Molecular Phylogenetics and Evolution xxx (xxxx) xxx–xxx

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

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