The origin of saffron:

Progenitors, areas and transcriptomics of economic traits

Dissertation

zur Erlangung des

Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

der

Naturwissenschaftlichen Fakultät I -Biowissenschaften-

der Martin-Luther-Universität

Halle-Wittenberg

Vorgelegt

von Frau Zahra Nemati

geb. am 23.05.1983 in Shiraz, Iran

verteidigt am 30.10.2018

Halle (Saale), Deutschland

Gutachter:
1. Prof. Dr. M. Röser
2. Dr. F. Blattner
3. Prof. Dr. T. Schmidt

Contents

1. General introduction

Saffron, Crocus sativus L.

Saffron, *Crocus sativus*, is the economically most important species within *Crocus* and, with up to 18,000 € per kilogram, the most expensive spice by weight (Rios et al. 1996; Ferrence and Bendersky 2004). The spice saffron derived from the dried stigmas of *C. sativus* with distinct colour, flavour and aroma. For producing 1 kg of saffron, 110,000 to 165,000 flowers are required (Gracia et al. 2009). However, the direct labour required for saffron's cultivation, harvesting and handling makes this spice expensive (Aytekin and Acikgoz 2008).

The use of *C. sativus* dates back to 2500 - 1500 BC according to historical and archaeological evidences (Negbi 1989; Ferrence and Bendersky 2004). The term saffron likely derived from Arabic *zafran* that means yellow and today saffron is the name for *C. sativus* as well as the spice obtained from its dried stigmas. Saffron was scientifically named by Linnaeus in 1762 as *Crocus sativus* var. *officinalis* (Grilli-Caiola and Canini 2010). Today, the greatest saffron producing countries are Iran, Spain and India (Fig. 1.1).

In recent years, in addition to the species' use as spice also therapeutic properties have been documented in saffron and also some other *Crocus* species seem to be of potential pharmaceutical interest. The amount of saffron used is really important when low doses of

saffron work as a stimulant but high doses of saffron might be toxic and dangerous (Winterhalter and Straubinger 2000).

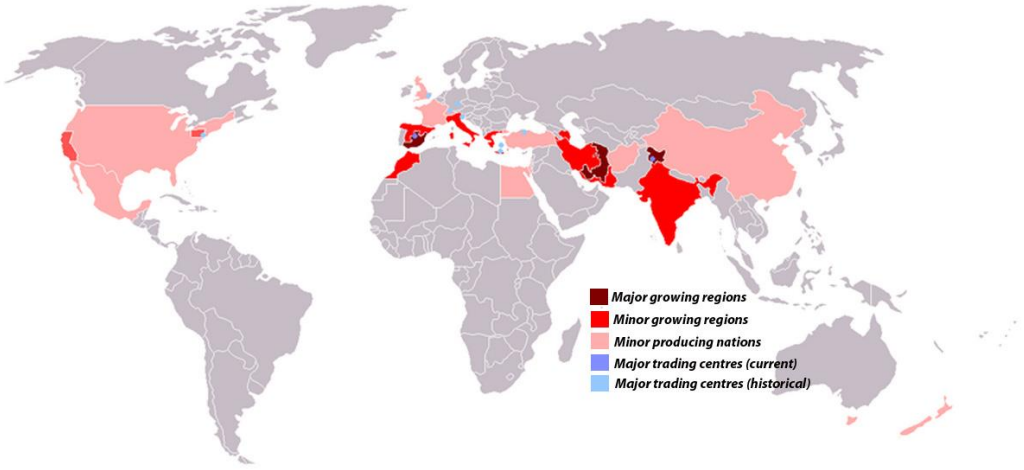


Figure 1.1. Saffron producing regions, producing nations (major and minor) and trading centres (present and past) (Taken from: https://en.wikipedia.org/wiki/Trade_and_use_of_saffron).

Saffron has been applied since long time against asthma, bronchospasm, insomnia, uterine bleeding, Alzheimer's disease, infertility, carminative, cardiovascular diseases, depression, cholesterol, diuretic, febrifuge, stimulant, stomach ailments, retina-degeneration, multiple sclerosis and different kinds of cancers (Abdullaev et al. 2004; Giaccio 2004; Chryssanthi et al. 2007; Dagostino et al. 2007; Fernandez 2007; De-Juan et al. 2009; Gresta et al. 2009; Poma et al. 2012; Makri et al. 2013; Siracusa et al. 2013). However, none of these treatments were tested in large clinical studies so far.

Saffron is an autumn flowering plant, growing from an underground globular shape corm. Depending on the size of mother corm, each mother corm produces 2-3 principal daughter corms from apical buds and several corms from lateral buds (Mathew 1982). Corms have the tunic cover of parallel fibres and are well-adapted to tolerate cold winters and hot summers (Agayev and Zarifi 2010). Leaves (from 6 to 9) are narrow, grass-like, dark green coloured and 30-50 cm long. Each shoot produces usually one or several, but even as many as 12 flowers, which are normally fragrant and purple in colour. Flowers have an underground ovary and a style of 9-10 cm length divided into three dark red branches stigmas, each one up to 30-40 mm long. Perianth consists of 6 petaloid tepals and three yellow anthers. (Fig. 1.2).

Figure 1.2. Different parts of the saffron plant (taken from Madan et al. 1996).

Crocus sativus is a sterile triploid that is propagated only vegetatively by means of corms (Brighton 1977; Mathew 1977). This vegetative cultivation offers advantages in maintaining the genetic characteristics of the plants, but impedes any breeding advances by genetic improvement through crossing of different lineages. In *C. sativus*, therefore many experiments have been performed in order to obtain seeds (Grilli-Caiola 1999) and in vitro-regenerated plants (Souret and Weathers 2000), although without success. An alternative approach would be to investigate the closest relatives of *C. sativus*, trying to elucidate how the triploid species has originated. Still, the wild progenitors and the mode of saffron evolution is a subject of speculation. *Crocus sativus* is thought to be an allopolyploid hybrid, i.e. the result of a cross between diploid parental species of which one contributed an unreduced gamete (Brighton 1977; Agayev 2002). An allotriploid origin was hypothesized by Tsaftaris et al. (2011) based on sequences of MADS-box genes with *C. cartwrightianus* and *C. thomasii* as putative progenitor species. The contributions of *C. cartwrightianus* to the formation of saffron were proposed in different studies (Grilli-Caiola et al. 2004; Harpke et al. 2013; Larsen et al. 2015). Apart of *C. cartwrightianus*, *C. thomasii* and *C. pallasii* also *C. mathewii* from Turkey, and *C. hausknechtii*, *C. michelsonii* and *C. almehensis* from Iran were mentioned as possible parents (Frello and Heslop-Harrison 2000; Grilli-Caiola et al. 2001, 2004; Alavi-Kia et al. 2008; Tsaftaris et al. 2011; Harpke et al. 2013; Alsayied et al. 2015; Larsen et al. 2015).

The genus Crocus

The monocot genus *Crocus* L. (Iridaceae) consists of perennial geophytes with several species grown as ornamentals and there is a large community of gardeners exploring the potential of species and/or hybrids as garden flowers. *Crocus sativus* though is the most important species in the genus due to the economic and the medicinal importance. *Crocus* is distributed from Western Europe and northern Africa, through Near East, Central Asia to western China with the centre of diversity on the Balkan Peninsula and in Asia Minor (Mathew 1982).

Since 200 years, various classifications of the genus *Crocus* have been proposed based on different morphological features. The most recent classification of the genus *Crocus* was done by Mathew (1982). He described two subgenera for *Crocus* consisting of subgenus *Crociris* (including only *C. banaticus*) and subgenus *Crocus* (with the two sections *Crocus* and *Nudiscapus*). He further defined six different series in the section *Crocus* and nine series in the section *Nudiscapus* based on the division of the style, corm tunic characters and flowering time. In contrast to all earlier authors he also introduced a subspecies concept for many species in different series. However, recent phylogenetic studies clearly show that the subspecies, at least in *C. biflorus*, are no valid units (Harpke et al. 2013; Schneider et al. 2013). Phylogenetic studies (Petersen et al. 2008; Harpke et al. 2013) furthermore revealed incongruence of Mathew's classification also on other taxonomic levels, as monophyly of series could only be confirmed for a part of them. For example, *C. baytopiorum* belonging to series *Verni* according to Mathew (1982) clearly groups with series *Crocus*. The genus *Crocus* includes currently over 200 recognized species (Rukšans 2017).

Crocus sativus, according to Mathew's system, belongs to series *Crocus*, which is distributed from Italy in the west to the Caucasus with the centre of diversity on the Balkan Peninsula (Fig. 1.3). The other species of this group are *C. asumaniae*, *C. baytopiorum*, *C. cartwrightianus*, *C. hadriaticus*, *C. macedonicus*, *C. mathewii*, *C. moabiticus*, *C. oreocreticus*, *C. pallasii* (with four subspecies), *C. thomasii* and *Crocus naqabensis*.

The species in series *Crocus* have corms of about 10-25 mm diameter, with fibrous tunics where fibres are parallel in the lower part, reticulate near the apex of corm, covering until neck up to 2 cm long. They are all autumn-flowering and usually with 1-6 flowers from white or light pinkish-violet to deep violet-blue or violet. The style is orange to red, deeply

divided into three long stigmatic branches, each branch 6-10 mm long (Mathew 1982; Grilli-Caiola and Canini 2010).

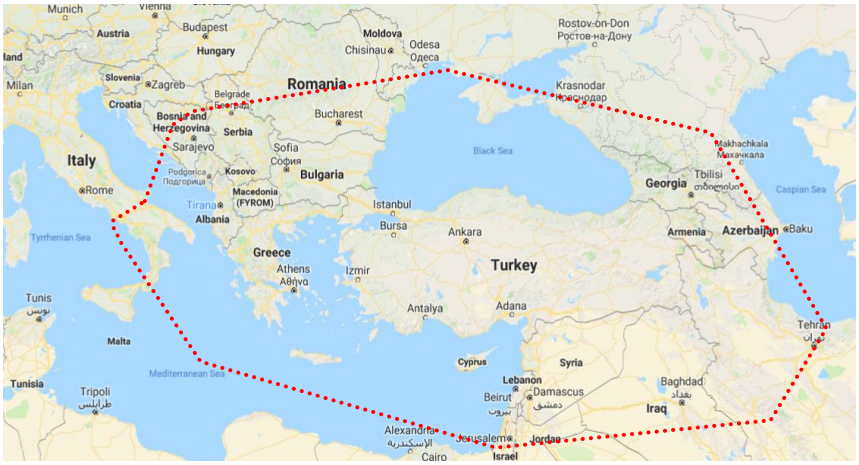


Figure 1.3. Distribution area of series *Crocus*; the dotted line outlines the distribution area.

The species in series *Crocus* are all wild and diploid except for the triploid *C. sativus* (2n = 3x = 24), which is male sterile and only propagated vegetatively (Brighton 1977; Mathew 1977). Generally, the series *Crocus* is known for varying chromosome numbers. Most common within the series are 2n = 16 (*C. cartwrightianus*, *C. hadriaticus*, *C. oreocreticus*, *C. pallasii* subsp. *hausknechtii*, *C. thomasii*) and 2n = 14 (*C. moabiticus*, *C. naqabensis*, *C. pallasii* subsp. *dispathaceus*, *C. pallasii* subsp. *pallasii*), respectively (Brighton et al. 1973; Karamplianis et al. 2013). Although *C. asumaniae* (2n = 26) and *C. mathewii* (2n = 70) have higher chromosome numbers (Brighton et al. 1973) their genome size suggests, that they might be also diploids (Erol et al. 2014). *Crocus pallasii* subsp. *hausknechtii* and *C. pallasii* subsp. *turcicus* are known with 2n = 12 chromosomes (Brighton, 1977; Sanei et al. 2007). The variety in the number of chromosomes within *C. pallasii* might indicate crossing barriers among subspecies and, thus, the probability of the presence of independent species.

Phylogenetic relationships among the taxa of series *Crocus* are still not clarified because either not all taxa were investigated within a single analysis or due to the low variability of the applied markers. Thus, it is also not clear who could be the parental species or at least closest relatives of saffron.

Identification of the position of saffron within the genus Crocus

For all further analyses first the species contributing their genomes to triploid *C. sativus* have to be identified. Up to now in most phylogenetic investigations, hybrids and their parental species were detected due to incongruence between phylogenetic trees of chloroplast markers and the nuclear ribosomal internal transcribed spacer region (ITS: ITS1 + 5.8S rDNA + ITS2) (Peterson et al. 2009, 2011; Tison et al. 2013). However, the biparentally inherited ITS region, as part of the rDNA multigene family, undergoes concerted evolution of the hundreds of tandem-repeated units through processes like unequal crossing over (Eickbush and Eickbush 2007) and gene conversion (Li 1997) that could result in essentially uniparental inheritance. In these cases phylogenetic incongruence does not always result from hybridization. If concerted evolution fails or is still incomplete, analysis of the ITS region provides the possibility to detect hybrids and to identify parental species via cloning and analysing the different ITS alleles separately (e.g., *Rosa*: Wissemann and Ritz 2005; *Gagea*: Peterson et al. 2009, 2011). Furthermore, biparentally inherited single copy nuclear gene regions serve well to detect hybridization events (*Gagea*: Peterson et al. 2011; *Hordeum*: Brassac et al. 2012), as gene conversion works much slower on these genome parts. In these cases single-copy genes of putative diploid progenitor species are analysed together with cloned copies derived from the polyploid under study. In the gene tree the different copies of the polyploid then group together with the copies from the respective diploid parent. They can be firstly uncovered by phylogenetic analyses of a combination of nuclear single-copy genes, ITS and ETS including all species of series *Crocus* together with some outgroup taxa from other groups of the genus. The inclusion of chloroplast markers will allow the determination of the crossing direction, i.e. identify the maternal parent. For the analysis of the phylogeny in series *Crocus*, I here used five nuclear and chloroplast loci including the nuclear ribosomal DNA (nrDNA) internal transcribed spacer (ITS) and external transcribed spacer (ETS), nuclear single-copy marker topoisomerase 6 (TOPO6) subunit B intron 11 and the chloroplast *trn*L-*trn*F and *mat*K-*trn*K regions.

Identification of parental and geographical origin of saffron

Advances in DNA sequencing are quickly reducing the gap between model and non-model organisms (Wheat 2010; Zimmer and Wen 2013). We screened genome-wide single-

nucleotide polymorphism (SNP) data obtained via genotyping-by-sequencing (GBS; Elshire et al., 2011). GBS is a method for sequencing genomic loci flanked by restriction enzyme sites for discovering and genotyping SNPs without any prior sequence information. It allows resolving fine-scale genetic divergence among intraspecific populations separated for less than 20,000 years (Emerson et al. 2010) and should be able to infer species history of recently originated or domesticated crops like saffron. Therefore, a phylogenetic analysis using GBS of the wild parental species from various populations each and *C. sativus* accessions from different countries may contribute in detecting where the origin of saffron was, what species/populations were involved, and how often it originated.

Transcriptome analysis of saffron and its putative parents

The dried stigma of *C. sativus* produces different unique apocarotenoids. Degraded carotenoids (crocin and crocetin) are the chemicals primarily causing colour while the flavour comes from the carotenoid oxidation of safranal and the bitter glucoside picrocrocin. A limited number of *Crocus* species can produce these apocarotenoids. Among them *C. sativus* produces these compounds in significant amounts. Therefore, it is likely that the unique gene combination of the polyploid is responsible for these characters. After identification of the parental species and the area of origin of saffron, we will be able to compare the wild progenitor with the domesticated species in detail. Through sequencing of the messenger RNA (RNA-seq) I want to understand the transcriptional differences between saffron and its parent. To detect the genes involved in differences, we will investigate transcriptome expression in the stigma of saffron as well as its progenitor. This approach helps to find the genes that are active during the time when the styles mature and at the same time provide insights in the expression level of such genes. These results are also important as the expression abundance of specific genes relevant to apocarotenoid biosynthesis could be identified and compared between saffron and its closest relatives.

Objectives

The parental species of probably allotriploid *C. sativus* taxon are still not identified and also the region of its origin remains unclear. To trace the evolution of saffron, I use a nested approach from (*i*) phylogeny of the series to (*ii*) detection of parents or closest relatives to (*iii*) the identification the origin of the crop, and finally (*iv*) of genes involved in important metabolite synthesis like apocarotenoids.

Regarding the phylogeny of the series, we want to answer the following questions: (*i*) What are the phylogenetic relationships among the taxa of series *Crocus*? (*ii*) Do the subspecies described within taxa of series *Crocus* need to be ranked as species or do they still belong to the same taxa? And (*iii*) who are the potential parents or at least closest relatives of saffron, *C. sativus*?

Afterwards, genome-wide single-nucleotide polymorphisms (SNPs) for a representative set of saffron and its closest relatives will be analysed using genotyping-by-sequencing (GBS) to infer (*i*) who the parents of saffron are and (*ii*) where saffron originated.

Finally, we use gene expression differences in the stigmas of saffron and its closest relatives inferred by RNA sequencing (RNA-seq) to understand the genetic changes that characterize saffron in comparison to its closest relatives.

Methods

This section will provide a short introduction to three methods used in this thesis including phylogenetic inference, GBS and transcriptome sequencing and analyses. In the first part, phylogenetic inference using nuclear and chloroplast markers will be described. In the second part, the application of GBS for detecting the location of origin of *C. sativus* will be introduced. Finally, the transcriptome sequencing method used for targeting genes that are most likely responsible for phenotypic changes of *C. sativus* and make saffron different from its parents will be shortly presented.

Phylogeny of series Crocus

The initial part of the study clarified phylogenetic relationships within series *Crocus*. This allowed identifying the closest relatives of saffron. To infer the phylogeny of series *Crocus*, I here used five nuclear and chloroplast loci including the nuclear ribosomal DNA (nrDNA) internal transcribed spacer (ITS) and external transcribed spacer (ETS), nuclear single-copy marker of topoisomerase 6 (TOPO6) subunit B intron 11 and the chloroplast *trn*L-*trn*F and *mat*K-*trn*K regions. I PCR amplified the loci for at least three individuals per species and did Sanger sequencing. In case of nuclear loci, I also cloned PCR products for all diploid taxa, and cloned amplicons of polyploid *C. sativus* and diploids where direct sequencing provides no clear sequences. Afterwards, sequences were quality checked, edited where necessary, and aligned separately for the five loci to produce the alignments.

Maximum parsimony (MP) and the model-based method Bayesian inference (BI) were used for the phylogenetic analysis in this study. Maximum parsimony (Fitch 1971) relies on the idea that the most straightforward interpretation is the best. Therefore, the shortest possible tree with the fewest character changes that explains the present-day sequences is considered best in this method. However, the evolutionary changes might be underestimated as only four different states occur in DNA and reverse or multiple mutations might occur so that it does not necessarily represent the true phylogeny when mutation rates vary largely between taxa. I used PAUP* 4a157 (Swofford 2002) to perform MP on both the separate and concatenated data matrices (if no incongruences were detected among the markers) using the heuristic search algorithm.

Bayesian phylogenetic inference (BI) in contrast to MP is based on nucleotide substitution models to recover a tree that represents the most likely clades regarding both branch lengths and topology. It computes the posterior probability of trees that involves a summation over all possible trees using Markov chain Monte Carlo (MCMC). I performed BI with MRBAYES 3.2.6 (Ronquist et al. 2012). Two Markov chain Monte Carlo (MCMC) analyses with four chains each are run for 4×10^6 generations, sampling a tree every 1000 generations. The initial 25% of trees are discarded as burn-in.

For the chloroplast regions, I used in addition a statistical parsimony network approach (in TCS (Clement et al. 2000) that reflects genealogical relationships of the chloroplast haplotypes. Apart of identifying the maternal parent of saffron, the chloroplast data could provide insights in geographic distribution of maternal genetic diversity (phylogeography).

To perform genotyping-by-sequencing (GBS) genomic DNA of eight to ten individuals of about 15 populations covering the entire distribution area of the parental species of saffron (defined within subproject 1) and a diverse set of saffron cultivars were genotyped according to the two-enzyme method described by Poland et al. (2012). Essentially, we restriction digested with the methylation-sensitive restriction enzyme *Msp*I (CCGG, NEB Inc.) together with the non-sensitive and rare-cutting enzyme *Pst*I-HF (CTGCAG, NEB Inc.) to arrive at restriction sites mostly situated in the under-methylated gene space of the genome. Hence for the GBS library preparation, the genomic DNA is cut with the two restrictions enzymes *Pst*I-HF and *Msp*I and the fragments were then ligated to common and barcode adapters (barcodes consisting of specific 7-bp DNA sequences), which allow the use of specific PCR primers necessary for Illumina sequencing. Furthermore, these unique barcodes allow pooling and splitting up fragments after sequencing. To assure that only fragments with both adapters were amplified the reverse primer is unable to bind unless the complementary sequence is filled in during the first round of forward elongation. After the enrichment, the indexed samples were loaded on an agarose gel for a size selection. Fragments of 150-600 bp were recovered, as smaller fragments of 130 bp or below can contain also primer dimers. Cluster formation and 1 × 100 bp single-end sequencing-by-synthesis were performed on the Illumina HiSeq 2500 platform (Fig. 1.4).

