., Taylor, J.W., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. (Eds.), PCR Protocols: A Guide to Methods and Applications. Academic Press, San Diego, California, USA, pp. 315–322.
Zohary, D., Hopf, M., Weiss, E., 2012. Domestication of Plants in the Old World, fourth ed. Oxford University Press, Oxford, United Kingdom.