Molecular Phylogenetics and Evolution 66 (2013) 617–627

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

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

Phylogeny of *Crocus* (Iridaceae) based on one chloroplast and two nuclear loci: Ancient hybridization and chromosome number evolution

Dörte Harpke [a,*], Shuchun Meng [a,b], Twan Rutten [a], Helmut Kerndorff [c], Frank R. Blattner [a]

[a] Leibniz Institute of Plant Genetics and Crop Research (IPK), D-06466 Gatersleben, Germany
[b] Vegetable Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China
[c] Casa da Eira, São Romão, Cx 548 A, 8150-058 São Bras de Alportel, Portugal

ARTICLE INFO

Article history:
Received 18 February 2012
Revised 30 August 2012
Accepted 8 October 2012
Available online 30 October 2012

Keywords:
B chromosome
Chromosome number
Crocus
Evolution
Internal transcribed spacer (ITS)
pCOSAt103
Phylogeny
*trn*l-F
Seed testa
Whole genome duplication

ABSTRACT

Crocus consists of about 100 species distributed from western Europe and northern Africa to western China, with the center of diversity on the Balkan Peninsula and in Asia Minor. Our study focuses on clarifying phylogenetic relationships and chromosome number evolution within the genus using sequences of the chloroplast *trn*L-F region, the nuclear ribosomal DNA internal transcribed spacer (ITS) region, and a part of the nuclear single-copy gene *pCOSAt103*.

In a combined dataset of ITS and *trn*L-F sequences, 115 individuals representing 110 taxa from both subgenera and all sections and series of *Crocus* were analyzed with Bayesian phylogenetic inference. For *pCOSAt103* 79 individuals representing 74 *Crocus* taxa were included, and for the majority of them PCR amplicons were cloned and up to eight clones per individual were sequenced to detect allopolyploidization events. *Romulea* species were included as outgroup in both analyses. Characteristics of seed surface structures were evaluated by scanning electron microscopy.

Phylogenetic analysis of ITS/*trn*L-F data resulted in a monophyletic genus *Crocus*, probably monophyletic sections *Crocus* and *Nudiscapus*, and inferred monophyly for eight of the 15 series of the genus. The *C. biflorus* aggregate, thought to be consisting of closely related subspecies, was found to be polyphyletic, the taxa occurring within three major clades in the phylogenetic tree. Cloning of *pCOSAt103* resulted in the detection of homoeologous copies in about one third of the taxa of section *Nudiscapus*, indicating an allotetraploid origin of this section. Reconstruction of chromosome number evolution along the phylogenetic tree using a probabilistic and a parsimony approach arrived at partly contradictory results. Both analyses agreed however on the occurrence of multiple polyploidization and dysploidy events. B chromosomes evolved at least five times independently within the genus, preferentially in clades characterized by karyotype changes.

1. Introduction

The genus *Crocus* consists currently of about 100 recognized species (Petersen et al., 2008; Kerndorff and Pasche, 2011; Kerndorff et al., 2011, 2012; Peruzzi and Carta, 2011) occurring from western Europe and northwestern Africa to western China (Supplemental Fig. S1) with the centre of species diversity in Asia Minor and on the Balkan Peninsula. Many crocuses are known as popular ornamentals. Saffron, the dried styles of *C. sativus*, is one of the world's most expensive spices by weight. The styles furthermore contain carotenoids inhibiting cancer cell proliferation (Abdullaev and Espinosa-Aguirre, 2004). This effect was also found in some other *Crocus* species showing that they are of potential pharmaceutical interest (Chryssanthi et al., 2007).