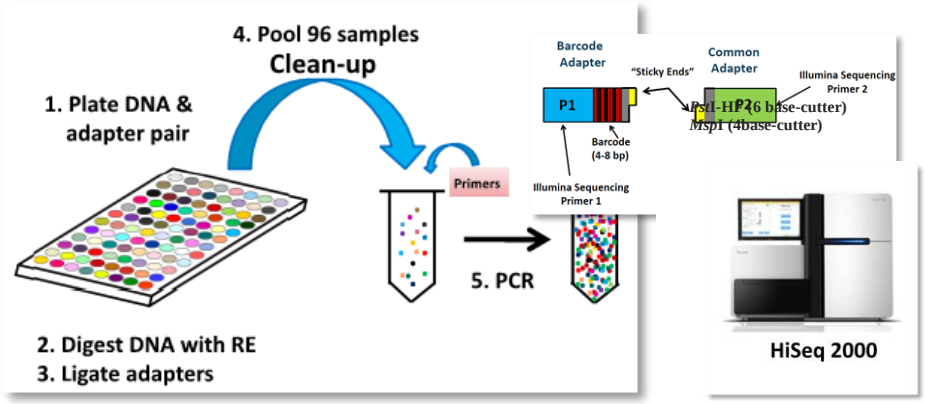


Figure 1.4. GBS library construction. Barcodes sets are restriction enzyme specific (taken from http://cbsu.tc.cornell.edu/lab/doc/GBS_Method_Overview1.pdf).

Generally raw reads obtained from GBS were aligned to a reference genome that is not the case for non-model organisms. In the absence of a reference genome, the assembly of the GBS data was done *de novo*.

I here used the IPYRAD 0.7.5 pipeline (Eaton and Overcast 2016) for *de novo* assembly and clustering of the GBS data. It consists of the seven steps when raw reads were sorted (step 1), quality trimmed (step 2), within-sample clustered (step3), heterozygosity estimated (step 4), consensus sequences (loci), created (step 5), across sample clustered (step 6). At the step 7, all clusters were filtered for a final alignment (Fig. 1.4).

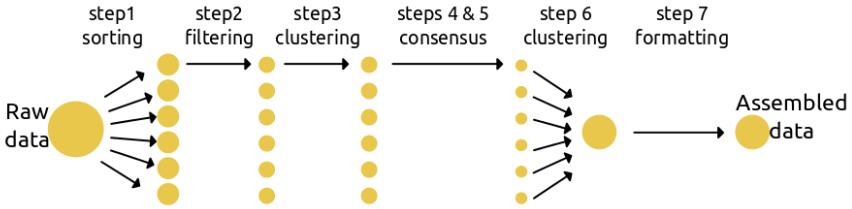


Figure 1.5. Workflow of GBS assembly by IPYRAD (taken from http://ipyrad.readthedocs.io/outline.html).

The dataset generated by IPYRAD was afterwards used as input file for the analysis of population structure and generating a phylogenetic tree. Generally crocuses are obligate outcrossing taxa so that we did not expect to find individuals having today the identical allele constitution as saffron. Therefore, we took into account not only overall similarity (e.g., grouping together in a phylogenetic tree) but looked for the allelic constitution of saffron and searched for parental populations possessing the highest amount of saffron-like alleles. This was done through principal component analysis (PCA) using SNPRelate (Zheng et al. 2012) and ggplot2 (Wickham 2009) R packages and STRUCTURE-like Bayesian assignment tests using R package of LEA (Frichot and Francois 2015). For population structure analysis, each individual is depicted by a vertical bar, mostly partitioned into coloured segments that the length of each segment shows the proportion of the individual's genome in relation to the ancestral populations.

Moreover, we used the SNP data to infer the phylogeny of the three species using maximum parsimony (MP) and Bayesian inference (BI) methods to understand the patterning of their genetic diversity and, for *C. sativus*, the data were also used to see if saffron originated only once (what can be expected based on the low genetic diversity encountered within the species) or if there exist different types of the polyploid.

Transcriptome sequencing and analyses of saffron and its parents

To arrive at an understanding of genes involved in biosynthesis of the unique apocarotenoids of saffron the transcriptomes of saffron and its progenitors were compared. For this step I analysed the transcriptomes of three species, all with multiple carefully staged individuals, with three biological and two technical replications. With this approach I wanted to (*i*) compare style tissue that is always in the same developmental stage to eliminate differences due to disparities in style development, and (*ii*) generalize over several individuals to integrate over individual differences. (*iii*) Harvesting were always done at the same time of the day to eliminate the influence of circadian expression differences.

To perform RNA sequencing, total RNA was extracted from *C. sativus* and its progenitors each with three biological replicates. The poly-A-containing mRNAs were isolated using poly-T oligo-attached magnetic beads. The cDNAs were constructed using the poly-A fragments as templates and then purified and ligated to adapters and PCR amplified. A paired-end 2×100 fragment were then constructed for each sample through PCR and sequenced on Illumina HiSeq 2500 platform.

For transcriptome analysis, I did adapter trimming of reads using CUTADAPT 1.15 (Martin 2011) and reads shorter than 75 bp after adapter removal were discarded. I here used TRINITY for *de no*vo assembly of transcriptome reads (Grabherr et al. 2011). This assembly was used as a transcriptome reference for aligning the reads of *C. sativus* and its progenitor to check for the sequence divergence and mapping rate between the species. BOWTIE2 (Langmead and Salzberg 2012) was applied for mapping reads to the reference transcriptome sequences to know about the mapping rate.

I used the align_and_estimate_abundance.pl script from TRINITY package to align cleaned reads from each libraries of *C. sativus* and its progenitors to the transcriptome generated by Jain et al (2016) using BOWTIE2 (Langmead and Salzberg 2012) and to estimate the transcript abundance using KALLISTO (Bray et al. 2016), I used the align_and_estimate_ abundance.pl script from the TRINITY package. I also applied the PtR script to generate the correlation matrix and Principal Component Analysis (PCA) plot to compare biological replicates across all the samples. Expression analysis using this matrix was executed using the run_DE_analysis.pl script from TRINITY, which uses the edgeR package (Robinson et al. 2010).

2. Phylogeny of the saffron-crocus species group, *Crocus* series *Crocus* (Iridaceae)

The content of this chapter is in press by Nemati et al. (2018) in *Molecular Phylogenetics and Evolution*, doi: https:// doi.org/10.1016/j.ympev.2018.06.036

Introduction

Crocus series *Crocus*, which is distributed from Italy in the West to the Caucasus in the East with the center of diversity on the Balkan Peninsula and Asia Minor (Larsen et al. 2015), consists of 16 autumn-flowering species. Among them is the triploid *C. sativus*, its dried, long three-lobed stigmas are used to produce the spice saffron. The other species of this group are *C. asumaniae*, *C. cartwrightianus*, *C. hadriaticus* (with three subspecies), *C. haussknechtii*, *C. kofudagensi*, *C. macedonicus*, *C. mathewii*, *C. moabiticus*, *C. naqabensis*, *C. oreocreticus*, *C. pallasii* (with four subspecies), and *C. thomasii* (Mathew 1982; Al-Eisawi 2001; Mathew et al. 2009). This group was confirmed to be monophyletic by molecular phylogenetic analyses (Petersen et al. 2008; Harpke et al. 2013), but molecular investigations suggest that also *C. baytopiorum*, that was included in series *Verni* by Mathew (1982), belongs to series *Crocus* (Harpke et al. 2013, 2015). Based on recent investigations in the genus *Crocus*, where it could be shown that several subspecies

represent indeed independent evolutionary lineages (Harpke et al. 2014) and the rather large genetic diversity of the subspecies of *C. pallasii* (Erol et al. 2014; Larsen et al. 2015), Rukšans (2017) ranked these taxa as species (i.e. *C. dispathaceus*, *C. haussknechtii*, *C. turcicus*) instead of *C. pallasii* subspecies.

The high interest in the origin of saffron resulted in several phylogenetic studies carried out within series *Crocus*. Still, relationships among the taxa of series *Crocus* are not yet clarified because either not all known taxa were investigated within single analyses (Grilli-Caiola et al. 2004; Zubor et al. 2004; Tsaftaris et al. 2011; Namayandeh et al. 2013; Erol et al. 2014; Nemati et al. 2014; Alsayied et al. 2015), too low variability of the analyzed sequence regions (Petersen et al. 2008), including only single/very few individuals per species (Petersen et al. 2008; Harpke et al. 2013), or the application of marker methods not appropriate to uncover relationships over distant taxa within the series (RAPD: Grilli-Caiola et al. 2004; AFLP/SSR: Erol et al. 2014; Nemati et al. 2014; Larsen et al. 2015; IRAP: Alsayied et al. 2015). Also the origin of saffron is not yet clear. Most often *C. cartwrightianus* (Grilli-Caiola et al. 2004; Harpke et al. 2013; Larsen et al. 2015), *C. hadriaticus* (Tsaftaris et al. 2011), *C. pallasii* (Harpke et al. 2013; Erol et al. 2014; Alsayied et al. 2015) and/or *C. thomasii* (Tsaftaris et al. 2011) were hypothesized to be its closest wild relative or progenitor.

Generally, the genus *Crocus* is known for variable chromosome numbers. Except for the triploid *C. sativus* (2n = 3x = 24), which is male sterile and only propagated vegetatively (Brighton 1977; Mathew 1977), the species in series *Crocus* are probably all diploid. The most common chromosome numbers within the series are 2n = 16 (*C. cartwrightianus*, *C. hadriaticus*, *C. haussknechtii*, *C. oreocreticus*, *C. thomasii*) and 2n = 14 (*C. dispathaceus*, *C. moabiticus*, *C. naqabensis*, *C. pallasii*), respectively (Brighton et al. 1973; Karamplianis et al. 2013). Although *C. asumaniae* (2n = 26) and *C. mathewii* (2n = 70) have higher chromosome numbers (Brighton et al. 1973) their genome sizes suggest, that they might be still diploids (Erol et al. 2014). *Crocus turcicus* is reported with 2n = 12 chromosomes and *C. haussknechtii* seems to occur also with this chromosome number (Brighton 1977; Sanei et al. 2007).

To obtain a resolved phylogeny we analyzed two chloroplast and three nuclear markers in all taxa of series *Crocus* with the aim to answer the following questions. (*i*) What are the phylogenetic relationships among the taxa of series *Crocus*? (*ii*) Do the *C. pallasii* subspecies, which were recently elevated to species rank, represent independent evolutionary lineages? For saffron and its closest relatives genotyping-by-sequencing

(GBS; Elshire et al. 2011) was carried out to address the question (*iii*) who are the closest relatives and potential progenitors of saffron, *C. sativus*?

Materials and Methods

Taxon sampling

Fifty two individuals, representing 16 species of series *Crocus* and the two outgroup species *C. vernus* and *C.* cf. *tauri* were included in the multi-locus analysis. Based on this analysis, 14 individuals comprising six *C. cartwrightianus*, two *C. hadriaticus*, three *C. oreocreticus* and three *C. sativus* were included in a GBS analysis to better define the position of *C. sativus* among its closest relatives. For seven of them the samples included in the multi-locus analysis could be used, while some different individuals had to be chosen when not enough DNA was available for the GBS analysis. Vouchers of the analyzed taxa were stored in the herbaria GAT, ISTF, MAIC and University of Niš (Table A2.1; Tables and figures marked by A can be found in the Appendix section).

Molecular methods

The extraction of genomic DNA and amplification of the nuclear ribosomal DNA (nrDNA) internal transcribed spacer (ITS) and external transcribed spacer (ETS), and the chloroplast *trn*L-*trn*F and *mat*K-*trn*K regions were carried out according to Harpke et al. (2014). As a nuclear single-copy marker we used the topoisomerase 6 (TOPO6) subunit B intron 11 (Blattner, 2016). All nuclear markers were direct Sanger sequenced. PCR products were cloned and sequenced [six clones per individual; following Harpke et al. (2015)] where direct sequencing revealed polymorphic sequence positions or length differences. All ITS sequences were checked for the presence of pseudogenes according to Harpke and Peterson (2008a, b). All newly obtained sequences were submitted to the EMBL nucleotide database and are accessible through sequence accession numbers LT991589-LT991638, LT991641-LT991747, LS398293-LS398342, LS398365-LS398415.

To obtain genome-wide single-nucleotide polymorphisms (SNPs), genotyping-by-sequencing (GBS) analyses were conducted for 14 individuals comprising *C. cartwright-ianus*, *C. hadriaticus*, *C. oreocreticus* and *C. sativus*. For the library preparation 200 ng of

genomic DNA were used and cut with the two restrictions enzymes *Pst*I-HF (CTGCAG, NEB Inc., Ipswich, UK) and *Msp*I (CCGG, NEB Inc.). Library preparation and individual barcoding followed Wendler et al. (2014). The libraries were sequenced on an Illumina HiSeq 2500 (100 bp single-end reads). Sequence reads for all Illumina runs have been deposited in the European Nucleotide Archive under ERR2570764-77, sample numbers ERS2473163-76 and experiment numbers ERX2570764-78.

Data analyses

Forward and reverse sequences were manually checked, edited where necessary, and assembled in one sequence for each locus and individual. In cases where cloning (ITS, TOPO6–11) revealed the presence of different alleles, cloned sequences were assigned to different haplotypes or consensus sequences were generated for cloned sequences differing by only one substitution. A tetra-nucleotide (AGGG) simple sequence repeat (SSR) within ITS1 was trimmed to the same size in all individuals since it is homoplastic within ser. *Crocus*, and sometimes the number of repeats varies even within individuals. The sequences from single loci were initially aligned using the CLUSTAL algorithm and alignments were adjusted manually where necessary.

Phylogenetic analysis using maximum parsimony (MP) was performed in PAUP* 4a157 (Swofford, 2002) on the separate and, as no major incongruences were detected among the markers, concatenated data matrices (Table A2.2) using the heuristic search algorithm. Bootstrap support values were obtained by 500 bootstrap re-samples using the same settings as before. Models of DNA evolution were evaluated in JMODELTEST 2.1.10 (Darriba et al. 2012) under the Bayesian information criterion (Table A2.2). Bayesian phylogenetic inference (BI) was performed with MRBAYES 3.2.6 (Ronquist et al. 2012) on the separate and concatenated sequences, for the latter with three data partitions (rDNA, TOPO6-11, chloroplast). Two Markov chain Monte Carlo (MCMC) analyses with four chains each were run for 4×10^6 generations, sampling a tree every 1000 generations. The initial 25% of trees were discarded as burn-in. The phylogenetic trees were visualized in FIGTREE v1.3.1 (Rambaut 2009) and finalized in a graphics program.

For the chloroplast loci *trn*L-F and *mat*K-*trn*K resolution of the backbone of the tree was rather low, even when combined and analysed together. To obtain chloroplast haplotype genealogy insertions/deletions (indels) in the alignment that likely originated from single

mutational events were reduced to single alignment positions. Variation at a di-nucleotide and a mono-nucleotide SSR in *mat*K-*trn*K was excluded from the analysis. This shortened alignment was subjected to statistical parsimony analysis in TCS (Clement et al. 2000) that is able to deal with low genetic diversity among sequences (Posada and Crandall 2001; Jakob and Blattner 2006). To take into account the information of indel positions, gaps were treated as a 5th character state. The number of shared alleles between the species was inferred using the R package POPGENREPORT (Gruber and Adamack 2015).

In the absence of a reference genome, the assembly of the GBS data was done *de novo*. Barcoded reads from the 14 samples were de-multiplexed using the CASAVA pipeline 1.8 (Illumina, Inc.). Adapter trimming of GBS sequence reads was performed with CUTADAPT 1.15 (Martin 2011) and reads shorter than 60 bp after adapter removal were discarded. Further filtering and clustering was carried out using the software pipeline IPYRAD 0.7.5 (Eaton 2014). The minimal number of samples per locus was set to 14, the clustering threshold of reads within and between individuals was set to 0.9. The maximum ploidy level was appointed as diploid since most analyzed individuals are diploids. For the other parameters the default settings of parameter files generated by IPYRAD were used.

For the analyses of GBS data, due to the restriction to 100 kbp alignment length in MRBAYES, 1009 loci out of 1409 were concatenated and a MP analysis (using the branch-and-bound algorithm) and BI analysis (running for 1×10^6 MCMC generations) were performed as described above.

Results

The concatenated alignment of both rDNA regions had a length of 1107 bp (Table A2.2) and was the most variable dataset of the three data partitions with 125 parsimony informative sites (PIS). In comparison, TOPO6–11 (744 bp) had only about half of PIS (60) of the rDNA data set and the concatenated chloroplast data (1191 bp) had the fewest PIS (17).

Aligning the five concatenated loci resulted in an alignment of 3042 bp length with 208 PIS (Table A2.2). For individuals with ITS or TOPO6–11 paralogues or homeologues, we created haplotypes (pseudo-individuals) by combining the different ITS/TOPO6–11 copies each with the otherwise uniform sequences of *trn*L-F, *mat*K-*trn*K, ETS, ITS and TOPO6–11.

In the BI tree (Fig. 2.1) *C. baytopiorum* is sister to all other taxa of series *Crocus.* BI showed that *C. naqabensis* is sister to *C. moabiticus*. Next in the progression of successively branching clades is *C. asumaniae*, then *C. pallasii* followed by a tritomy. The tritomy consists of *C. kofudagensis*, a clade comprising the former *C. pallasii* subspecies (*C. dispathaceus*, *C. haussknechtii*, *C. turcicus*) and *C. mathewii* as sister to *C. macedonicus*, and the species of core series *Crocus* (*C. cartwrightianus*, *C. hadriaticus*, *C. oreocreticus*, *C. sativus*, *C. thomasii*). Apart of two exceptions all species analyzed here are monophyletic. *Crocus hadriaticus* (Peloponnese) is sister to a clade comprising another *C. hadriaticus* population (from Pindus Mts.) and *C. thomasii*. *Crocus oreocreticus* is sister to the clade formed by *C. cartwrightianus* and *C. sativus*. In all *C. sativus* individuals analyzed we found two sequence types of the single-copy locus TOPO6-11, which we interpret as fixed homeologs of the triploid. The pseudo-individuals with one TOPO6-11 copy (indicated as copy A in Fig. 2.2) are sister of a large polytomy comprising *C. cartwrightianus* and the *C. sativus* pseudo-individuals characterized by the second TOPO6-11 copy (indicated as copy B in Fig. 2.2). The close relationship of both homeologs with *C. cartwrightianus* sequences indicates that only this species contributed genetic material to the triploid. In contrast, the different TOPO6-11 and ITS haplotypes in the other species seem to be allelic variants (*C. cartwrightianus*) or paralogs (*C. kofudagensis*, *C. turcicus*). The strict consensus MP tree (Fig. A2.1) of the combined dataset was mainly compatible with the BI result. The BI and MP trees of the rDNA data set (the BI tree is provided as Fig. A2.2) were similar to the trees of the combined data set, but generally support values of clades were lower and relationships within the clade comprising *C. cartwrightianus*, *C. oreocreticus* and *C. sativus* were poorly resolved. Although the phylogenetic trees obtained from TOPO6-11 did not contradict the results of the combined analyses, they resulted in low resolution within series *Crocus* (Fig. A2.3).

The phylogenetic analysis of the combined chloroplast loci resulted in very low resolution of the series *Crocus* taxa (Fig. A2.4). Therefore, we also analyzed these sequences with a statistical parsimony network approach. The haplotype genealogy (Fig. 2.2) revealed that all *C. cartwrightianus* samples studied here share the same chloroplast haplotype with *C. sativus*. The only difference between some individuals of *C. cartwrightianus* and *C. sativus* were observed in a mono-nucleotide repeat that was excluded from the TCS analysis. The *C. cartwrightianus*/*C. sativus* haplotype was found four mutational steps away from *C. hadriaticus*. The investigated *C. hadriaticus* individuals from two different locations (Peloponnese, Pindus Mts.) share the same haplotype, being separated by one step from *C.*

oreocreticus. The *C. pallasii* chloroplast type grouped four to seven steps away from the haplotypes of its former subspecies. Those taxa are also separated from each other by at least two mutational steps except for *C. turcicus* and *C. haussknechtii*. They were found to share an identical haplotype in the shortened alignment. However, they differ in the number of repeats of a di-nucleotide microsatellite (AT), which was excluded from the analysis. *Crocus macedonicus* and *C. asumaniae* are placed at internal branching points close to the root of the network, indicating that these are ancient chloroplast haplotypes within the series. In contrast to the BI tree, the *C. naqabensis* haplotype was found to group close to *C. turcicus*, separated by six mutational steps from *C. moabiticus*, its closest relative according to the analysis of the combined data. *Crocus kofudagensis* was separated by two mutational steps from *C. asumaniae*. Both taxa also differ in the number of repeats in the excluded SSR regions.

Concatenation of 1009 loci obtained by GBS resulted in an alignment of 99,948 bp length with 1567 variable sites of which 753 were parsimony informative. *Crocus hadriaticus* was defined as outgroup according to the multi-locus analysis. *Crocus oreocreticus* grouped as sister to a clade comprising *C. cartwrightianus* and *C. sativus* (Fig. 2.3). Within this clade, *C. sativus* accessions grouped together as sister to one *C. cartwrightianus* individual from Crete. The topology of the MP tree (Fig. A2.5) was identical to the BI tree.

Within the GBS data set *C. sativus* possessed the highest heterozygosity ranging from 0.011 to 0.013 followed by *C. cartwrightianus* (0.008 to 0.01). *Crocus sativus* shares about 93% of its alleles with *C. cartwrightianus*, 88% with *C. oreocreticus* and 83% with *C. hadriaticus*. Although *C. sativus* is a triploid, in most cases only two alleles are present.

