

Article

# Phenotypic and Genetic Variation of an Interspecific *Centaurium* Hybrid (*Gentianaceae*) and Its Parental Species

Tijana Banjanac <sup>1,\*</sup>, Sanja Đurović <sup>2</sup>, Mihailo Jelić <sup>2</sup>, Milan Dragičević <sup>1</sup>, Danijela Mišić <sup>1</sup> ,  
Marijana Skorić <sup>1</sup>, Jasmina Nestorović Živković <sup>1</sup> and Branislav Šiler <sup>1</sup> 

<sup>1</sup> Institute for Biological Research “Siniša Stanković”, University of Belgrade, Bulevar despota Stefana 142, 11060 Belgrade, Serbia

<sup>2</sup> Faculty of Biology, University of Belgrade, Studentski trg 16, 11000 Belgrade, Serbia

\* Correspondence: tbanjanac@ibiss.bg.ac.rs

Received: 29 June 2019; Accepted: 12 July 2019; Published: 14 July 2019



**Abstract:** Interspecific hybridization is one of the major actuators of evolutionary changes in plants. As the result of allopolyploid hybridization, offspring may gain different ploidy levels in comparison to parental species, which can provide them instant reproductive isolation. Two tetraploid sister species, *Centaurium erythraea* and *C. littorale*, readily cross-fertilize, resulting in hybrids of various ploidy. In northern Serbia, two stable populations of a hexaploid taxon *C. pannonicum* have been documented. It has been proposed previously that this taxon emerged after an interspecific hybridization event between two tetraploid sister-species: *C. erythraea* and *C. littorale* subsp. *compressum*. The existing populations of the hybridogenic taxon, as well as neighboring populations of the two parental taxa were here characterized by both morphometrics and molecular markers (EST-SSR and *trnL-F*). Three leaf and two flower characteristics were found to be informative in delimitation of the parental taxa and in their discernment from hybrid individuals, the latter having intermediate values. Eight microsatellite markers were found to have good ability to distinguish studied taxa, placing *C. pannonicum* in closer relationship with *C. erythraea*. Conversely, *trnL-F* plastid marker nominated *C. littorale* subsp. *compressum* to be the donor of the *C. pannonicum* plastid DNA. Reproductive isolation of the hexaploid hybrid individuals from the parental species should be examined as the next logical step in describing the new species.

**Keywords:** *Centaurium*; interspecific hybridization; allopolyploidy; speciation; morphometry; EST-SSR; *trnL-F*

## 1. Introduction

The significance of interspecific hybridization in the “origin” of plant species and evolution has been debated for decades, with opposing views on it regarding it as a creative evolutionary force or as an evolutionary noise [1]. Since the reproductive barriers between morphologically well-described congeneric species are often weak, interspecific hybridization can occur quite frequently, which has been a focus of many botanists for a very long time [2].

It is a well-known fact that a new hybrid line can be formed through allopolyploidization or homoploid hybrid speciation, the former being more common and more feasible [1]. A recent article [3] suggests that 11% of the 47 therein studied plant species most likely have an allopolyploid origin, while Mallet [4] estimates that about 25% of plant species are known to hybridize with at least one other species.

Huge numbers of plant species have been reported to have passed through at least one whole genome duplication event [5], and immense angiosperm diversity is assumed to originate from

recurrent polyploidization events [6]. In between, polyploids may step into interspecific crossing, which may result in occurrence of hybrids harboring multiple sets of parental genomes.