Over the last 200 years a multitude of classifications of the genus were proposed, based on differential weighting of a number of morphological characteristics. Haworth (1800) recognized two sections using the presence (section *Pilgrim*) or absence (section *Depilati*) of hairs in the throat of the flower, a feature that by now is known to vary even within species. Sabine (1829) presented a revision of the genus based on the absence or presence of a prophyll (basal spathe) and corm tunic characteristics, with a further subdivision due to the presence of a single or double proper spathe (bract and bracteole). Herbert (1847) adopted for his primary grouping the presence or absence of a basal spathe, but subdivided his main groups into sections by the character of the corm tunic. Baker (1873) classified the genus by the subdivisions of the stigmata. Maw (1886) preferred Herbert's system of classification and divided the genus into two divisions: *Involucrati* (two sections, further separated by spring and autumn-flowering habit), and *Nudiflori* (four sections, further separated by spring and autumn-flowering habit). The most recent classification of Mathew (1982)

* Corresponding author.
E-mail address: harpke@ipk-gatersleben.de (D. Harpke).

1055-7903/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2012.10.007

is mainly based on Maw's but without considering flowering time. He divided the genus into two subgenera, the monotypic subgenus *Crociris* consisting only of *C. banaticus* (=*C. iridiflorus* Heuff. ex Reichb.) and subgenus *Crocus*. The latter he divided into two sections (section *Crocus* and section *Nudiscapus*) based on the presence or absence of the prophyll, and used the division of the style, corm tunic features and flowering time to define 15 different series (section *Crocus*: six series; section *Nudiscapus*: nine series). In contrast to all earlier authors he also introduced a subspecies concept for many species in different series.

Few attempts were made up to now to clarify phylogenetic relationships within single series (ser. *Crocus*: Grilli Caiola et al., 2004) or the whole genus (Frello et al., 2004; Petersen et al., 2008) with molecular methods. Frello et al. (2004) studied the genomic organization of a tandemly repeated DNA sequence family in 55 representative *Crocus* species. Their results did not support Mathew's classification what led them to discuss the possibility of far-reaching hybridization and fast speciation within the genus. In the first phylogenetic study comprising the whole genus Petersen et al. (2008) sequenced five chloroplast regions in 131 individuals representing most species and subspecies described at that time. The analysis supported seven of the series as described by Mathew (1982) but revealed the subgenera and the remaining series as non-monophyletic. Particularly the members of series *Biflori* and *Reticulati*, comprising about one third of the taxa within the genus, grouped in different clades within the phylogenetic tree. As the study was solely based on plastid marker regions it could not detect potential chloroplast capture or incomplete lineage sorting, mechanisms which are not rare and might result in inference of wrong species relationships even at deep nodes of phylogenetic trees (e.g., Bänfer et al., 2006; Jakob and Blattner, 2006; Kiefer et al., 2009). Chloroplast sequences were also used to evaluate the possibilities of DNA barcoding for taxon recognition in *Crocus*, although not all species could be safely distinguished by this method (Seberg and Petersen, 2009).

Crocus is an extreme example for varying chromosome numbers ranging from $2n = 6$ to $2n = 70$ with up to eleven supernumerary B chromosomes (Brighton et al., 1973; Goldblatt and Takei, 1997; Özhatay, 2002). Even within the same species chromosome numbers might vary (Brighton et al., 1973). The evolution of the karyotypes seems to be characterized by extensive dysploidy combined with polyploidization events (Rudall et al., 1984; Goldblatt and Takei, 1997). Through comparison with the closely related genus *Syringodea* ($x = 6$) and on the basis of *Crocus* species defined as primitive (Mathew, 1982), Goldblatt and Takei (1997) suggested a basic chromosome number of $x = 6$ for the genus. Chromosome numbers within *Crocus* might be furthermore a result of a reduction to $x = 4$ by descending dysploidy in several lineages and polyploidization events based on the basic numbers of $x = 3$–6 (Goldblatt and Takei, 1997).