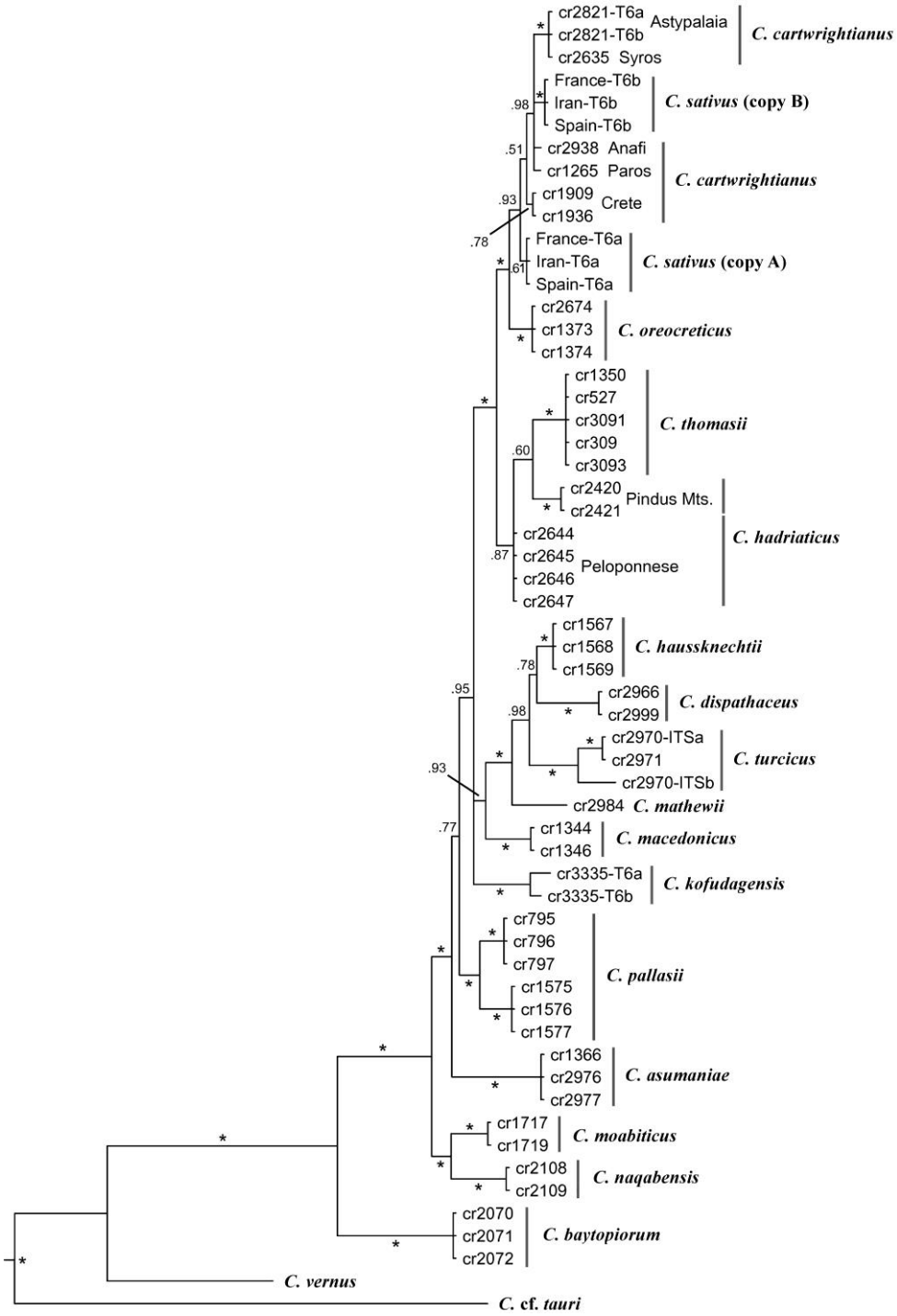


Figure 2.1. Phylogenetic tree derived from Bayesian phylogenetic inference (BI) of the combined sequences of nrDNA external (ETS) and internal transcribed spacer (ITS), the single-copy locus TOPO6-11, and the chloroplast regions *trn*L-*trn*F and *mat*K-*trn*K of all series *Crocus* and two outgroup taxa. Numbers along branches depict BI posterior probabilities, with values ≥0.99 indicated by asterisks. Different haplotypes found within single individuals are depicted by either T6a-b, when based on TOPO6-11 variation, or ITSa-b, if variation in the ITS region was detected by cloning and sequencing of PCR amplicons.

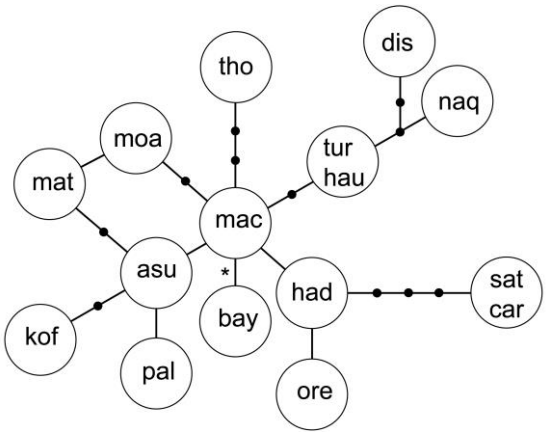


Figure 2.2. Numerical parsimony network derived from sequence analysis of two combined chloroplast regions (*trn*L-F, *mat*K-*trn*K). Black dots represent empty haplotypes, i.e. intermediate types not found in the studied individuals. Taxon names are abbreviated by the first three letters of the species epithets. The asterisk marks the root of the network according the result of the phylogenetic analysis of the combined data of all sequenced loci (Fig. 2.1).

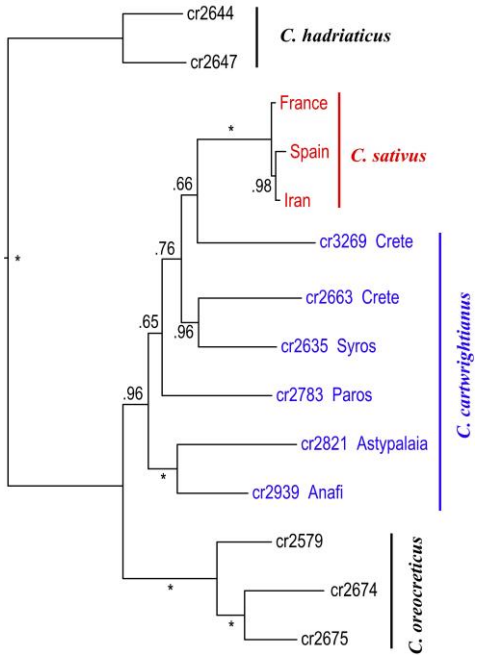


Figure 2.3. Phylogenetic tree derived from Bayesian phylogenetic inference (BI) of genotyping-by-sequencing data of 1009 loci. Numbers along branches depict BI posterior probabilities, with values ≥0.99 indicated by asterisks. *Crocus hadriaticus* was defined as outgroup.

Discussion

Phylogenetic relationships within ser. Crocus

Using a five-locus dataset with a combined alignment length of 3042 bp, we here infer phylogenetic relationships of all taxa of series *Crocus*, providing higher resolution compared to the previous studies (Petersen et al. 2008; Harpke et al. 2013; Larsen et al. 2015) and also including multiple individuals for most species. Our results confirm the importance of multi-locus studies for the inference of phylogenetic relationships in series *Crocus*. Considering variability, ITS and TOPO6 intron 11 were the most informative loci. Both are characterized by several single-nucleotide polymorphisms and insertions/deletions. Therefore, these markers allowed resolving previously unclear taxon relationships. In contrast, the chloroplast markers and ETS showed a rather low level of variation among taxa. The low variability of chloroplast DNA in series *Crocus* was already shown in the study of Seberg and Petersen (2009), where even the use of six chloroplast regions did not provide good resolution within narrow species groups of *Crocus*.

Our BI analysis of the combined data revealed that *C. pallasii* in its traditional circumscription is not monophyletic. Former *C. pallasii* subsp. *dispathaceus*, subsp. *haussknechtii*, and subsp. *turcicus* are closer related to *C. macedonicus* and *C. mathewii* than to *C. pallasii* subsp. *pallasii*. Therefore, we confirm that they should be treated as separate species. Their evolutionary independency is further supported by different chromosome numbers (Brighton et al. 1973, 1977) and genome sizes (Erol et al. 2014). In addition, within *C. dispathaceus* morphological differences and different genome sizes between populations might suggest the presence of separate evolutionary lineages (Erol et al. 2014). The taxonomical status of these populations has still to be evaluated. Unfortunately, the situation here is complicated since the type of *C. dispathaceus* (Bowles 1952) probably came from the war zone of today's Syria, making a closer examination of these populations currently impossible. *Crocus turcicus* material used in this study was collected from the type locality. However, the distribution area of the taxon or possible undefined relatives seems wider, and also here no widespread sampling could be made recently due to security considerations. For the Al-Hasakah region in northeastern Syria *C. macrobolbos* was described (Gombault 1956) but later became a synonym for *C. pallasii* subsp. *turcicus* (Mathew, 1982). Examination of the type specimens (MNHN-P-P00622947 to MNHN-P-P00622957) shows that *C. macrobolbos* is probably not identical to *C. turcicus* but seems to

be another species belonging to series *Crocus*. However, detailed molecular and morphological analyses are necessary to infer the taxonomic status of these taxa.

The two *C. hadriaticus* populations we investigated are not monophyletic, although they still share the same chloroplast haplotype. Larsen et al. (2015) found *C. hadriaticus* from the Peloponnese separated from their samples from the Pindus Mts., too. Based on morphology, *C. hadriaticus* was already split into three subspecies (Mathew 2000) mainly due to throat and segment colors. However, Larsen et al. (2015) observed color variation co-occurring within populations, making a sub-specific rank for the units described by Mathew (2000) impossible. Further detailed morphological and molecular investigations are necessary to characterize the populations from the Peloponnese and mainland Greece to clearly define genetic borders and the taxonomical status of the populations from the different regions.

Up to now, *C. naqabensis* has not been included in any phylogenetic study. This species is native to Jordan (Ras en Naqb). Apart from this species, *C. pallasii* subsp. *haussknechtii* was also reported for Jordan for the same location (Brighton 1977; Mathew 1982; Kerndorff 1988) and/or region (Kerndorff 1988; Peri 2016). According to recent investigations the *crocus*es at the type location do not exit any more (Peri 2016). We analyzed material growing some kilometers north of the type locality (same elevation, same habitat) and our results reveal that these individuals are different from *C. haussknechtii* from Iran (Kurdistan, type location). Despite the differences in morphology as well as the number of chromosomes, Iranian *C. haussknechtii* (2n = 12) was found to be closely related to *C. dispathaceus* (2n = 14), while the individuals from Jordan grouped together with *C. moabiticus* in the nuclear dataset. In the chloroplast dataset they appear closer to *C. turcicus* (2n = 12) and *C. haussknechtii* than to *C. moabiticus*. The incongruence between chloroplast and nuclear marker might indicate a hybrid origin, which chloroplast loci could not show due to their uniparental maternal inheritance. However, by considering the differences in the nuclear loci of Jordanian taxa to all other former *C. pallasii* subspecies, it is clear that the material of the population in Jordan represents an independent taxon. Therefore, we suggest that specimens from Jordan previously considered as *C. pallasii* subsp. *haussknechtii* probably all represent *C. naqabensis*.

Another species studied here for the first time is *C. kofudagensis*, whose position as an independent species is clearly confirmed by our molecular data. However, the closest relative of *C. kofudagensis* remains unclear. Based on morphology, Rukšans (2017) suggested an affiliation to *C. asumaniae*. This could not be confirmed in our combined

analysis. There is a geographical distance of *c.* 200 km between *C. asumaniae* and *C. kofudagensis* populations. Moreover, these areas are separated by a series of high mountain chains (Western Taurus Mountains). Although the closest relative of *C. kofudagensis* according the chloroplast data is *C. asumaniae*, both taxa are probably separated for an extended timespan, as they already accumulated two sequence mutations and additional variation at the two excluded SSR motives in the analyzed chloroplast markers.

A high degree of infra-specific genetic variability and heterozygosity in TOPO6–11 was detected in *C. cartwrightianus*. This is in agreement with the results of Larsen et al. (2015). Their AFLP and microsatellite analyses showed that *C. cartwrightianus* is characterized by high genetic diversity. Their data also revealed a close affiliation of *C. cartwrightianus* and *C. sativus* with *C. oreocreticus*. The close relationship of the three species is confirmed by our multi-locus and GBS analyses.

Phylogenetic affiliation of C. sativus

Phylogenetic analyses of all datasets we used placed *C. sativus* within *C. cartwrightianus* indicating this species as closest relative of saffron. In *C. sativus* we found also two TOPO6–11 copies, which were identical in all investigated saffron individuals and are interpreted as homeologs fixed in this species. These sequences are very similar to the alleles found in *C. cartwrightianus* and group in the phylogenetic tree within the *C. cartwrightianus* clade. This suggests that both parental genotypes of triploid saffron belonged to *C. cartwrightianus*, as otherwise at least one of the homeologs should fall outside of this clade. The latter is supported by the GBS analysis, where *C. sativus* shares most of its alleles with *C. cartwrightianus*. The number of shared alleles would probably even increase with the inclusion of more individuals of *C. cartwrightianus*, which would allow sampling more of the genetic diversity present in this species.

Our data support the result of Larsen et al. (2015) who also found *C. cartwrightianus* to be the species most similar to *C. sativus*. Several other studies hypothesized *C. cartwrightianus* to be an ancestor of *C. sativus* (Grilli-Caiola et al. 2004; Zubor et al. 2004; Frizzi et al. 2007; Petersen et al. 2008; Ahrazem et al. 2010). However, up to now it was not clear if *C. cartwrightianus* was the only parent or if other species had contributed to it through allopolyploid formation. While karyological investigations suggested saffron to be

an autotriploid (Brigthon, 1977; Ghaffari, 1986) recent investigations of nuclear low-copy genes hypothesize allotriploidy (Tsaftaris et al. 2011; Harkpe et al. 2013). Thus, *C. thomasii*, *C. pallasii*, *C. mathewii* from Turkey, *C. serotinus* Salisb., *C. haussknechtii*, *C. michelsonii* B.Fedtsch. and *C. almehensis C.*D.Brickell & B.Mathew from Iran were mentioned as possible parents (Frello and Heslop-Harrison 2000; Grilli-Caiola et al. 2001, 2004; Alavi-Kia et al. 2008; Tsaftaris et al. 2011; Gismondi et al. 2013; Erol et al. 2014). Most probably, this resulted from the inclusion of only very limited *C. cartwrightianus* materials in these studies. Furthermore, the phylogenetic markers used were mostly not variable or informative enough. In contrast to these studies, we here included individuals of all known species of series *Crocus* and had *C. cartwrightianus* from five different locations in Greece in our analysis. The high variability among different *C. cartwrightianus* individuals and the high degree of heterozygosity found within *C. sativus* makes an autopolyploid origin of saffron possible, i.e. different genotypes of *C. cartwrightianus* hybridized resulting in the triploid.

Conclusion

Our phylogenetic analyses resolved the relationships among all taxa of the series. *Crocus hadriaticus* and the former *C. pallasii* subspecies appeared polyphyletic. The latter deserve elevating the subspecies to species rank, while for *C. hadriaticus* a detailed study of species boundaries is necessary. Multi-locus and also genome-wide single nucleotide polymorphism data obtained through genotyping-by-sequencing placed *C. sativus* within *C. cartwrightianus* with no indication that other *Crocus* species contributed to the evolution of the triploid. Our analyses thus made an autotriploid origin of *C. sativus* from *C. cartwrightianus* very likely.

3. Identification of parental and geographical origin of saffron

Introduction

The geographic origin of saffron has been a topic for long-lasting speculations naming different locations like Crete, Egypt, Greece, India, Iran and Spain as possible domestication centres. Also the mode of saffron evolution has been a subject of speculations since a long time. Some researchers believed that saffron is a triploid clone derived from one species through autotriploidy and therefore it is not a true species (Brigthon 1977; Ghaffari 1986). Alternatively, it is thought to be an allopolyploid hybrid, i.e. the result of a cross between diploid parental species of which one contributed an unreduced gamete (Brighton 1977; Agayev 2002). However, none of the aforementioned studies were successful to provide strong support for their claim mainly due to either lack of enough variable markers or an adequate sampling strategy. The result of our molecular phylogeny using chloroplast and nuclear markers presented in the previous chapter provides the hypothesis that only *C. cartwrightianus* is the parental species of *C. sativus* but those different genotypes of *C. cartwrightianus* hybridized resulting in the triploid. Therefore, we assumed an autopolyploid origin either through diploids producing reduced and unreduced gametes or through a cross between a di- and a tetraploid parent. Crocuses have no specialized morphological structures promoting long-distance dispersal of seeds (mostly the seeds shed and then are moved by ants). Thus, we expected that there might

be a good chance to identify an area where saffron-like allele combinations still occur within the local *C. cartwrightianus* population(s). By identifying an area where most of the saffron alleles occur and which is different from allele composition in other areas it could function as a proxy for the region of origin of saffron. Currently next-generation sequencing (NGS) clearly holds promise for fast and cost-effective detection of genome-wide SNP data, for phylogeography and phylogenetics (McCormack et al. 2013). One of the most promising techniques of library preparation for NGS is genotyping-by-sequencing (GBS) (Elshire et al. 2011). This technique uses restriction enzymes to reduce genome complexity and genotype several DNA samples. Therefore, we here used this technique to infer the area(s) where saffron initially originated by identifying regions/populations that are genetically most similar to saffron. As *Crocus* species are obligate outcrossing taxa, we did not expect to find individuals having today the same allele constitution as saffron. Therefore, we had to take into account not only overall similarity (e.g., grouping together in a phylogenetic tree) but look for the allelic constitution of saffron and search for populations within today's distribution area of *C. cartwrightianus* possessing the highest amount of saffron-like alleles. To test our claim that only *C. cartwrightianus* contributing to saffron, we here also analyse several individuals of all wild species in the series *Crocus* with the same basic chromosome number as *C. sativus*, i.e. including *C. cartwrightianus, C. hadriatius, C. thomasii, C. pallasii* and *C. oreocreticus* using bi-parentally inherited single-copy loci to infer where possible homeologs of saffron can be found.

Materials and Methods

Nuclear single-copy loci

Taxon sampling

Twenty-two individuals comprising 12 *C. cartwrightianus*, one *C. hadriaticus*, one *C. oreocreticus*, one *C. thomasii*, one *C. pallasii* subsp. *pallasii* and one *C. sativus* were included in an analysis of single-copy loci to find the positions of *C. sativus* homeologs (Table A3.1).

Molecular methods

DNA was extracted from the fresh leaves using the DNeasy Plant Mini kit (Qiagen). We here used 14 single-copy or low copy-number genes from the nuclear genome of *C. sativus* and found in five genes indications for the presence of more than one copy while in the other genes no intra-individual polymorphisms were detected in saffron. We PCR amplified and cloned amplicons of the five polymorphic genes and analyzed phylogenetic relationships at these loci in order to identify the species that the homeologs of saffron were derived from (Table A3.2).

Data analyses

Phylogenetic analysis using neighbor-joining (NJ) was performed in PAUP* 4a157 (Swofford, 2002) calculating GTR distances. Bootstrap support values were obtained by 500 bootstrap re-samples using the same settings as before.

Genotyping-by-sequencing

Taxon sampling

One hundred and fourteen individuals comprising 96 *C. cartwrightianus* from 19 populations, ten *C. sativus* individuals and eight individuals of *C. oreocreticus* as outgroup species were included in a GBS analysis to accurately define the area of origin of *C. sativus* (Table A3.3).

Molecular methods

DNA was extracted using the DNeasy Plant Mini kit (Qiagen). To prepare the libraries for GBS, we followed the two-enzyme method described by Poland et al. (2012). Essentially, we digested 200 ng of each genomic DNA with the methylation-sensitive restriction enzyme *Msp*I together with the non-sensitive and rare-cutting enzyme *Pst*I to arrive at restriction sites mostly situated in the under-methylated gene space of the genome. Afterwards, barcode and common adapters for Illumina sequencing were ligated to the

fragments. After PCR amplification to enrich the desired fragments, 1×100 bp single-end sequencing was performed on the Illumina HiSeq 2500 platform.

Data analyses

Raw GBS reads were assembled and clustered using the program IPYRAD 0.7.5 (Eaton and Overcast, 2016). Raw sequence reads of *C. cartwrightianus*, *C. sativus* and *C. oreocreticus* were demultiplexed, quality trimmed, within and between samples clustered and the output files filtered for a clustering threshold of 90%, and minimum number of sample per locus of 100 which means only the loci that are present in at least 90% of samples were called.

Bayesian Inference (BI) analyses were performed in MRBAYES 3.2.6 (Ronquist et al. 2012b) using the models inferred by JMODELTEST (Posada 2008). Each analysis consisted of two independent analyses each running four sequentially heated chains for 1 x 10^6 generations and sampling a tree every 1000 generations. The first 25% of sampled trees were discarded as burn-in and a consensus tree was computed in MRBAYES. To see if the phylogenetic relationships inferred by BI are robust regarding different analysis algorithms, a parsimony analysis (MP) of all taxa was conducted in PAUP* 4a152 using a heuristic search. Node support was evaluated by 1000 bootstrap re-samples with the same settings.

To estimate the ancestry coefficients of C. *cartwrightianus* and *C. sativus* individuals, I further performed STRUCTURE-like Bayesian assignment tests with the LEA package in R (Frichot and Francois 2015) for each parental population (K = 3-8). The data are used to identify the degree of genetic relatedness of *C. cartwrightianus* and *C. sativus* individuals in order to find the *C. cartwrightianus* populations most closely resembling the SNP composition within *C. sativus* and to clarify simultaneously the number of times that saffron originated based on the genetic variation between *C. sativus* individuals. I also performed a principal component analysis (PCA) on all SNPs using the R package SNPRelate version 1.12.2 (Zheng et al. 2012) and the R package ggplot2 version 2.2.1 (Wickham 2009).

WGS/Genome skimming to obtain chloroplast genomes

Taxon sampling

Fresh leaves one *C. sativus* individual, one *C. cartwrightianus* individual from southern Attica and one *C. cartwrightianus* individual from Crete was collected in order compare their chloroplast genomes.

Molecular methods

DNA was extracted from the fresh leaves using the DNeasy Plant Mini kit (Qiagen). To construct the sequencing library, 200 ng of genomic DNA was used as input material for the following steps. DNA was sheared to produce fragments with an average size of 300-400 base pairs (bp) followed by adaptor and barcode ligation. Size selection was then performed and fragment-size distribution and DNA concentration were evaluated on an Agilent BioAnalyzer (Agilent Technologies USA). Afterwards, the fragments belonging to *C. cartwrightianus* from Crete were sequenced on the Illumina HiSeq 1000 platform (2 x 100 bp paired end) while for *C. cartwrightianus* from southern Attica and *C. sativus*, a 2 x 250 bp single-end sequencing was performed on the Illumina HiSeq 2500 platform.

Data analyses

We used CUTADAPT 1.15 (Martin 2011) to remove adapters and perform quality trimming (minimum length of 75 bp). For genome skimming, we first generated *de novo* genome assemblies of each individual using CLC 4.3.0 (CLC bio) and plastid scaffolds were then identified using BLAST searches. The scaffolds were mapped against the plastid chromosome of *Iris sanguinea* (GenBank accession number KT626943.1). The reads were then mapped against the *I. sanguinea* plastid chromosome using BOWTIE2 v2.2.9 (Langmead and Salzberg 2012) and manually examined by visualization using SAMTOOLS v 1.8 (Li et al. 2009). In order to identify the maternal progenitor of saffron, as the chloroplast genome is inherited maternally, the chloroplast genomes of *C. sativus* and *C. cartwrightianus* from Crete and Attica were aligned and compared and primers were designed from conserved sequences adjacent to deletion/insertion regions. The same regions of *C. sativus* individuals and *C. cartwrightianus* individuals from the entire distribution area were then PCR amplified and visualized on agarose gels.

Flow cytometry

We used flow cytometry to estimate genome size and ploidy level of 91 *C. cartwrightianus* samples from southern Attica in order to get information about the frequency of individuals with a ploidy level of more than two. Since flow cytometry needs a reference measurement against that other measurements can be compared, we here selected *Vicia faba* as the reference standard since it had with 26.66 pg 2C an approximately three-fold genome size in comparison to *C. cartwrightianus*. Nuclei were extracted from 20- to 100-mg portions of fresh leaves of *C. cartwrightianus* samples and *V. faba* following Dolezel et al. (1989). Shortly, leaves from each sample were chopped together with the reference using a sharp razor blade in 250 µl extraction buffer solution (Galbraith et al. 1983). First, the reference standard was processed alone and the standard peak was positioned by changing the gain values for the photomultiplier tubes. The instrument settings were saved and used for all measurements using this standard on that day. The solution containing cell particles was then passed through a filter to remove cell fragments and large particles. Afterwards, isolated nuclei in lysis buffer were stained using 1 ml DAPI buffer (Sigma-Aldrich, Germany). After 5 min incubation time the nuclear suspension was analysed with a CyFlow Space flow cytometer (Sysmex Partec GmbH, Germany).

Results

Single-copy loci phylogenies

To test our hypothesis that no other species contributed to the formation of the triploid, we here sequenced five single-copy genes with more than one copy in saffron. The presence of two different gene copies in the triploid might indicate that different species were involved in its formation. We here expected that the alleles found in the triploid should all group together with alleles occurring in *C. cartwrightianus* if only this species had contributed to the formation of *C. sativus*. The phylogenetic analysis using five single copy genes showed that there is high genetic variation in *C. cartwrightianus*. We here found four out of ten alleles present in saffron to be identical and two to be very similar to alleles occurring in *C. cartwrightianus*, three grouped together with only *C. cartwrightianus* alleles and only one fell in a clade where among *C. cartwrightianus* alleles also alleles of *C.*

hadriaticus and *C. thomasii* were found (Fig. 3.1). Therefore, this result supported the hypothesis that *C. cartwrightianus* is the sole progenitor of *C. sativus*.

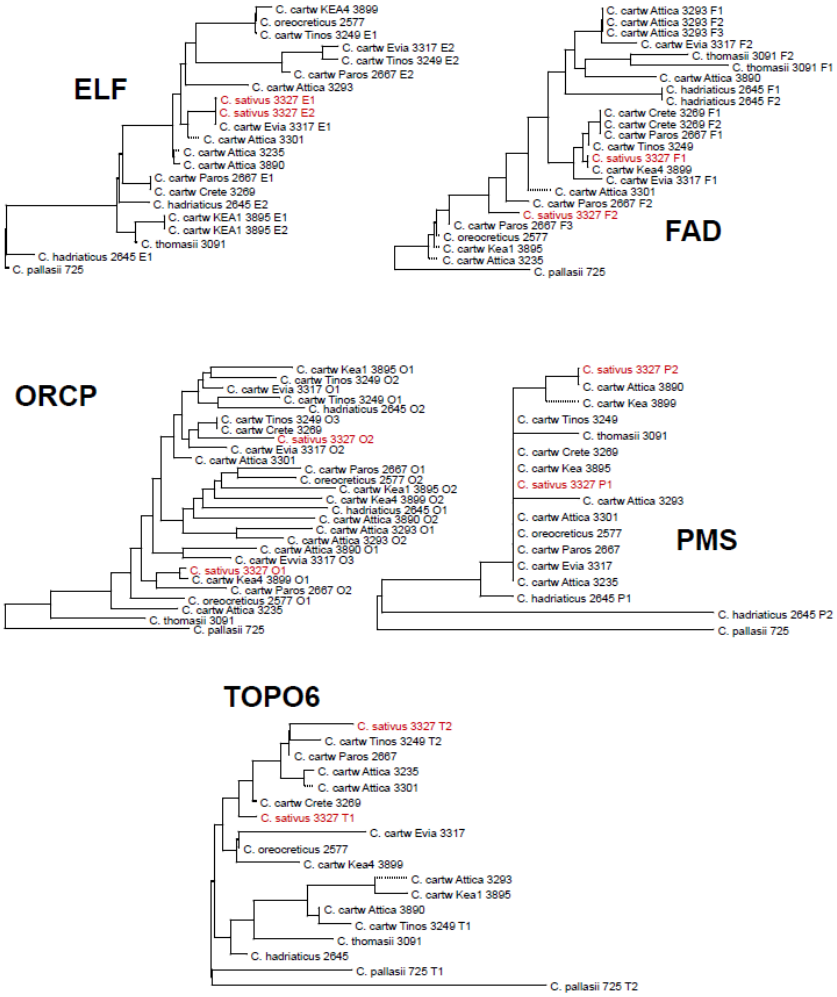


Figure 3.1. Neighbor-joining analyses of five single-copy loci in close relatives of *Crocus sativus*. *Crocus pallasii* is specified as outgroup in all analyses. Different copies (paralogs) present in single individuals are indicated by the first letter of the gene name followed be the numbers 1-3.

Genotyping-by-sequencing

To follow up on the result regarding the progenitor of saffron and in order to infer the geographic location of saffron domestication, we screened genome-wide single-nucleotide polymorphism (SNP) data obtained via genotyping-by-sequencing. The phylogenetic tree derived from 3353 GBS loci obtained by maximum parsimony (MP) showed that the samples of *C. sativus* grouped within *C. cartwrightianus* populations from the northern

part of the species' distribution range and formed the sistergroup of the Attic individuals. Together they are sister of *C. cartwrightianus* individuals from Evia and Kea. We further found a high degree of variation and admixture in the individuals collected from the islands of Anafi, Mykonos, Paros, Syros, and Tinos, which formed four clades but did not show a pattern according to their geographic origin (Fig. 3.2A).

For a population structure analysis, we identified an optimum of eight ancestral population (K = 8). This analysis also supported the contribution of only *C. cartwrightianus* from Attica in *C. sativus*, as we found the similar allelic patterns presented in *C. sativus* in *C. cartwrightianus* from Attica (Fig. 3.2B). It further revealed a high degree of heterozygosity within *C. sativus* but nearly no variation among individuals. In contrast to saffron, we found a high degree of admixture and heterozygosity within and among *C. cartwrightianus* individuals and populations.

The close genomic similarity of Attic *C. cartwrightianus* and *C. sativus* was also confirmed in a principal component analysis of the GBS dataset where we found an overlap between *C. sativus* and *C. cartwrightanus* from southern Attica (Fig. 3.3).

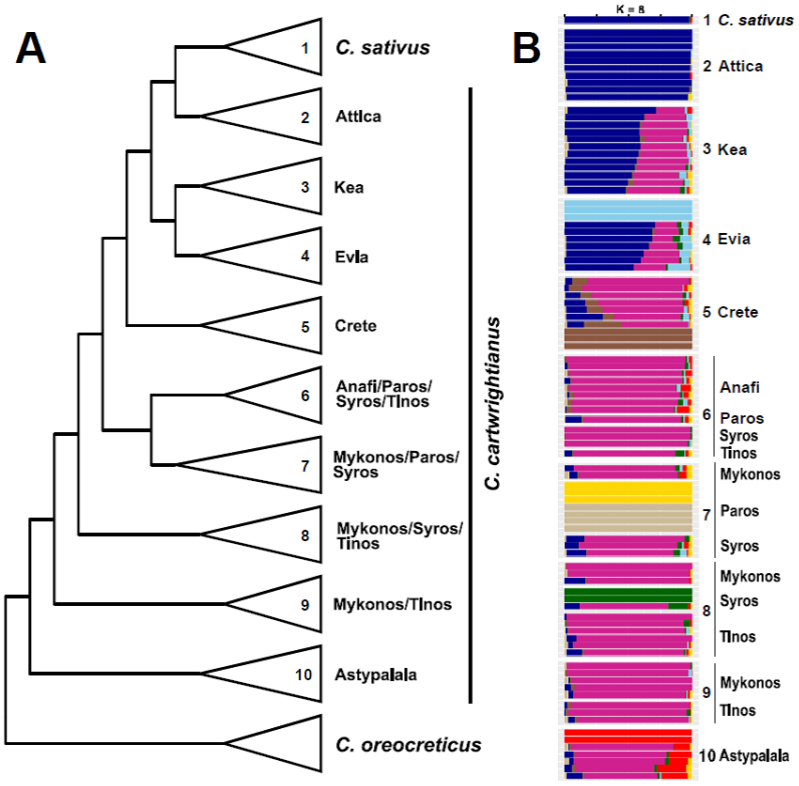


Figure 3.2. Schematic representation of the phylogenetic relationships of *Crocus cartwrightianus* with regard to *C. sativus.* A Scheme summarizing the six most parsimonious trees based on

genome-wide DNA data of a genotyping-by-sequencing analysis of 114 individuals. *Crocus oreocreticus* was defined as outgroup taxon. Saffron (*C. sativus*) groups in *C. cartwrightianus* as a sister to the individuals from Attica. **B** Result of a Bayesian population assignment analysis for K = 8.

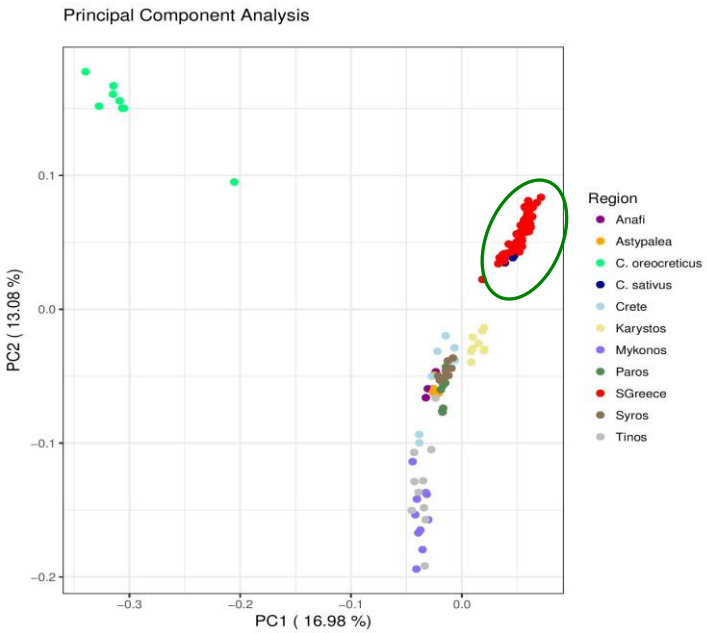


Figure 3.3. Principal component analysis of the GBS dataset. Saffron groups with individuals from Attica (highlighted by the green outline).

WGS/Genome skimming

To identify the maternal parent of saffron, we analyzed DNA differences in the maternally inherited chloroplast genome. To obtain initial information about potentially polymorphic and informative loci, we used genome skimming, that is low-coverage whole genome shotgun sequencing, to be able to assemble the chloroplast genomes of *C. sativus* and two *C. cartwrightianus* individuals from the southern (Crete) and northern (Attica) borders of the species' distribution area. By comparing the chloroplast genomes, we found a 84 bp insertion/deletion in the *trn*S(GCU)–*trn*G(UCC) intergenic spacer with the shorter allele occurring in *C. sativus* and the Attic *C. cartwrightianus* individuals (Fig. 3.4).

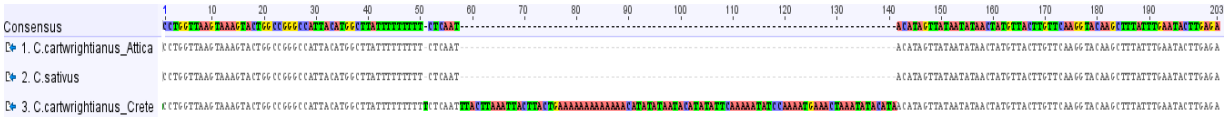


Figure 3.4. Eighty-four base pair deletion in the *trn*S(GCU)–*trn*G(UCC) intergenic spacer of the chloroplast genome of *C. sativus* and Attic and Crete *C. cartwrightianus*.

Therefore, we designed primers for this locus and performed screening PCR amplicon sizes in *C. sativus* together with *C. cartwrightianus* individuals from entire distribution area. We here found that *C. sativus* possesses the short allele while in *C. cartwrightianus* it occurs only in Attica and in one out of two populations from the island of Kea. In all other populations of *C. cartwrightianus* only the longer allele was detected (Fig. 3.5). Sequencing of this locus then confirmed complete sequence identity between saffron and the Attic wild saffron.

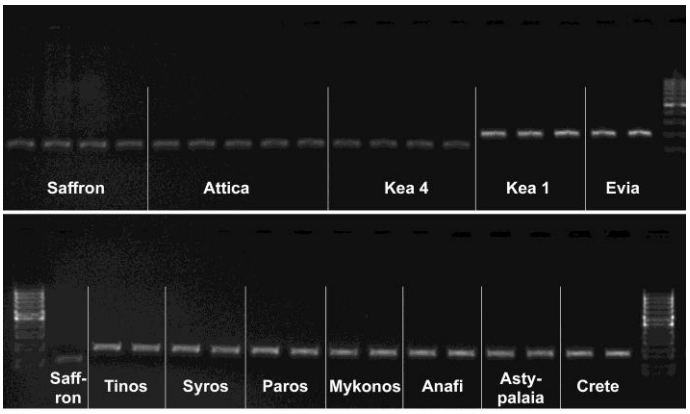


Figure 3.5. Screening PCR amplicon sizes of the *trn*S(GCU)–*trn*G(UCC) intergenic spacer in representative individuals of *C. sativus* and *C. cartwrightianus* from its entire distribution area. We found the short allele in *C. sativus* and *C. cartwrightianus* in Attica and in one out of two populations from the island of Kea while all other populations of *C. cartwrightianus* possess only the longer allele.

Flow cytometry

To be able to infer the mode of origin of the triploid *C. sativus*, i.e. either if it evolved through a cross between a di- and a tetraploid parent or through the combination of a reduced with an unreduced gamete within diploids, we analyzed genome sizes for 91 *C. cartwrightianus* individuals in the Attica area. We found that all of them have a 2C genome size of 7.06 ±0.09 pg indicating that they are all diploid.

Discussion

To identify the area(s) where saffron initially originated we here studied 3353 loci obtained from GBS of *C. sativus*, *C. cartwrightianus* from different populations in Greece as well as *C. oreocreticus* that was used as an outgroup. The GBS analyses resulted in identifying the Attica region in mainland Greece as the most likely area for the evolution of triploid *C. sativus*.

Furthermore, these analyses also revealed a high degree of fixed heterozygosity within *C. sativus* individuals but nearly no variation between them. Therefore, we assume that *C. sativus* is an autopolyploid that originated in southern Attica by combining two different genotypes of *C. cartwrightianus*. The result obtained from GBS analyses also showed a high degree of heterozygosity and admixture within and between *C. cartwrightianus* populations growing in Anafi, Mykonos, Paros, Syros and Tinos. From this we assume that probably the central Aegean is the area of origin of *C. cartwrightianus* and this species reached later Attica/Kea where *C. sativus* then originated in Attica.

Screening amplicon sizes for the chloroplast region *trn*S(GCU)–*trn*G(UCC) in *C. sativus* together with *C. cartwrightianus* individuals from all populations, we found that *C. sativus* possesses the short allele while in *C. cartwrightianus* it occurs only in Attica and in one out of two populations from the island of Kea. Kea is directly adjacent of the southern tip of Attica and was connected to the mainland repeatedly when during Quaternary the sea level dropped (Lambeck 1996). Sequencing of this locus confirmed complete sequence identity between saffron and Attic wild saffron. This revealed that the maternal parent of the saffron crocus occurred in the Attica/Kea area of Greece. Furthermore, since we only found the deletion in chloroplast genome of *C. cartwrightianus* populations from Attica/Kea, we assume that this mutation originated in this area and it could also support our assumption that *C. cartwrightianus* occurred in Attica/Kea only after the central Aegean islands.

To be able to infer the mode of origin of the triploid, i.e. if it evolved through a cross between a di- and a tetraploid parent or through the combination of a reduced with an unreduced gamete within diploids, we analyzed genome sizes of 91 Attic *C. cartwrightianus* individuals. In case that saffron arose from crossing between tetraploid and diploid individuals, we would expect to find relatively high number of tetraploid individuals in the present Attic *C. cartwrightianus* populations. We found only diploid individuals and we therefore conclude that *C. cartwrightianus* in Attica seems to be mostly

diploid and that tetraploid plants are rare or absent. Therefore, we assume that saffron probably arose through crossing between diploid Attic *C. cartwrightianus* individuals with a reduced and an unreduced gamete. This result is in agreement with the observation that *C. sativus* is a clonal lineage that arose most probably only once showing that this was a rare event. If there were high numbers of tetraploid individuals within the diploid *C. cartwrightianus* populations, triploid plants should originate several times that would broaden the genetic basis of saffron through time. This is, however, not the case since saffron has a relatively narrow genetic base (Nemati et al., 2014; Larsen et al., 2015). Although we here could not determine the role of triploidy in the development of the typical traits of saffron, the identification of the parental origin of saffron now contributes to overcoming the low genetic variation occurring in saffron, since it will allow recreating new saffron genotypes from different Attic *C. cartwrightianus* individuals.

Conclusion

Here we provide evidence that diploid *Crocus cartwrightianus*, a species occurring in southern mainland Greece and on the Aegean Islands, is the sole progenitor of the saffron crocus. Phylogenetic analyses of nuclear loci and genome-wide DNA polymorphisms together with chloroplast genome comparisons indicate that saffron is genetically most similar to the Attic *C. cartwrightianus* populations. Based on our data we postulate that *C. sativus* is an autopolyploid that originated in southern Attica by combining two different genotypes of *C. cartwrightianus*.

4. Transcriptome analysis of saffron and its parents

Introduction

The results obtained in the previous chapters showed that saffron originated in southern Attica and probably was domesticated there. Saffron was afterwards propagated by vegetative reproduction through corms (Brighton 1977; Mathew 1982; Fernandez 2004). Since at least 2300 years, farmers selected and replanted the corms of the saffron plants with extra-long, reddish and more aromatic stigmas (Douglas and Perry 2003). In fact the accumulation of crocin, picrocrocin, and safranal in saffron's stigmas gives unique colour, flavour, and aroma to the plant, respectively. In addition to saffron, *C. cartwrightianus* also seems to produce these valuable compounds, as it possesses relatively long reddish to orange stigmas. However, *C. sativus* with its more intensive red stigmas seems to produce higher amounts of apocarotenoids compared with *C. cartwrightianus*.

By the identification of the parental origin of saffron in the previous studies, we have for the first time the chance to compare the transcriptome and transcription activity in the stigmas of wild progenitors and the domesticated saffron crocus through RNA-seq. It will allow us to begin investigating evolutionary and developmental mechanisms contributing to the superiority of *C. sativus* over *C. cartwrightianus* in terms of the apocarotenoids production.

In this study, we compare for the first time the transcriptomes of cultivated and wild saffron to understand the molecular basis of apocarotenoid biosynthesis. To achieve our

goal, the expression profiles of all the transcripts in stigmas of *C. sativus* and the Attic *C. cartwrightianus* were analysed and several genes showing differential gene expression were identified. We further analysed the expression patterns of the putative genes involved in apocarotenoid biosynthesis. These data provide a resource to understand presence and the expression pattern of these important genes in the stigma of wild saffron and are a first step to identify traits and genotypes necessary for future improvement of saffron.