In order to further clarify species relationships within the genus and analyze chromosome number evolution in *Crocus* in a phylogenetic framework, we sequenced the internal transcribed spacer region (ITS) of the nuclear ribosomal DNA (rDNA) and the chloroplast *trn*L intron plus *trn*L-F intergenic spacer (*trn*L-F). Both regions have proven to be useful phylogenetic markers in a multitude of studies (Baldwin et al., 1995; Pleines et al., 2009). However, chloroplasts are inherited mostly maternally in angiosperms, and the rDNA is a multigene family consisting of hundreds of tandem-repeated units organized in single or multiple clusters in the genome (NOR), resulting in concerted evolution of the repeats through processes like unequal crossing over (Eickbush and Eickbush, 2007) and gene conversion (Li, 1997). As these processes can also result in effectively uniparental inheritance of ITS, thus masking traces of hybridization (Álvarez and Wendel, 2003), we sequenced as a third marker a part of the single-copy nuclear Conserved Ortholog Set (COS) gene *pCOSAt103* (a Mg-protoporphyrin IX monomethyl ester cyclase, AGI-ID: AT3G56940; see Li et al., 2008). These sequences were primarily used to infer hybridization events and determine parental taxa of allopolyploids with higher reliability in comparison to ITS.

2. Materials and methods

2.1. Plant materials

In our ITS analysis we included 142 individuals representing 125 taxa (88 species, 37 subspecies according to Mathew's classification) and all taxonomic groups of *Crocus* (Supplemental Table S1). A subset of 115 individuals (110 taxa) was used in the analysis of *trn*L-F, and 79 individuals (74 taxa) for *pCOSAt103*. In addition, we included three species of the genus *Romulea* to function as outgroup taxa, as *Romulea* together with *Syringodea* are closest relatives of *Crocus* (Goldblatt et al., 2006, 2008; Petersen et al., 2008). The seed testa structure of 113 individuals representing all taxonomic groups was examined using a scanning electron microscope. Here we present only the major types of surface structures. The entire dataset of testa structure analyses will be published in detail elsewhere. Voucher specimens of all analyzed individuals were deposited at the herbaria Athens (ATH), Belgrad (BEOU), DNA-Bank/Berlin Dahlem (B), Chania (MAIC), Gatersleben (GAT), Amman (IABH), or Teheran (TARI). *Crocus biflorus* subspecies were already treated here as species according to a new taxonomic concept reflecting phylogenetic relationships in this group (Kerndorff et al., in preparation).

2.2. Molecular methods

Genomic DNA was extracted from about 10 mg of silica-dried leaf material with the DNeasy Plant DNA Extraction Kit (Qiagen) according to the protocol of the manufacturer. DNA concentration and quality were afterwards checked on 0.8% agarose gels.

The ITS region (ITS1, 5.8S rDNA, ITS2) was amplified using the primers ITS-A and ITS-B (Blattner, 1999). PCR was performed with 1.5 U Taq DNA Polymerase (Qiagen) in the supplied reaction buffer, 0.2 μM of each dNTP, 50 pmol of each primer, Q-Solution (Qiagen) with a final concentration of 20%, and about 20 ng of total DNA in 50 μl reaction volume in a GeneAmp PCR System 9700 (Perkin–Elmer). Amplification was performed with 3 min initial denaturation at 95 °C and 35 cycles of 30 s at 95 °C, 45 s at 56 °C and 30 s at 70 °C, followed by a final extension for 8 min at 70 °C. PCR products were purified using NucleoFast 96 PCR plates (Macherey–Nagel) following the manufacturer's protocol, and eluted in 30 μl water. Both strands of the PCR products were directly sequenced with Applied Biosystems BigDye Terminator technology on an ABI 3730xl automatic DNA sequencer using either the primers from PCR amplifications or the nested sequencing primers ITS-SF and ITS-SR (Blattner et al., 2001). Forward and reverse sequences from each directly sequenced amplicon were inspected, manually edited where necessary, and combined in single consensus sequences.

To obtain sequence information from the chloroplast genome we sequenced the *trn*L-F region, consisting of the *trn*L(UAA) gene with its intron and the intergenic spacer between the *trn*L and *trn*F(GAA) genes. PCR amplification and sequencing followed Jakob and Blattner (2006). Forward and reverse sequences from each amplicon were inspected, manually edited where necessary, and combined in single consensus sequences.

As *Crocus* was hypothesized to probably consist of many hybrid taxa (Frello et al., 2004), we sequenced *pCOSAt103* as an additional single-copy nuclear marker for representatives of all major clades resulting from the ITS/*trn*L-F analysis. This marker was amplified