Materials and Methods

RNA isolation and sequencing

For RNA sequencing, three *C. sativus* individuals, three *C. cartwrightianus* individuals from southeast and three individuals from southwest of Attica were selected. Stigmas of *C. cartwrightianus* specimens were freshly harvested and stored in RNAlater RNA Stabilization Reagent (Qiagen). For *C. sativus*, stigmas were freshly harvested from individuals potted in the greenhouse and immediately frozen in liquid nitrogen. All stigmas were harvested in the same developmental stage and stored at -80 °C until RNA extraction.

Total RNA was extracted using RNeasy Mini Kit (Qiagen) according to manufacturer's instructions. Quantity and quality of isolated RNA was assessed in a BioAnalyzer (Agilent Technologies). Samples with the standard $1.8 \leq OD260/280 \leq 2.1$, $OD260/230 \geq 1.8$ and RNA integrity number (RIN) over 7 were then selected. Finally, RNA sequencing libraries were prepared in duplicate for *C. sativus* and *C. cartwrightianus* individuals. Libraries were sequenced on the Illumina HiSeq 2500 platform, producing 2 x 100-nucleotide paired-end reads.

TRINITY-based de novo transcriptome assembly

The quality of the raw reads as well as the GC content of the sequences was checked with FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) version 0.11.2. To use high-quality data for *de novo* assembly, the quality and adaptors trimming of the reads was performed by the CUTADAPT 1.15 software (Martin 2011) with a quality cut-off of 15 and the minimum reads length of 75 bp. Therefore, reads that are too short even

before adapter removal were also discarded. The high quality reads were then *de novo* assembled with the TRINITY 2.4.0 software using default parameters (Grabherr et al. 2011). The quality of the assembly was also checked by mapping the reads of each *C. sativus* and *C. cartwrightianus* individual to the *C. sativus* transcriptome assembly using BOWTIE2 (Langmead and Salzberg 2012).

Differential expression analysis

To estimate expression abundance of nine *C. sativus* and *C. cartwrightianus* libraries, high quality reads of each library were first aligned to the *de novo* transcriptome assembly using BOWTIE2 (Langmead and Salzberg 2012) and then transcript abundance was quantified using KALLISTO (Bray et al. 2016). To compare biological replicates across all the samples and generate the correlation matrix and a principal component analysis (PCA) plot we used the PtR script from TRINITY (Grabherr et al. 2011). Differential expression analysis of gene expression data was executed using the run_DE_analysis.pl script and define_clusters_ by_cutting_tree.pl script with the Ptree method in TRINITY, which uses the edgeR package 3.0 (Robinson et al. 2010).

We carried out a second analysis using transcripts of 78 genes involved in apocarotenoid biosynthetic pathway in *C. sativus* reported by Jain et al. (2016) to compare their expression levels in *C. sativus* and *C. cartwrightianus* individuals. For this, we used BOWTIE2 (Langmead and Salzberg 2012) for mapping of the reads, KALLISTO (Bray et al. 2016) for the gene-level count data. The DE features were partitioned into clusters with similar expression patterns using the define_clusters_by_cutting_tree.pl script.

Results

TRINITY-based de novo transcriptome assembly

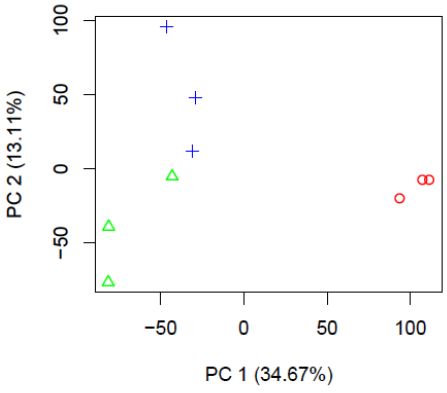
In order to compare gene expression patterns of *C. sativus* and its parent, we performed RNA sequencing for three sources of material including *C. cartwrightianus* from southeast and southwest Attica and *C. sativus* each with three biological replicates. A total of 261,499,984 paired-end reads (100 bp) were obtained after sequencing all nine libraries on the Illumina HiSeq 2500 platform. Raw reads were quality trimmed using the CUTADAPT 1.15 program (Martin 2011) and the adaptors were removed from the reads that totally

yields over 24 Gbp of reads. We here used 73,424,392 reads belonging to saffron individuals for *de novo* assembly of *C. sativus* using the TRINITY software package with default parameters. The resulting assembly consisted of 104,076 TRINITY transcripts with an N50 value of 584 bp. We further checked the quality of the *de novo* assembly by mapping *C. sativus* and *C. cartwrightianus* reads to the assembly using BOWTIE 2 (Langmead and Salzberg 2012). We obtained a high mapping rate ranging from 89% to 94% in a *C. cartwrightianus* individual from southwest of Attica and a *C. sativus* individual respectively, demonstrating the high quality of the assembly.

Differential expression analyses

Following transcriptome assembly, differential expression analysis was performed for the transcripts across all nine individuals of *C. sativus* and *C. cartwrightianus*. First of all, KALLISTO was applied for the transcript abundance estimation. We then checked for the correlation between the replicates for all the samples using the PtR script. PCA analysis and correlation matrix showed a good correlation between the replicate sets for each of the nine samples (Fig. 4.1A and B)

A



B

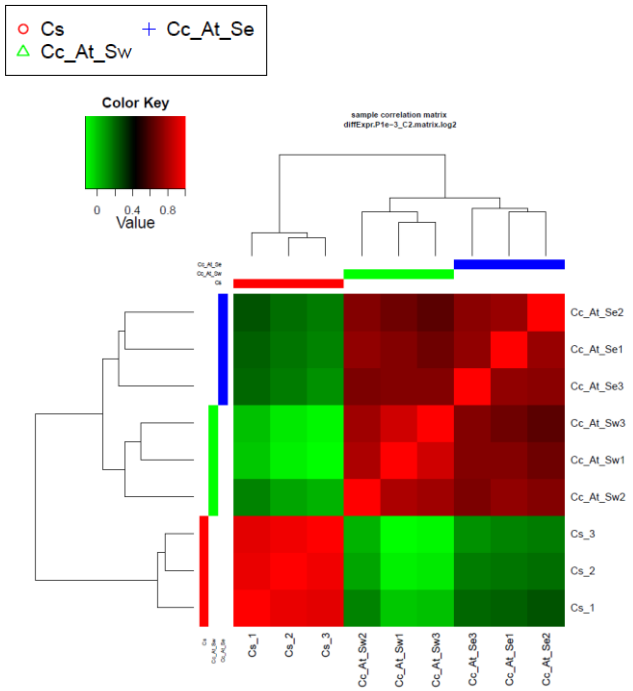


Figure 4.1. Correlation analyses showing the relationship between samples and replicates. (**A**) Principal component analysis and (**B**) heatmap showing relationship between all samples as well as replicates.

EdgeR was used to identify the differentially expressed transcripts for all the pairwise comparisons between the nine samples and a hierarchical clustering of *C. sativus* and *C. cartwrightianus* individuals based on overall patterns of gene expression. At a significance threshold of P = 0.001, we obtained 4502 genes expressed significantly different in at least one sample. The heatmap clearly showed higher transcriptome diversity in the individuals of *C. cartwrightianus* than in saffron (Fig. 4.2).

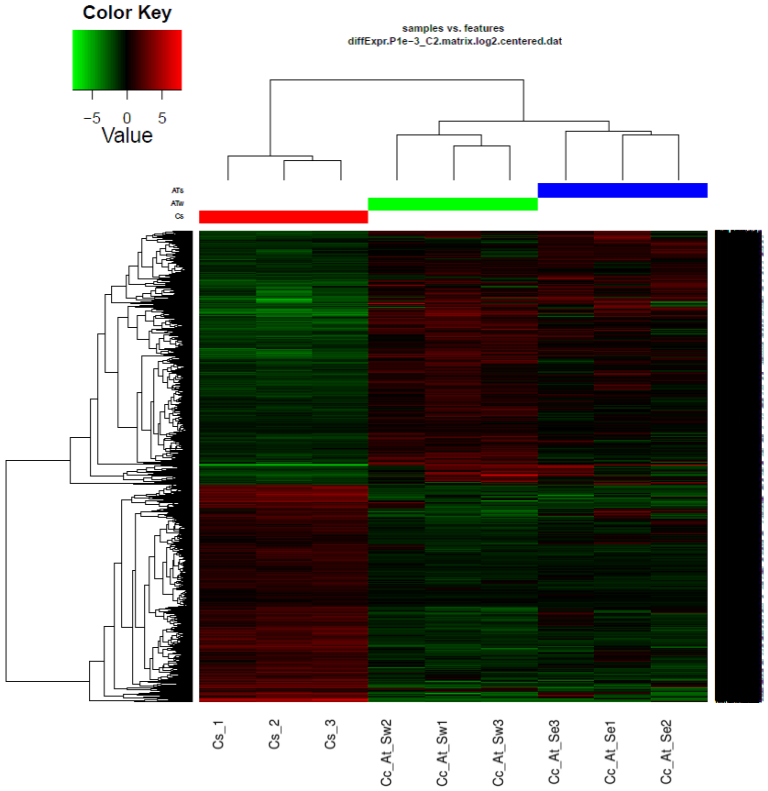


Figure 4.2. Hierarchical clustering of saffron (*C. sativus*) and *C. cartwrightianus* individuals from southeast and southwest Attica based on overall patterns of gene expression. On top is the clustering of individuals given (red = *C.s.*, green = *C.c.* southeast Attica, blue = *C.c.* southwest Attica), to the left the clustering of genes sorted according to their expression differences. Heatmap colours are red for up-regulated and green for down-regulated genes in the dataset. Transcriptome diversity is higher in the individuals of *C. cartwrightianus* than in saffron.

We also estimated the gene expression pattern of 78 transcripts reported by Jain et al. (2016) to be involved in apocarotenoid biosynthesis of saffron to partition the DE genes

into clusters with similar expression patterns and to identify the number of significantly (P-value ≤ 0.01 and at least two-fold change) up- or down-regulated transcripts in each sample compared to all other. We found seven genes out of 78 (~10%) genes related to apocarotenoid biosynthesis that were expressed differentially between samples. These genes are orthologues of zeaxanthin epoxidase (CsTc006236), UDP-glucosyltransferase (CsTc020060), carotenoid isomerase (CsTc091265), lycopene beta-cyclase (CsTc018497), phytoene synthase (CsTc009491), carotene beta-hydroxylase (CsTc000418) and nine-cis-epoxy carotenoid dioxygenase (CsTc035409) (Fig. 4.3).

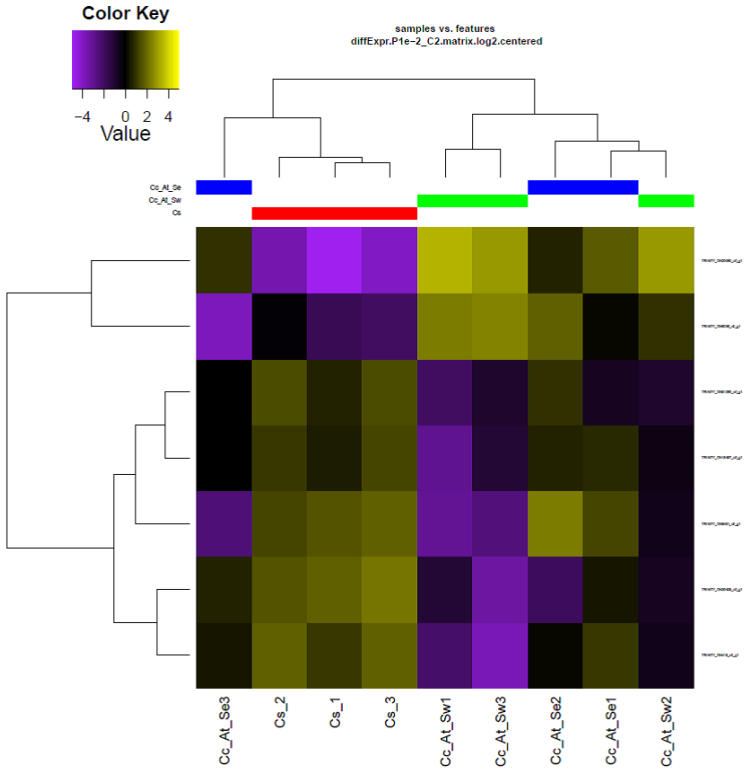


Figure 4.3. Hierarchical clustering of seven differentially expressed genes relevant to the apocarotenoid biosynthesis pathway of saffron reported by Jain et al. (2016) at the threshold P-value of 0.01. A *C. cartwrightianus* individual from southeast of Attica (blue) shows similar expression values as in saffron for these genes. Heatmap colours are yellow for up-regulated and purple for down-regulated genes in the dataset.

Discussion

We performed a *de novo* transcriptome assembly of *C. sativus* and it was used as a reference for mapping the reads of *C. cartwrightianus* and *C. sativus* individuals for clustering of these individuals based on overall patterns of their gene expression. We here

found higher expression diversity in the individuals of *C. cartwrightianus* than in saffron as expected, due to asexual propagation of saffron and the allelic diversity caused by the outcrossing nature of *C. cartwrightianus*.

We further estimated the expression pattern of *C. sativus* and *C. cartwrightianus* individuals in terms of 78 genes of putative high importance for the apocarotenoid biosynthetic pathway of *C. sativus*. At a threshold of significance of P-value of 0.01, we found seven genes showing similar expression values in one *C. cartwrightianus* individual from southeast of Attica as in saffron. However, they are differentially regulated in other individuals. Among these genes, there is only one gene (CsTc020060) down-regulated in all *C. sativus* individuals while up-regulated in all *C. cartwrightianus* individuals (Fig. 4.3). This gene was reported to be related to UDP-glucosyltransferase, which is an important enzyme involved in the last step of the apocarotenoid pathway where crocetin is converted to crocin and the pigments in the saffron stigmas are accumulating. This could cause an important difference in metabolite accumulation between wild and cultivated saffron.

Although we here could not find a single *C. cartwrightianus* individual with the completely identical gene expression pattern of *C. sativus*, we found one Attic *C. cartwrightianus* individuals showing a more similar expression pattern of the genes involved in apocarotenoid biosynthesis to saffron. Therefore, to arrive at a better understanding of differences in the pigment accumulation in the stigmas of saffron and to identify *C. cartwrightianus* individuals with higher content of colourful compounds, it is necessary to perform in the future parallel analyses of important secondary compounds and stigma transcriptome levels (RNA-seq) on several saffron and *C. cartwrightianus* individuals with low and high content of the redish pigments.

From the result we obtained in this study a hypothesis derives, that has to be tested in the future to understand what makes saffron exceptional, and to finally unravel saffron evolution. The hypothesis is that triploidy and the sterility it caused was a way to safeguard a very favourable allele composition (regarding aroma and colour of the styles it causes) from being broken up by recombination after it occurred in *C. cartwrightianus*. In case this hypothesis holds true, it would be expected to find all the relevant characters/compounds already on the diploid level, although maybe only rarely. Selection for genotypes with different colour and aroma intensity then could be a way to improve diversity within saffron, either with or without a final step towards triploidisation and sterility to fix new superior gene combinations in certain lineages.

Conclusion

Transcriptome analyses of *C. sativus* and the Attic *C. cartwrightianus* individuals show that there is higher expression diversity in *C. cartwrightianus* in comparison to the saffron crocus but also that Attica harbours *C. cartwrightianus* individuals with expression patterns similar to saffron regarding the genes involved in the apocarotenoid biosynthesis. However, not all similar expression patterns occur necessarily within single individuals. From the results obtained in this study we hypothesise, that triploidy and the sterility it caused was a way to safeguard a very favourable diploid allele composition (regarding aroma and colour of the styles it influences) from being broken up by recombination after it occurred in *C. cartwrightianus*.

Abstract

Crocus sativus, is the economically most important species within *Crocus*. Its long stigmas are used to produce saffron, the worldwide most expensive spice. *Crocus sativus* is a male-sterile triploid that is propagated only vegetatively by means of the bulb-like corms. The vegetative cultivation offers advantages in maintaining the genetic characteristics of the plants, but impedes any breeding advances by genetic improvement through crossing of different lineages. An alternative approach was to infer the closest relative(s) of *C. sativus* and trying to understand how the triploid species evolved, which then could open up the possibility to recreate saffron and thus broaden the genetic and phenotypic basis of cultivated lineages. To trace the evolution of triploid saffron, we performed a nested approach from phylogeny of the series through the detection of parents or closest relatives and their distribution area, to the identification and comparison of genes involved in important metabolite synthesis like apocarotenoids.

To obtain a resolved phylogeny of the series, we firstly analysed two chloroplast (*trn*L-F, *mat*K-*trn*K) and three nuclear (TOPO6, nrDNA ETS and ITS) markers in 53 individuals belonging to all taxa of series *Crocus*. All loci were amplified and directly Sanger sequenced. We also cloned amplicons where direct sequencing provided no clear sequences. The haplotype network using chloroplast markers revealed that all *C. cartwrightianus* samples studied here share the same chloroplast haplotype with *C. sativus*. Aligning the five concatenated loci resulted in an alignment of 3045 bp length and revealed that only this species contributed genetic material to the triploid. Our results also showed *C. hadriaticus* and the former *C. pallasii* subspecies to be polyphyletic. The latter deserve elevating the subspecies to species rank, while for *C. hadriaticus* a detailed study of species boundaries seems necessary.

As a second step, we performed phylogenetic analyses of five nuclear loci, screening PCR amplicon sizes of one chloroplast locus and genome-wide DNA polymorphisms of *C. cartwrightianus* individuals collected from its entire distribution area in Greece, together with *C. sativus* and *C. oreocreticus* as the closest relative of both species, to arrive at an understanding of the origin of saffron. Our results indicate that saffron is genetically most similar to the Attic *C. cartwrightianus* populations. Furthermore, the data obtained from the phylogenetic tree and population structure analyses based on genome-wide DNA polymorphisms indicate that there is a high degree of admixture between and within *C.*

cartwrightianus populations from different Greek locations especially the ones belonging to the central Aegean islands. We assume *C. cartwrightianus* originated in the central Aegean. In contrast to *C. cartwrightianus,* we found very low genetic variation in *C. sativus* from across the world, as all analysed individuals show the same genetic pattern and grouped in the same cluster. Based on our data we postulate that *C. sativus* is an autopolyploid that originated only once in southern Attica by combining two different genotypes of *C. cartwrightianus*. Sterility and vegetative propagation prevented afterwards segregation of these traits, resulting in cultivation of a unique, long-term stable clonal lineage.

Finally, I studied gene expression differences in the stigmas of saffron and its closest relative *C. cartwrightianus* by RNA sequencing (RNA-seq) to understand the genetic changes that characterize saffron in comparison to its closest relatives, *C. cartwrightianus* from southwest and southeast of Attica. After sequencing we assembled the reads *de novo* and estimated gene expression patterns. From 4502 genes found to be differentially expressed between the individuals at the P value of 0.001, I found transcriptome diversity in *C. cartwrightianus*, even among individuals from the same region. This is a result that we expected due to the high allelic diversity found in the wild, outbreeding species. Furthermore, I used 78 transcripts attributed to genes involved in the apocarotenoid biosynthesis pathway of saffron in the public transcriptome dataset as template for mapping our reads. Here I found some *C. cartwrightianus* individuals from southern Attica showing more similar gene expression pattern to saffron in terms of putative genes involved in apocarotenoid biosynthesis of saffron than the other studied individuals. This indicates that there might be a chance to obtain new and different saffron types in the future.

Zusammenfassung

Crocus sativus ist die ökonomisch wichtigste Art in der Gattung *Crocus*. Die langen Narben der Art sind das Ausgangsmaterial für die Produktion von Safran, dem weltweit teuersten Gewürz. *Crocus sativus* ist eine männlich sterile triploide Art, die nur durch ihre zwiebelartigen unterirdischen Organe vermehrt werden kann. Dadurch bleiben ihre genetischen Eigenschaften bestehen, gleichzeitig verhindert die vegetative Vermehrung natürlich alle Versuche die Kulturpflanze durch Kreuzungen und Züchtung zu verbessern. Ein anderer Ansatz ist es die nächsten Verwandten von *C. sativus* zu bestimmen und zu verstehen wie die triploide Art entstanden ist. Dieses Wissen kann dann verwendet werden um den Safrankrokus neu zu synthetisieren und so eine breitere genetische Vielfalt für die Kulturpflanze zu erhalten. Um die Evolution von triploidem Safran nachzuvollziehen beschreibe ich hier einen hierarchischen Ansatz von der Analyse der Phylogenie der Serie zu der *C. sativus* gehört, über die Studie potentieller Eltern und ihrer geographischen Vorkommen, hin zur Identifikation wichtiger Gene des Sekundärstoffwechsels und ihr Vergleich zwischen Wild- und Kulturform des Safran.

Um eine aufgelöste Phylogenie der ser. *Crocus* zu erhalten wurden zunächst zwei Abschnitte des Chloroplastengenoms (*trn*L-*trn*F und *mat*K-*trn*K) und drei aus dem Kerngenom (TOPO6, rDNA ITS und ETS) in 53 Individuen sequenziert. Teilweise wurden die PCR-Produkte auch zunächst kloniert, wenn intraindividuelle Polymorphismen vorhanden waren. Die Chloroplastenmarker zeigten, dass *C. sativus* den selben Haplotyp wie *C. cartwrightianus* besitzt. Eine gemeinsame Analyse der fünf Markerregionen (Alignmentlänge: 3045 bp) zeigte, dass wohl nur *C. cartwrightianus* zur Entstehung von *C. sativus* beitrug. Weiterhin ergaben die Analysen, dass es sich bei den Unterarten von *C. pallasii* um eigenständige Arten handelt. Für die innerartliche Variation in *C. hadriaticus* sind weitere detaillierte Studien notwendig um den Status der beschriebenen Unterarten aufklären zu können.

In einer zweiten Analyse wurden fünf Einzelgene aus dem Kerngenom, eine polymorphe Chloroplastenregion sowie genomweite Einzelbasenpolymorphismen analysiert um die zuvor aufgestellte Hypothese zu testen. Auch diese Studie zeigte, dass nur *C. cartwrightianus* der Vorfahre von *C. sativus* ist. Als Region der Entstehung des Safrankrokus konnte Attika, die Region in Griechenland südlich von Athen identifiziert werden. In *C. cartwrightianus* wurde sehr hohe genetische Diversität gefunden, *C. sativus*

hingegen ist innerartlich kaum differenziert. Dies deutet auf eine einmalige Entstehung von *C. sativus* durch die Kombination zweier unterschiedlicher Genotypen von *C. cartwrightianus* hin. Sterilität und vegetative Vermehrung verhinderten danach, dass die typischen Charakteristika von Safran durch Rekombination und Segregation wieder auseinander gerissen wurden.

Im letzten Schritt dieser Arbeit verglich ich die Genexpression in den Griffeln und Naben von *C. sativus* mit zwei verschiedenen *C. cartwrightianus* Typen durch RNA-Sequenzierung. Insgesamt wurden 4502 Gene gefunden, die signifikant unterschiedlich transkribiert wurden. I der Wildart waren klare Unterschiede zwischen den Individuen zu finden, während im Safrankrokus solche Unterschiede relativ klein waren. Dieses Ergebnis war durch die hohe genetische Diversität und das obligate Auskreuzen dieser Art zu erwarten. In 78 Genen des Apokarotin-Metabolismus, die somit für die Entstehung der wichtigen Inhaltsstoffen von Safran entscheidend sind, zeigte sich, dass einige der untersuchten *C. cartwrightianus* Individuen aus dem Süden Attikas deutlich ähnlicher zu *C. sativus* sind als die anderen verglichenen Exemplare. Dies zeigt an, dass es in Zukunft möglich sein kann, durch gezielte Kreuzungen verschiedener *C. cartwrightianus* Genotypen, neue und unterschiedliche Safrantypen zu erzeugen.

References

Abdullaev, F., Espinosa-Aguirre, J. 2004. Biomedical properties of saffron and its potential use in cancer therapy and chemoprevention trials. *Cancer Detection and Prevention* 28, 426-432.

Agayev, Y.M. 2002. New features in karyotype structure and origin of saffron, *Crocus sativus* L. *Cytologia* 67, 245-252.

Agayev, Y.M., Zarifi, E. 2010. Peculiar evolution of saffron (*Crocus sativus* L.): prosperity and decline. *Acta Horticulturae* 850, 29-34.

Ahrazem, O., Trapero, A., Gomez, M.D., Rubio-Moraga, A., Gomez-Gomez, L. 2010. Genomic analysis and gene structure of the plant carotenoid dioxygenase 4 family: a deeper study in *Crocus sativus* and its allies. *Genomics* 96, 239-250.

Alavi-Kia, S.S., Mohammadi, S.A., Aharizad, S., Moghaddam, M. 2008. Analysis of genetic diversity and phylogenetic relationships in *Crocus* genus of Iran using inter retrotransposon amplified polymorphism. *Biotechnology and Biotechnological Equipment* 22, 795-800.

Al-Eisawi, D. 2001. Two new species of Iridaceae, *Crocus naqabensis* and *Romulea petraea* from Jordan. *Arab Gulf Journal of Scientific Research* 19, 167-169.

Alsayied, N.F., Fernandez, J.A., Schwarzacher, T., Heslop-Harrison, J.S. 2015. Diversity and relationships of *Crocus sativus* and its relatives analysed by inter-retroelement amplified polymorphism (IRAP). *Annals of Botany* 116, 359-368.

Aytekin, A., Acikgoz, A.O. 2008. Hormone and microorganism treatments in the cultivation of saffron (Crocus sativusL.) plants. *Molecules* 13, 1135-1147.

Blattner, F.R. 2016. TOPO6: a nuclear single-copy gene for plant phylogenetic inference*. Plant Systematics and Evolution* 302, 239-244.

Bowles, E.A. 1952. A Handbook of *Crocus* and *Colchicum* for Gardeners*. Bodley Head*, London.

Brassac, J., Jakob, S.S., Blattner, F.R. 2012. Progenitor-derivative relationships of *Hordeum* polyploids (Poaceae, Triticeae) Inferred from sequences of TOPO6, a nuclear low-copy gene region. *PLoS One* 7, e33808.

Bray, N.L., Pimentel, H., Melsted, P., Pachter, L. 2016. Near-optimal probabilistic RNA-seq quantification . *Nature Biotechnology* 34, 525-527.

Brighton, C.A., Mathew, B., Marchant, C.J. 1973. Chromosome counts in the genus *Crocus* (Iridaceae). *Kew Bulletin* 28, 451-464.

Brighton, C.A. 1977. Cytology of *Crocus sativus* and its allies (Iridaceae). *Plant Systematics and Evolution* 128, 137-157.

Chryssanthi, D.G., Lamari, F.N., Iatrou, G., Pylara, A., Karamanos, N.K., Cordopatis, P. 2007. Inhibition of breast cancer cell proliferation by style constituents of different *Crocus* species. *Anticancer Research* 27, 357-362.

Clement, M., Posada, D., Crandall, K.A. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9, 1657-1660.

Dagostino, N., Pizzichini, D., Chiusano, M.L., Giuliano, G. 2007. An EST database from saffron stigmas. *BMC Plant Biology* 7, 53.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9, 772.

De-Juan, J.A., Lopez, C.H., Munoz, R.M., Picornell, M.R. 2009. Yield and yield components of saffron under different cropping systems. *Industrial Crops and Products* 30, 212- 219.

Dolezel, J., Binarova, P., Lucretti, S. 1989. Analysis of nuclear DNA content in plant cells by flow cytometry. *Biologia Plantarum* 31, 113-120.

Douglas, M., Perry, N. 2003. Growing saffron-the world's most expensive spice. *Crop and Food Research*, 20, 1-4.

Eaton, D. 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* 30, 1844-1849.

Eaton, D., Overcast, I. 2016. ipyrad: interactive assembly and analysis of RADseq data sets. Available at: http://ipyrad.readthedocs.io/.

Eickbush, T.H., Eickbush, D.G. 2007. Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics* 175, 477-485.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6, e19379.

Emerson, K.J., Merz, C.R., Catchen, J.M., Hohenlohe, P.A., Cresko, W.A., Bradshaw, W.E., Holzapfel, C.M. 2010. Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences USA* 107, 16196-16200.

Erol, O., Kaya, H.B., Sik, L., Tuna, M., Can, L., Tanyolac, M.B. 2014. The genus *Crocus*, series *Crocus* (Iridaceae) in Turkey and 2 East Aegean islands: a genetic approach. *Turkish Journal of Biology* 38, 48-62.

Erol, E., Harpke, D. & Yildrim, H. 2015. A new *Crocus* L. (Iridaceae) species from SE Turkey, based on morphological and molecular data. *Phytotaxa* 239, 223-232.

Fernandez, J.A. 2004. Biology, biotechnology and biomedicine of saffron. *Recent Research Development in Plant Science*. 2, 127-159.

Fernandez, J.A. 2007. Genetic resources of saffron and allies (*Crocus* spp.) *Acta Horticulturae* 739, 167-189.

Ferrence, S.C., Bendersky, G. 2004. Therapy with saffron and the goddess at Thera. *Perspectives in Biology and Medicine* 47, 199-226.

Fitch, W.M. 1971. Toward defining the course of evolution: Minimal change for a specific tree topology. *Systematic Zoology* 20, 406-416.

Frello, S., Heslop-Harrison, J.S. 2000. Repetitive DNA sequences in *Crocus vernus* Hill (Iridaceae): the genomic organization and distribution of dispersed elements in the genus *Crocus* and its allies. *Genome* 43, 902-909.

Frichot, E., Francois, O. 2015. LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution* 6,925-929.

Frizzi, G., Miranda, M., Pantani, C., Tammaro, F. 2007. Allozyme differentiation in four species of the *Crocus cartwrightianus* group and in cultivated saffron (*Crocus sativus*). *Biochemical Systematics and Ecology* 35, 859-868.

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.

Ghaffari, S.M. 1986. Cytogenetic studies of cultivated *Crocus sativus* (Iridaceae). *Plant Systematics and Evolution* 153, 199-204.

Giaccio, M. 2004. Crocetin from saffron: An active component of an ancient spice. *Critical Reviews in Food Science and Nutrition* 44, 155-172.

Gismondi, A., Fanali, F., Martínez Labarga, J.M., Grilli-Caiola, M., Canini, A. 2013. *Crocus sativus* L. genomics and different barcode applications. *Plant Systematics and Evolution* 299, 1859-1863.

Gombault, R. 1956. Contribution à la connaissance de la flore de la Djézireh syrienne. *Bulletin de la Societe Botanique de France* 103, 460.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., 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.

Gracia, L., Perez-Vidal, C., Gracia-Lopez, C. 2009. Automated cutting system to obtain the stigmas of the saffron Flower. *Biosystems Engineering* 104, 8-17.

Gresta, F., Avola, G., Lombardo, G.M., Siracusa, L., Ruberto, G. 2009. Analysis of flowering, stigmas yield and qualitative traits of saffron (*Crocus sativus* L.) as affected by environmental conditions. *Scientia Horticulturae*, 119, 320-324.

Grilli-Caiola, M. 1999. Reproduction biology in saffron and its allies. in: Negbi, M. (ed.), Saffron. *Harwood Academic Publishers* 31-44.

Grilli-Caiola, M., Di-Somma, D., Lauretti, P. 2001. Comparative study of pollen and pistil in *Crocus sativus* L. (Iridaceae) and allied species. *Annali di Botanica* 1, 93-103.

Grilli-Caiola, M., Caputo, P., Zaier, R. 2004. RAPD analysis in *Crocus sativus* L. accessions and related *Crocus* species. *Biologia Plantarum* 48, 375-380.

Grilli-Caiola, M., Canini, A. 2010. Looking for saffron's (*Crocus sativus* L.) parents. in: Husaini, A.M. (ed.), Saffron. *Functional Plant Science and Biotechnology* 4, 1-14.

Gruber, B., Adamack, A.T. 2015. Landgenreport: a new R function to simplify landscape genetic analysis using resistance surface layers. *Molecular Ecology Resources* 15, 1172-1178.

Harpke, D., Peterson, A. 2008a. 5.8S motifs for the identification of pseudogenic ITS regions. *Botany* 86, 300-305.

Harpke, D., Peterson, A. 2008b. Extensive 5.8S nrDNA polymorphism in *Mammillaria* (Cactaceae) with special reference to the identification of pseudogenic ITS regions. *Journal of Plant Research* 121, 261-270.

Harpke, D., Meng, S., Kerndorff, H., Rutten, T., Blattner, F.R. 2013. Phylogeny of *Crocus* (Iridaceae) based on one chloroplast and two nuclear loci: ancient hybridization and chromosome number evolution. *Molecular Phylogenetics and Evolution* 66, 617-627.

Harpke, D., Peruzzi, L., Kerndorff, H., Karamplianis, T., Constantinidis, T., Ranđelović, V., Ranđelović, N., Jušković, M., Pasche, E., Blattner, F.R. 2014. Phylogeny, geographic

distribution and new taxonomic circumscription of the *Crocus reticulatus* species group (Iridaceae). *Turkish Journal of Botany* 38, 1182-1198.

Harpke, D., Carta, A., Tomović, G., Ranđelović, V., Ranđelović, N., Blattner, F.R., Peruzzi, L. 2015. Phylogeny, karyotype evolution and taxonomy of *Crocus* series *Verni* (Iridaceae). *Plant Systematics and Evolution* 301, 309-325.

Jain, M., Sricastava, P.L., Verma, M., Ghangal, R., Garg, R. 2016. *De novo* transcriptome assembly and comprehensive expression profiling in *Crocus sativus* to gain insights into apocarotenoid biosynthesis. *Scientific Report* 6, 22456.

Jakob, S.S., Blattner, F.R. 2006. A chloroplast genealogy of *Hordeum* (Poaceae): long-term persisting haplotypes, incomplete lineage sorting, regional extinction, and the consequences for phylogenetic inference. *Molecular Biology and Evolution* 23, 1602-1612.

Karamplianis, T., Tsiftsis, S., Constantinidis, T. 2013. The genus *Crocus* (Iridaceae) in Greece: some noteworthy floristic records and karyotypes. *Phytologia Balcanica* 19, 53-66.

Kerndorff, H. 1988. Observations on *Crocus* (Iridaceae) in Jordan with special reference to *Crocus moabiticus*. *Herbertia* 44, 33-53.

Kerndorff, H., Pasche, E. 2004. Two new taxa of the *Crocus biflorus* aggregate (Liliiflorae, Iridaceae) from Turkey. *Linzer Biologische Beiträge* 36, 5-10.

Langmead, B., Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9, 357-359.

Lambeck, K. 1996. Sea-level change and shore-line evolution in Aegean Greece since Upper Palaeolithic time. *Antiquity* 70, 588-611.

Larsen, B., Orabi, J., Pedersen, C., Ørgaard, M. 2015. Large intraspecific genetic variation within the Saffron-Crocus group (*Crocus* L., Series *Crocus*; Iridaceae). *Plant Systematics and Evolution* 301, 425-437.

Li, W. 1997. Molecular Evolution. Sinauer Associates, Sunderland, Massachusetts.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 genome project data processing subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078-2079.

Madan, C.I., Kapur, B.M., Gupta, U.S. 1966. Saffron. *Economic Botany* 20, 377-385.

Makri, O.E., Ferlemi, A.V., Lamari, F.N., Georgakopoulos, C.D. 2013. Saffron administration prevents selenite-induced cataractogenesis. *Molecular Vision* 19, 1188-1197.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10-12.

Mathew, B. 1977. *Crocus sativus* and its allies (Iridaceae). *Plant Systematics and Evolution* 128, 89-103.

Mathew, B. 1982. The Crocus. A Revision of the Genus *Crocus* (Iridaceae). *Timber Press*, Portland.

Mathew, B. 2000. *Crocus hadriaticus* subsp. *parnassicus. Bulb Newsletter* 32, 7.

Mathew, B., Petersen, G., Seberg, O. 2009. A reassessment of *Crocus* based on molecular analysis. *Plantsman* 8, 50-57.

McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., Brumfield, R.T. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution* 66, 526-538.

Namayandeh, A., Nemati, Z., Kamelmanesh, M.M., Mokhtari, M., Mardi, M. 2013. Genetic relationships among species of Iranian *Crocus* (*Crocus* spp.). *Crop Breeding Journal* 3, 61-67.

Nemati. Z., Zeinalabedini, M., Mardi, M., Pirseyediand, S.M., Marashi, S.H., Nekoui, S.M.K. 2012. Isolation and characterization of a first set of polymorphic microsatellite markers in saffron, *Crocus sativus* (Iridaceae). *American Journal of Botany* 99, e340-e343.

Nemati, Z., Mardi, M., Majidian, P., Zeinalabedini, M., Pirseyedi, S.M., Bahadori, M. 2014. Saffron (*Crocus sativus* L.), a monomorphic or polymorphic species? *Spanish Journal of Agricultural Research* 12, 753-762.

Negbi, M. 1989. Theophrastus on geophytes. *Botanical Journal of the Linnean Society* 100, 15-43.

Petersen, G., Seberg, O., Thorsøe, S., Jørgensen, T., Mathew, B. 2008. A phylogeny of the genus *Crocus* (Iridaceae) based on sequence data from five plastid regions. *Taxon* 57, 487-499.

Peterson, A., Harpke, D., Peruzzi, L., Levichev, I.G., Tison, J., Peterson, J. 2009. Hybridization drives speciation in *Gagea* (Liliaceae). *Plant Systematics and Evolution* 278, 133-148.

Peterson, A., Levichev, I.G., Peterson, J., Harpke, D., Schnittler, M. 2011. New insights into the phylogeny and taxonomy of Chinese species of *Gagea* (Liliaceae) – speciation through hybridization. *Organisms Diversity and Evolution* 11, 387-407.

Poma, A., Fontecchio, G., Carlucci, G., Chichiriccò, G. 2012. Anti-inflammatory properties of drugs from saffron crocus. *Anti-Inflammatory and Anti-Allergy Agents Medicinal Chemistry* 11, 37-51.

Posada, D., Crandall, K.A. 2001. Intraspecific gene genealogies: trees grafting into networks. *Ecology and Evolution* 16, 37-45.

Robinson, M.D., McCarthy, D.J., Smyth, G.K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139-140.

Rambaut, A. 2009. FigTree version 1.3.1. http://tree.bio.ed.ac.uk/soft ware/figtree/

Rios, J.L., Recio, M.C., Giner, R.M., Manez, S. 1996. An update review of saffron and its active constituents. *Phytotherapy Research* 10, 189-193.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61, 539-542.

Rukšans, J 2017. The World of Crocuses. *Latvian Academy of Sciences*, Riga.

Sanei, M., Rahimyan, H., Agayev, Y., Soheilivand, S. 2007. New cytotype of *Crocus pallasii* subsp. *haussknechtii* from west of Iran. *Acta Horticulturae* 739, 107-111.

Schneider, I., Kerndorff, H., Pasche, E. 2013. Chromosome numbers of Turkish *Crocus* (Liliiflorae, Iridaceae) and their geographical distribution. *Feddes Repertorium* 123, 73-79.

Seberg, O., Petersen, G. 2009. How many loci does it take to DNA barcode a crocus? *PLoS One* 4, e4598.

Siracusa, L., Gresta, F., Avola, G., Albertini, E., Raggi, L., Marconi, G., Ruberto, G. 2013. Agronomic, chemical and genetic variability of saffron (*Crocus sativus* L.) of different origin by LC-UV–vis-DAD and AFLP analyses. *Genetic Resources and Crop Evolution* 60, 711-721.

Souret, F., Weathers, P. 2000. *Crocus sativus* L. (saffron): cultivation, in vitro culture, secondary metabolite production and phytopharmacognosy, *Journal of Herbs, Spices & Medicinal Plants* 6, 99-116.

Swofford, D.L. 2002. PAUP*: Phylogenetic Analysis Using Parsimony (* and other Methods), version 4.0. *Sinauer Associates,* Sunderland.

Tison, J., Peterson, A., Harpke, D., Peruzzi, L. 2013. Reticulate evolution of the critical Mediterranean *Gagea* sect. *Didymobulbos* (Liliaceae) and its taxonomic implications. *Plant Systematics and Evolution* 299, 413-438.

Tsaftaris, A., Pasentsis, K., Makris, A., Darzentas, N., Polidoros, A., Kalivas, A., Argiriou, A. 2011. The study of the E-class SEPALLATA3-like MADS-box genes in wild-type and mutant flowers of cultivated saffron crocus (*Crocus sativus* L.) and its putative progenitors. *Plant Physiology* 168, 1675-1684.

Wendler, N., Mascher, M., Nöh, C., Himmelbach, A., Scholz, U., Ruge-Wehling, B., Stein, N. 2014. Unlocking the secondary gene-pool of barley with next-generation sequencing. *Plant Biotechnology Journal* 12, 1122-1131.

Wheat, C.W. 2010. Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing. *Genetica* 138, 433-451.

Wickham, H. 2009. ggplot2: Elegant graphics for data analysis. useR. *Springer-Verlag*, New York.

Winterhalter, P., Straubinger, M. 2000. Saffron–renewed interest in an ancient spice. *Food Reviews International*. 16, 39-59.

Wissemann, V., Ritz, C.M. 2005. The genus *Rosa* (Rosoideae, Rosaceae) revisited: molecular analysis of nrITS-1 and *atp*B-*rbc*L intergenic spacer (IGS) versus conventional taxonomy. *Botanical Journal of the Linnean Society* 147, 275-290.

Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., Weir, B.S. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28, 3326-3328.

Zimmer, E.A., Wen, J. 2013. Reprint of: Using nuclear gene data for plant phylogenetics: progress and prospects. *Molecular Phylogenetics and Evolution* 66, 539-550.

Zubor, A.A, Suranyi, G., Gyori, Z., Borbely, G., Prokisch, J. 2004. Molecular biological approach of the systematics of *Crocus sativus* L. and its allies. *Acta Horticulturae* 650, 85-93.

Acknowledgement

First of all, I would like to thank my great supervisor Dr. Frank Blattner for the continuous support of my Ph.D. study and related research, for his patience, motivation, immense knowledge, and great efforts to explain things clearly. His guidance helped me in all the time of research and writing of this thesis, often suggesting ways to go on with my work.

I am deeply indebted to Dr. Dörte Harpke for her invaluable help, the stimulating discussions and all the fun we have had in the last four years. Without her and her knowledge of *Crocus* species, this work could not be accomplished.

I am hugely appreciative to Dr. Helmut Kerndorff, especially for sharing his taxonomic expertise about crocuses so willingly, and for helping me in the collecting plant materials.

I am also thankful to Prof. Dr. Martin Röser at the Martin-Luther-Universität Halle-Wittenberg for agreeing to supervise my thesis.

I would like to thank all the members of the ETX working group for relaxed and interesting working atmosphere. My friends and colleagues Dr. Jonathan Brassac, Dr. Nadine Bernhardt, Dr. Simon Pfanzelt and all the others who were visitors during these almost four years.

I would like also to thank the precious help in the lab from Petra Oswald and Birgit Kraenzlin and Christina Koch, as well as in the greenhouse where Ina Faustman took care of my plants.

My thanks also go to the PhD student Board of the IPK for the organization of workshops, study trips, conferences and other events.

I am grateful to all the people who made my time at IPK enjoyable between the events and parties at the Club and the various travels. I'm thinking to Dörte Harpke, Leila Fazlikhani, Kamatchi Ulagappan, Daniela Impe, Milica Milcovic, Razieh Rahmati, Narges Riahi, Somayeh Sardouei, Solmaz Khosravi, Christian Hertig and probably many others.

Special thanks to my great family: parents and sisters for supporting me spiritually throughout my life. Your prayer for me was what sustained me thus far.

Last but not the least; I would like to thank my beloved husband, Ramin Tahmasebi, for his unbelievable support and encouragement throughout this study and my life in general.

I would like to dedicate this thesis to the most important people in my life; my parents and my husband.

Appendix

Table A2.1.

Studied plant materials

Taxon **Series *Crocus* L.**	Origin and sample ID[1]	Herbarium, ID
C. asumaniae B.Mathew & T.Baytop	Turkey, Antalya (cultivated); cr1366	GAT, 47271
	Turkey, Antalya; cr2976, cr2977	ISTF, 41072
C. baytopiorum B.Mathew	Turkey, Denizili (cultivated); cr2070-cr2072	GAT, 7187
C. cartwrightianus Herb.	Greece, Crete; cr1909, 2663*	GAT, HKEP1517
	Greece, Crete; cr1936, cr3269*	GAT, HKEP1525,
	Greece, Syros; cr2635	GAT, HKEP1443
	Greece, Paros; cr1265, cr2783*	GAT, HKEP1445
	Greece, Anafi; cr2938, cr2939*	GAT, HKEP1601
	Greece, Astypalaia; cr2821	GAT, HKEP1609
C. dispathaceus Bowles	Turkey, Mersin; cr2966, cr2999	ISTF, 40663
	Turkey, Mersin; cr2970, cr2971	ISTF, 40664
C. hadriaticus Herb.	Greece, Pindus (cultivated); cr2420, cr2421	GAT, 47247
	Greece, Peloponnese; cr2644-cr2647	GAT, 7220
C. haussknechtii (Boiss. & Reut. ex Maw) Boiss.	Iran, Sanandaj (cultivated); cr1567-cr1569	–, Kv140
C. kofudagensis Rukšāns	Turkey, Antalya (cultivated); cr3335	GAT, JJJ-024
C. macedonicus Rukšāns	Greece, Macedonia; cr1344, cr1346	GAT, 20314
C. mathewii Kernd. & Pasche	Turkey, Antalya; cr2984	ISTF, 40183
C. moabiticus Bornm.	Jordan (cultivated); Rik11; cr1717, cr1719	–, Rik11
C. naqabensis Al-Eisawi & Kiswani	Jordan, Ma'an Governorate (cultivated); cr2108, cr2109	GAT
C. oreocreticus B.L.Burtt	Greece, Crete; cr1373, cr1374	GAT, KR23

	Greece, Crete; cr2674, cr2579*, cr2675*	GAT, HKEP1585
C. pallasii Goldb.	Ukraine, Crimea (cultivated); cr1575-cr1577	GAT, 47264
C. pallasii Goldb.	Macedonia, Kumanovo; cr795-cr797	Niš, s.n.
C. sativus L.	Iran (cultivated); cr1573	GAT
	Spain (cultivated); cr3330	GAT, 163
	France (cultivated); cr3327	GAT, 160
C. thomasii Ten.	Croatia, Dalmatia;, cr1350	–, IS0501
	Italy, Apulia; cr527	GAT, HKEP1326
	Italy, Apulia; cr3091-cr3093	GAT, HKEP1614
C. turcicus (B.Mathew) Rukšāns	Turkey, Gaziantep; cr2970, cr2971	ISTF, 40665
Outgroups		
C. cf. *tauri* Maw	Turkey, Upper Euphrates; cr38	GAT, 7143
C. vernus Hill	Switzerland, Airolo; cr305	GAT, 30292

[1]Accessions indicated as 'cultivated' were regrown from seeds or corms collected in the wild or obtained through commercial sources in case of *C. sativus*. * DNA substitute for GBS analyses

Table A2.2.

Characteristics of the included data partitions

Character partition	**Alignment length**	**Variable sites**	**Parsimony informative sites[1]**	**MP tree number[2]**	**Consis-tency index**	**Retention index**	**Model of DNA evolution**
nrDNA ETS + ITS	1107	234	125	9	0.824	0.899	GTR+Γ
TOPO6-11	744	100	60	165	0.931	0.952	HKY
Chloroplast *trn*L-*trn*F + *mat*K-*trn*K	1191	50	17	50,000	1.0	1.0	F81
Combined	3042	384	208	1072	0.834	0.899	TIM+I+Γ

[1]The combined dataset has more informative sites as the sum of individual data partitions, as the number of analyzed taxa is higher due to multiple haplotypes occurring within some individuals; [2]Maximum number of trees retained set to 50,000

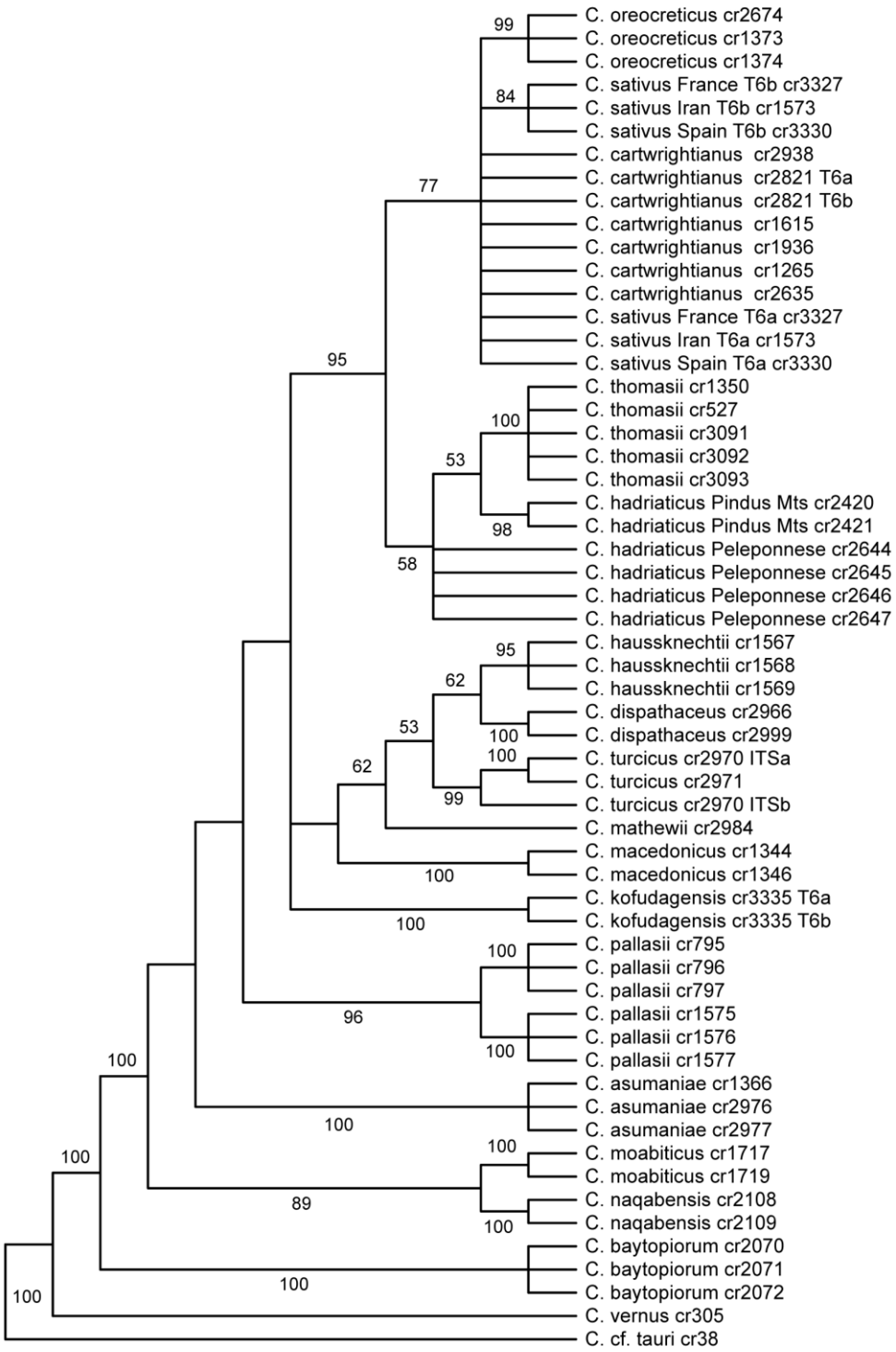


Figure A2.1. Strict consensus MP tree of the combined sequences of nrDNA external (ETS) and internal transcribed spacer (ITS), the single-copy locus TOPO6-11, and the chloroplast regions *trn*L-*trn*F and *mat*K-*trn*K of all series *Crocus* and two outgroup taxa.

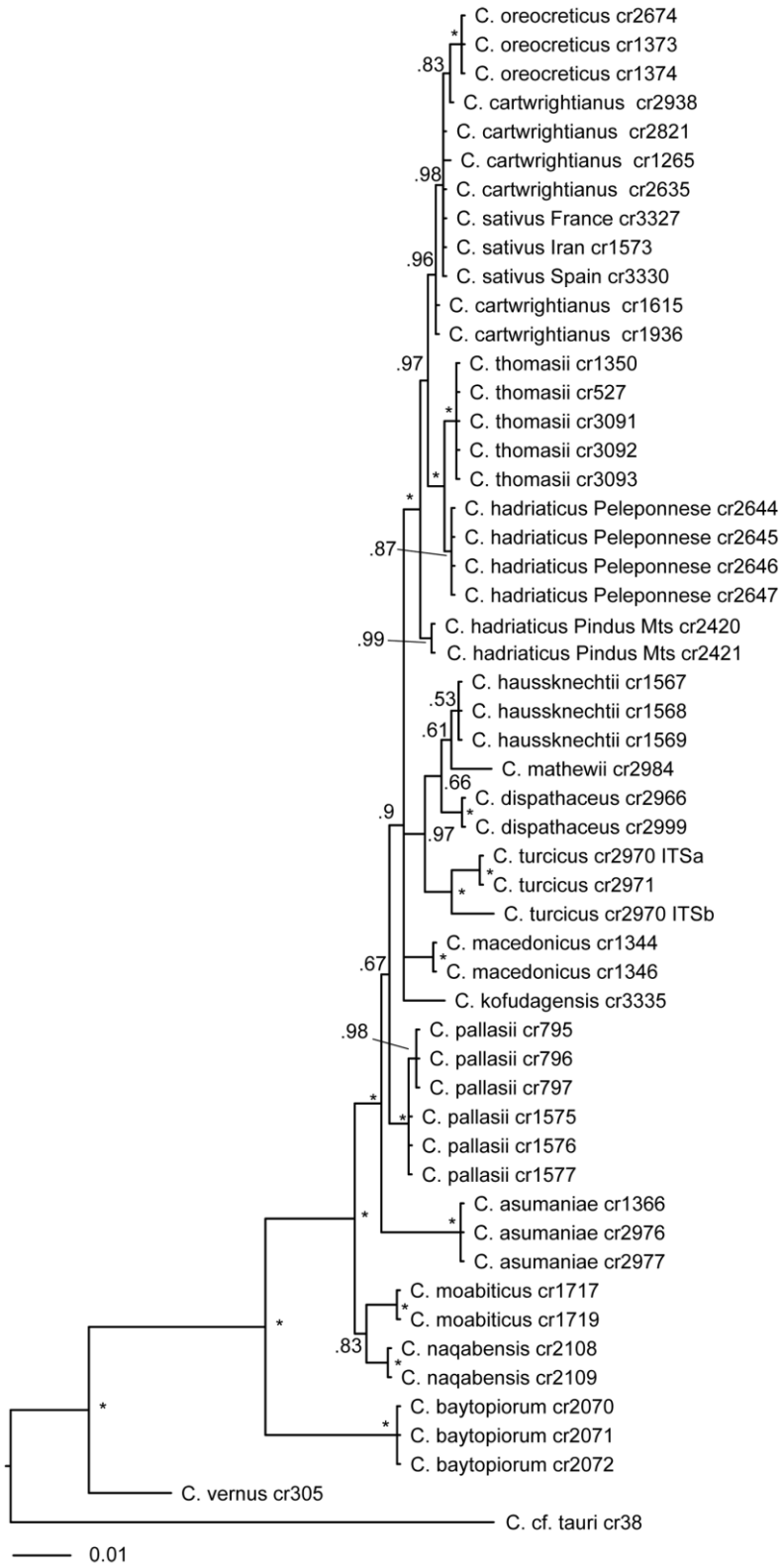


Figure A2.2. BI tree of the nrDNA external (ETS) and internal transcribed spacer (ITS). Different ribotypes found within single individuals detected by cloning and sequencing of PCR amplicons are depicted by ITSa-b.

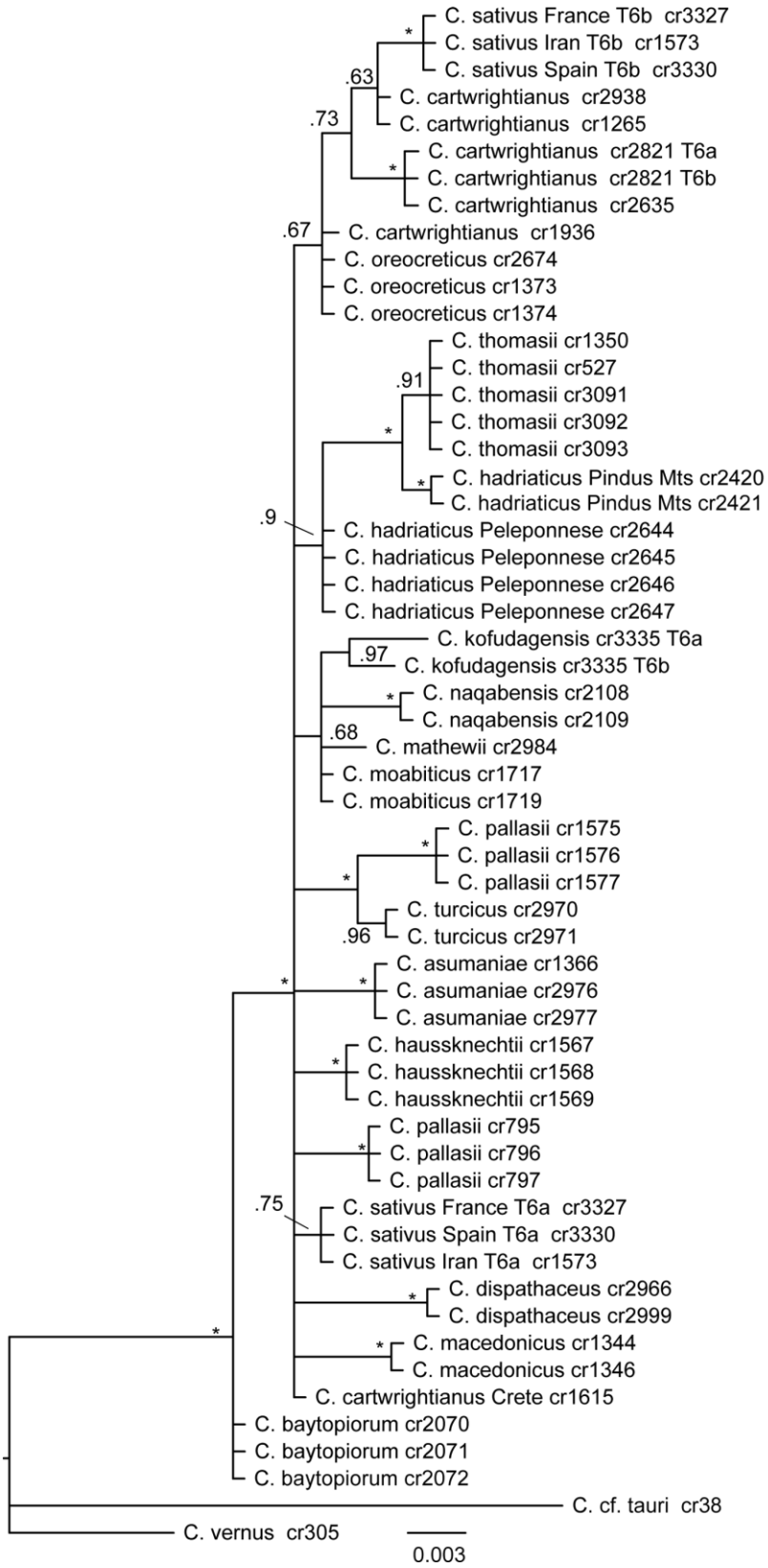


Figure A2.3. BI tree of the TOPO6-11 region. Different haplotypes found within single individuals detected by cloning and sequencing of PCR amplicons are depicted by T6a-b.

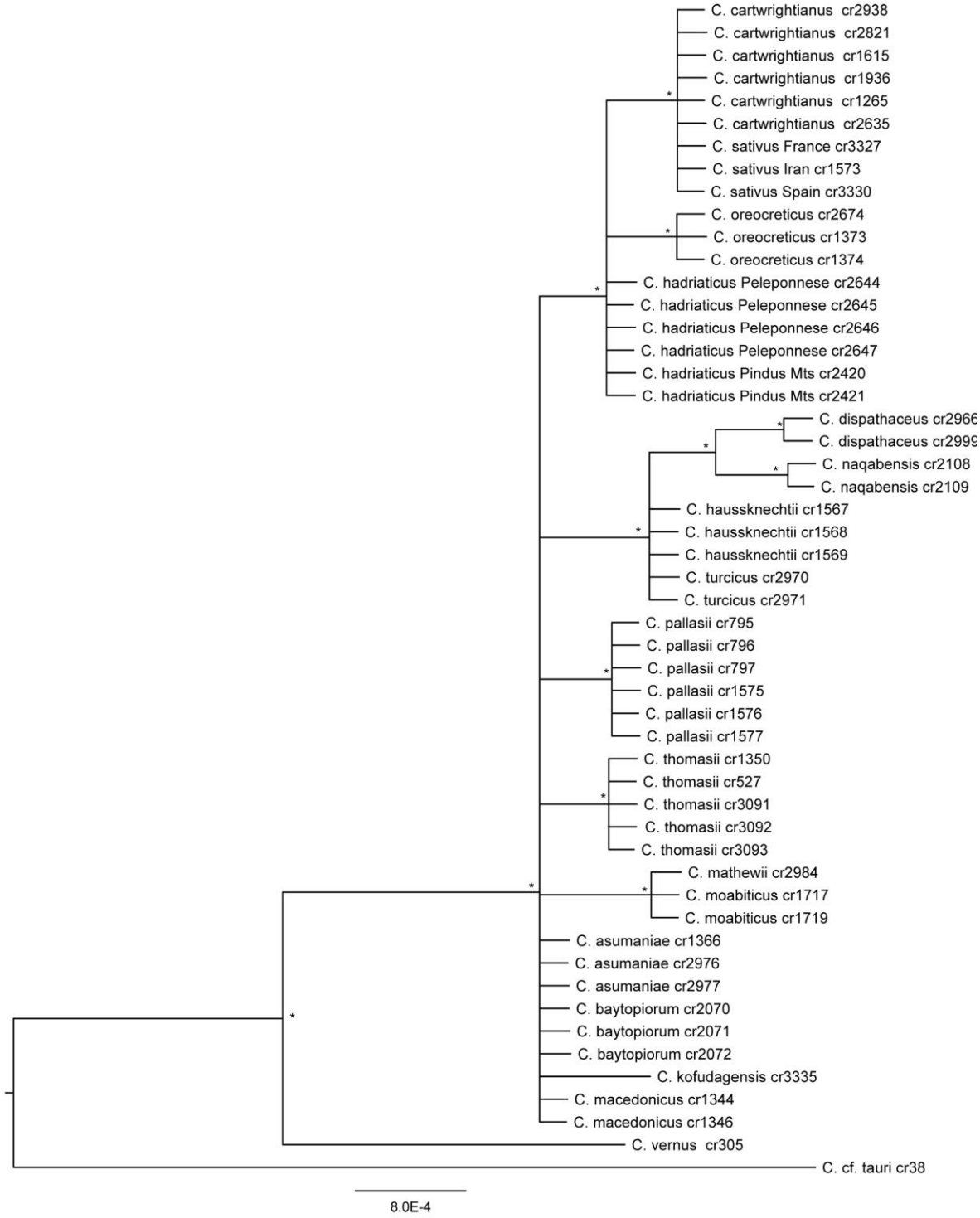


Figure A2.4. BI tree of the combined chloroplast loci (*trn*L-F, *mat*K-*trn*K).

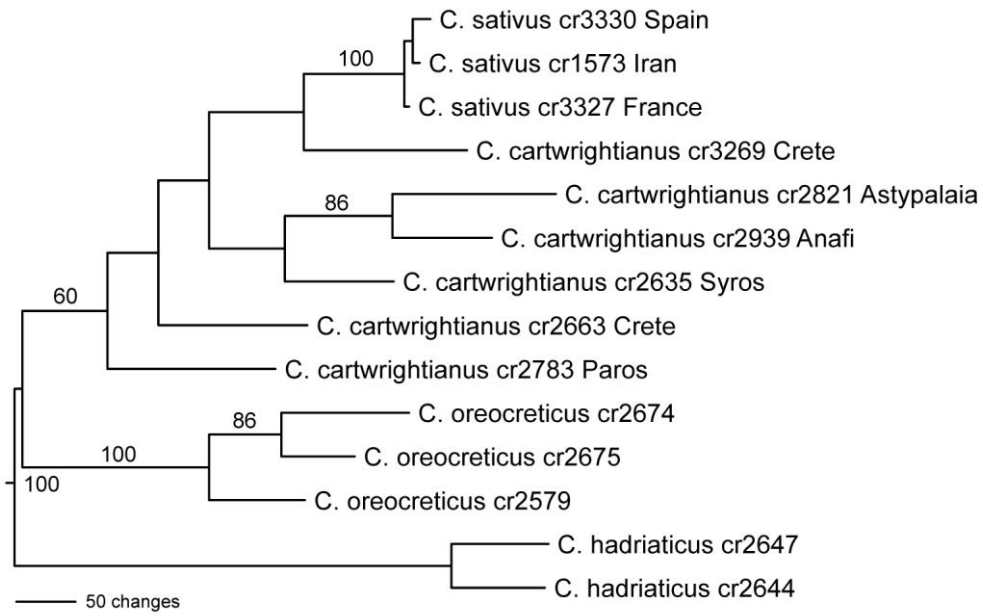


Figure A2.5. Single most parsimonious tree based on the sequences obtained through a genotyping-by-sequencing analysis of the closest relatives of saffron, *C. sativus.* Numbers along branches provide bootstrap values ≥50%.

Table A3.1

Studied plant materials used for single-copy loci analyses

Taxon	Origin and sample ID[1]	Herbarium, ID
C. cartwrightianus Herb.	Crete; 3269	GAT, HKEP1525
	Tinos; 3249	GAT, ZNTIN
	Paros; 2667	GAT, HKEP1445
	Kea; 3895	GAT,KEA1
	Kea; 3899	GAT,KEA4
	Evia; 3317	GAT, HKEP1664
	Attica;3890, 3893	GAT, DH1802
	Attica; 3235	GAT, HKEP1606
	Attica; 3301	GAT, HKEP1592b
	Attica; 3293, 3390	GAT, NB_2017-002A
C. sativus L.	France (cultivated); cr3327	GAT,160
C. thomasii Ten.	Italy, Apulia; cr3091	GAT, HKEP1614
C. hadriaticus Herb.	Greece, Pindus (cultivated); cr2645	GAT, HKEP1580
C. oreocreticus B.L.Burtt	Greece, Crete; cr2577	GAT, HKEP1585
Outgroup		
C. pallasii Goldb	Macedonia, Kumanovo; cr725	Niš, s.n

Table A3.2

Analyzed genome regions

Locus	Primer name	Primer sequence (5'–3')	Ampli-con size (bp)	Annealing temperature/ elongation time
Nuclear loci				
ORTHO005659, FAD dependent oxidoreductase; intron 5, exon 6, intron 6	FAD_ex5fwd FAD_ex7rwd	ACGTTTGCTCTCATCATCTGGC CCTAGAGCAATCTCAACACCTGC	~700	68 °C/25 sec
ORTHO008517, FAD/NAD(P)-binding oxidoreductase	orcp_f orcp_r	GGCTTCCAAATCTTCCTCACCGC AAAGCCTCTGCAGCTCTTCTCCC	~1100	68 °C/60 sec
ORTHO006661, mismatch repair endonuclease PMS; intron 3	pms_ex3f pms_ex4r	GAAACCAGAACAAAAGATGAGCC TGGGCATATCAACAGGTCTACC	~750	62 °C/30 sec
ORTHO006442, Topoisomerase 6 subunit B; intron 11	croc_Top6_11f croc_Top6_12r	CAGTTTCTTCAACACGAATTT[1] ATGGATTCTAACTATTTGTTGGG[1]	~750	59 °C/25 sec
ORTHO004935, Elongation factor 2A; intron 2, exon 3, intron 3	eIF2A_Ex1f eIF2A_Ex4r	GATGAATCTATTGCTTGCCGAATG[2] GTCATATATTGTAGCCTTCGCAGG[2]	~750	62 °C/30 sec
Chloroplast loci				
*trn*S(GCU) – *trn*G(UCC) IGS	croc_trnSf croc_trnS-Gr	CCTGGTTAAGTAAAGTACTGG TCTCAAGTATTCAAATAAAGC	119/203	54 °C/20 sec

[1]Nemati et al. (2018); [2]Erol et al. (2015)

Table A3.3

Characteristics of the included data partitions

Character partition	Alignment length	Variable sites	Parsimony informative sites[1]	MP tree number[2]	Consis-tency index	Retention index	Model of DNA evolution
TOPO6-11	611	38	9	6	0.898	0.871	K81uf+G
FAD_ex5	620	59	29	9	0.805	0.881	HKY+I
Pms	652	15	4	6	1.938	0.750	HKY+I
Orcp	1011	89	42	315	0.540	0.557	HKY+I+G
eIF2A_Ex1	684	34	27	17	0.795	0.898	HKY+I

Table A3.4

Studied plant materials used for GBS analyses

Taxon	Origin and sample ID[1]	Herbarium, ID
C. cartwrightianus Herb.	Crete; cr2662, cr2663, cr2664, cr3468, cr3472, cr3473	GAT, HKEP1517
	Crete; cr3269, cr3270, cr3466	GAT, HKEP1525
	Crete; cr2679	GAT, P9430
	Anafi; cr2791, cr2792, cr2793, cr2795, cr3905, cr3907	GAT, HKEP1601
	cr2790, cr2939	GAT, HKEP1603
	Syros; cr2635, cr2636, cr2637, cr2638, cr2639, cr3462, cr3463, cr3476, cr3477	GAT, HKEP1443
	Tinos; cr3249, cr3250, cr3251, cr3252, cr3253, cr3254, cr3255, cr3256, cr3257, cr3258	GAT, ZNTIN
	Paros; cr2665, cr2666, cr2667, cr2780, cr2781, cr2782, cr2783, cr3471	GAT, HKEP1445
	Evia; cr3309, cr3310, cr3311, cr3312, cr3313, cr3314, cr3315, cr3316, cr3317	GAT, HKEP1664
	Mykonos; cr3259, cr3260, cr3261, cr3262, cr3263, cr3264, cr3265, cr3266, cr3267, cr3268	GAT, ZNMYK
	Kea; cr3895, cr3896, cr3897, cr3898, cr3908, cr3909, cr3910	GAT, KEA1
	Kea; cr3899, cr3900, cr3901, cr3902, cr3914, cr3915, cr3916	GAT, KEA4
	Astypalaia; cr2817, cr2818, cr2819, cr2820, cr2821 cr2944, cr2945	GAT, HKEP1609
	Attica; cr3282, cr3285	GAT, HKEP1592
	Attica; cr3300, cr3305	GAT, HKEP1592b
	Attica; cr2806, cr3231	GAT, HKEP1605
	Attica; cr3275, cr3279	GAT, HKEP1607
	Attica; cr3294	GAT, NB_2017-002A
	Attica; cr3894	GAT, DH1802
C. sativus L.	Spain; cr3329, cr3330	GAT, 163
	France; cr3326, cr3327, cr3328	GAT, 160
	France; cr3331, cr3332, cr2436	GAT, 164
	Iran; S10	GAT, S
	Germany; cr2441	GAT, G
Outgroups		
C. oreocreticus B.L.Burtt	Greece, Crete; cr2576, cr2577, cr2578, cr2579, cr2674, cr2675, cr3381, cr3383	GAT, HKEP1585

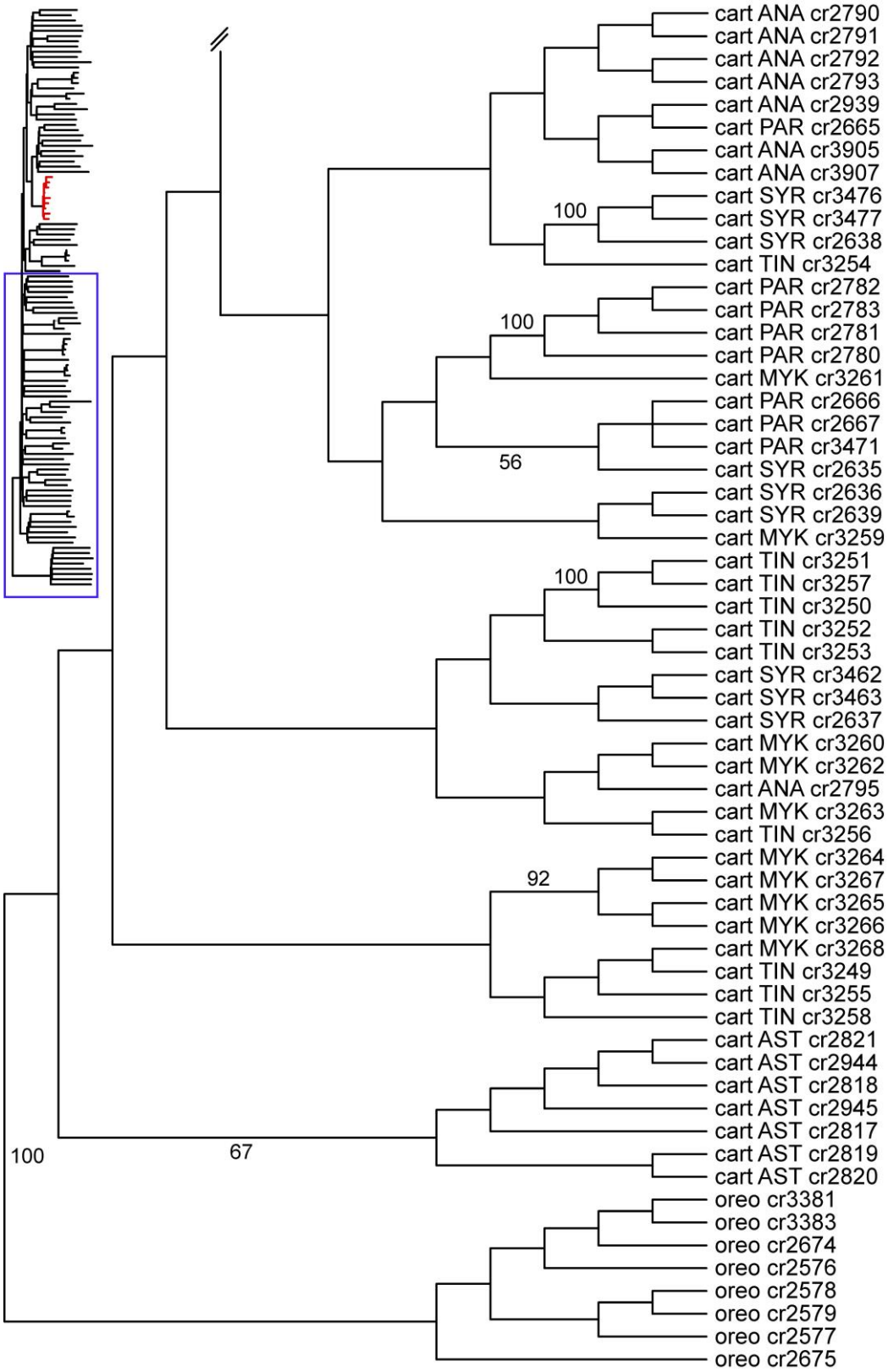
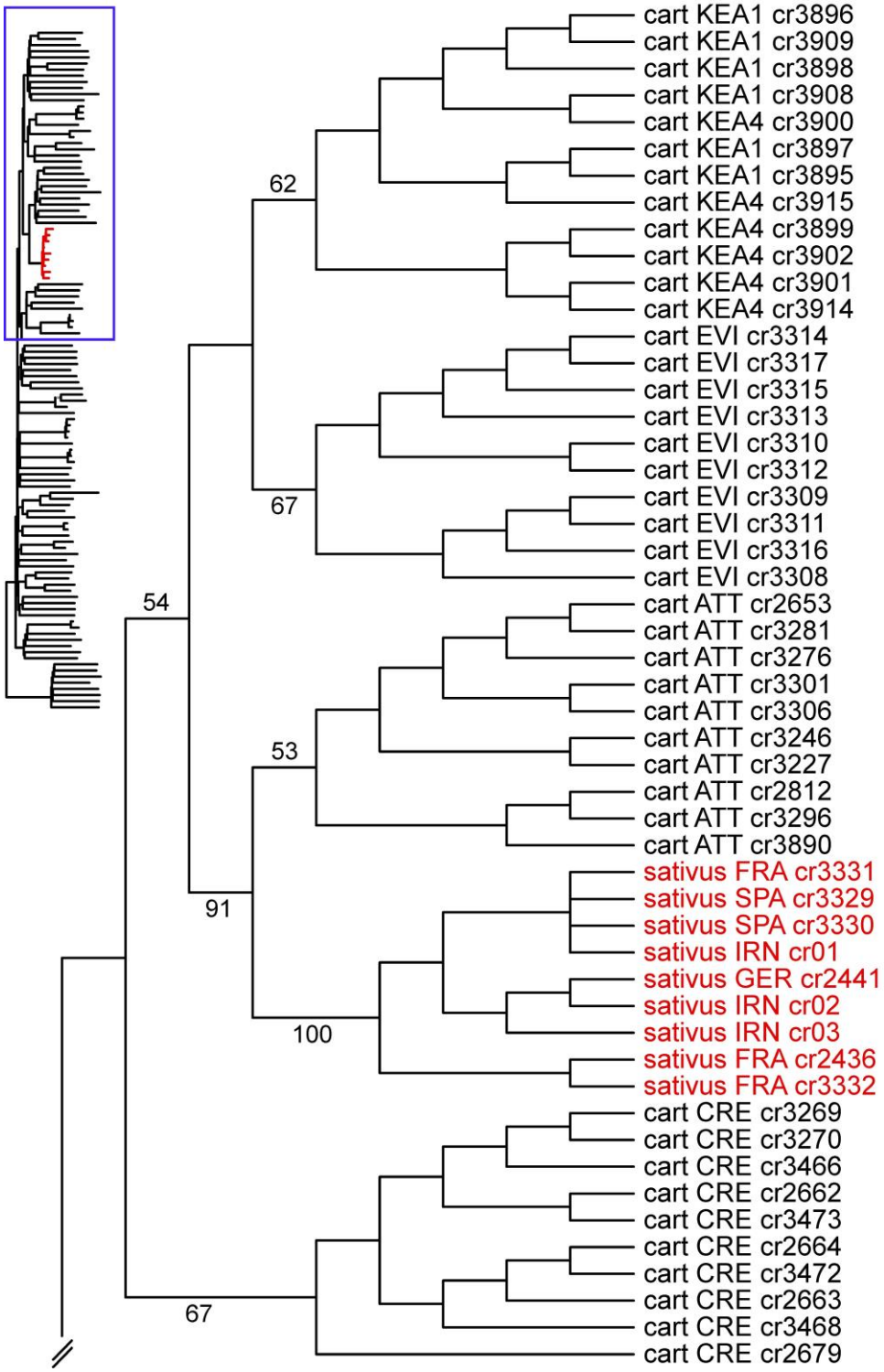


Figure A3.1. Consensus MP tree based on genome-wide DNA data of a genotyping-by-sequencing analysis of 111 individuals. *Crocus oreocreticus* was defined as outgroup taxon. Saffron (*C. sativus*) groups within *C. cartwrightianus* as sister to individuals from Attica.

Figure A3.1. Continued.



Curriculum Vitae & Publications

Personal information

Name: Zahra Nemati

Date of birth: May 23rd 1983

Place of birth: Shiraz, Iran

Nationality: Iranian

Recidency: Hans-Stubbe Str. 10

D-06466 Gatersleben

Germany

Study and professional experience

Since 10/2014

PhD student

at Martin-Luther University of Halle-Wittenberg

and Taxonomy and Evolutionary Biology at the Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben

"The origin of saffron: Progenitors, areas and transcriptomics of economic traits"

(DAAD scholarship)

05/2010-11/2013

Researcher assistant "Study of biodiversity and evolution of plants" at the Azad University of Shiraz, Iran.

10/2006-12/2009

Degree: Master of Science

Study of Agricultural Biotechnology

Biotechnology at the Ferdowsi University of Mashhad

and Genomics at the Agricultural Biotechnology Research Institute of Iran (ABRII), Karaj, Iran

"Isolation and characterization of a first set of polymorphic microsatellite markers in saffron, *Crocus sativus* (Iridaceae)"

09/2001 – 06/2005

Degree: Bachelor of Science

Study of Agronomy and Plant Breeding at the Azad University of Arsanjan, Iran

Extracurricular activities

June 2016 – Present

PhD Student Board member (organisation of workshops, Plant Science Student Conference 2016 & 2018)

Publications

Nemati, Z., Zeinalabedini, M., Mardi, M., Pirseyediand, S.M., Marashi, S.H., Nekoui, S.M.K. 2012. Isolation and characterization of a first set of polymorphic microsatellite markers in saffron, *Crocus sativus* (Iridaceae). *American Journal of Botany* 99, e340-e343.

Namayandeh, A., Nemati, Z., Kamelmanesh, M., Mokhtari, M., Mardi, M. 2013. Genetic relationships among species of Iranian crocus (*Crocus* spp.). *Crop Breeding Journal* 3, 61-67.

Nemati, Z., Zeinalabedini, M., Majidian, P., Eftekharian, A., Kiani, D. 2014. Phylogenetic relationships among Iranian and Spanish date palms (*Phoenix dactylifera* L.) revealed by microsatellite markers. *Journal of Horticultural Science and Biotechnology* 89, 114-120.

Nemati, Z., Mardi, M., Majidian, P., Zeinalabedini, M., Pirseyedi, S.M., Bahadori, M. 2014. Saffron (*Crocus sativus* L.), a monomorphic or polymorphic species? *Spanish Journal of Agricultural Research* 12, 753-762.

Nemati, Z., Blattner, F.R., Kerndorff, H., Erol, O., Harpke, D. 2018. Phylogeny of the saffron-crocus species group, *Crocus* series *Crocus* (Iridaceae). *Molecular Phylogenetics and Evolution*, doi: https:// doi.org/10.1016/j.ympev.2018.06.036

Nemati, Z., Harpke, D., Kerndorff, H., Gemicioglu, A., Blattner, F.R. submitted. Saffron (*Crocus sativus*) originated in mainland Greece from wild *Crocus cartwrightianus*.

Presentations (selected)

Nemati, Z., Mardi, M., Pirseyedi, S.M., Ghaffari, R., Marashi, S.H., Khayam, S.M. Isolation and characterization of microsatellite markers and their application for diversity assessment in saffron (*Crocus sativus*). 3rd International Symposium on Saffron, Kozani, Greece, 20-23 May 2009.

Nemati, Z., Majidian, P., Zeinalabedini, M., Eftekharian, A. Comparison of genetic variation among saffron (*Crocus sativus* L.) and different allies using microsatellite markers. 8th Iranian Biotechnology Congress, Tehran, Iran, 6-8 July 2013.

Nemati, Z., Harpke, D. Identification of parental species in saffron (*Crocus sativus*). 3rd Leibniz PhD Symposium: Keep it simple–Science communication, Berlin, Germany, 24-25 September 2015.

Nemati, Z., Harpke, D., Blattner, F.R. Identification of the origin of saffron (*Crocus sativus*). The 12th Plant Science Student Conference, Gatersleben, Germany, 4-7 July 2016.

Nemati, Z., Harpke, D., Blattner, F.R. The origin of saffron (*Crocus sativus*). The 13th Plant Science Student Conference, Halle, Germany, 6-9 June 2017.

Nemati, Z., Harpke, D., Kerndorff, H., Blattner, F.R. Triploidy results in safeguarding secondary metabolite pathways in saffron from recombination. The 14th Plant Science Student Conference, Gatersleben, Germany, 19-22 June 2018.

Posters (selected)

Nemati, Z., Majidian, P., Zeinalabedini, M., Eftekharian, A. Comparison of genetic variation among saffron (*Crocus sativus* L.) and different allies using microsatellite markers. 8th Iranian Biotechnology Congress, Tehran, Iran, 6-8 July 2013.

Majidian, P., Zeinalabedini, M., Nemati, Z. Utilization of next-generation sequencing in horticultural crops. 8th Iranian Horticultural Congress, Hamedan, Iran, 25-27 August 2013.

Nemati, Z., Harpke, D., Blattner, F.R. The origin of saffron (*Crocus sativus*)–molecular phylogenetic analyses to identify its parents. Plant Genome Evolution Conference, Sitges, Spain, 1-3 October 2017.

Eidesstattliche Erklärung / *Declaration under Oath*

Ich erkläre an Eides statt, dass ich die Arbeit selbstständig und ohne fremde Hilfe verfasst, keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.
I declare under penalty of perjury that this thesis is my own work entirely and has been written without any help from other people. I used only the sources mentioned and included all the citations correctly both in word or content.

_____ _____

Datum / Date Unterschrift des Antragstellers / *Signature of the applicant*

Erklärung über bestehende Vorstrafen und anhängige Ermittlungsverfahren /
Declaration concerning Criminal Record and Pending Investigations

Hiermit erkläre ich, dass ich weder vorbestraft bin noch dass gegen mich Ermittlungsverfahren anhängig sind. / *I hereby declare that I have no criminal record and that no preliminary investigations are pending against me.*

_____ _____

Datum / Date Unterschrift des Antragstellers / *Signature of the applicant*