The genus *Centaurium* Hill (family *Gentianaceae*) includes about 25 annual or biannual species having either diploid or polyploid genomes [7]. They are of Mediterranean origin, and were mostly spread to the north and west during the last few glaciations [8]. Different species and subspecies of this genus prefer similar habitats and are often found in sympatry, where they are “pleasing” to hybridize between themselves [9–13]. Interspecific hybridization represents one of the most important causes of phenotypic variations within the genus, causing appearance of hybrid swarms of morphologically intermediate appearance. Such prominent morphological polymorphism has led botanists to define an excessive number of species names within the genus [7,14–19], which further triggered nomenclatural problems and divergence in taxonomic approaches [7,20–22]. Back in 1948, Anderson [23] noticed that “the gene flow from one species to another may go far beyond any point which could be detected by ordinary morphological techniques”. Being additionally influenced by the environment, the morphological characteristics of the genus *Centaurium* are poorly featured in species determination and delimitation [7,11,16,20,24]. However, combination of morphological characteristics and newly developed molecular techniques may provide a solid basis for overcoming these difficulties. Morphometric methods are highly welcome in modern botanical studies to contribute in covering a wide picture of species variation, thus supporting or rebutting results obtained by cutting-edge molecular techniques. So far, different methods have been used in population genetics, taxonomy, and hybridization studies within the genus *Centaurium*, such as morphometrics and cytogenetics [11–13,25,26], or phytochemical profiling [9,27–32]. In modern phylogenetics studies, molecular markers are, of course, the most commonly applied [33]. Accordingly, three papers [8,24,34] introduced a comprehensive phylogenetic survey which indicated the polyphyletic origin of the genus *Centaurium* and proposed a new classification, while RAPD [9], ITS and plastid markers (*trnL-F* regions; [10]) have been used to study the interspecific hybridization between several species.

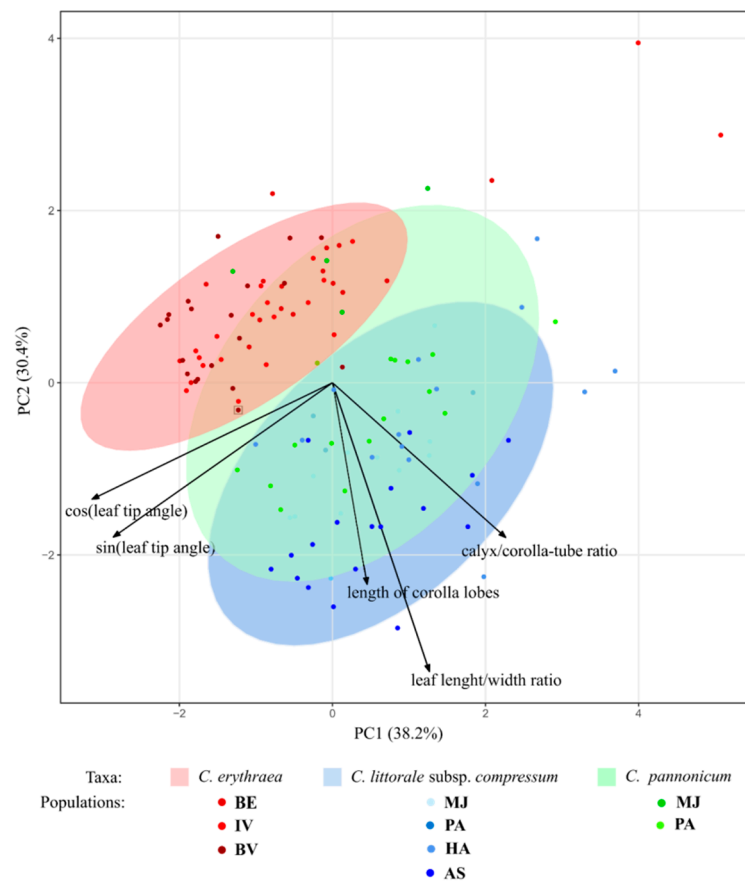
Here we studied the background and the outcomes of the interspecific hybridization between two sister-species: *C. erythraea* Rafn and *C. littorale* (Turner) Gilmour (in this particular case, *C. littorale* subsp. *compressum* (Hayne) Kirschner) (following up-to-date knowledge and achievements of the most relevant taxonomic and nomenclature source, Euro + Med Plantbase [19], the taxon *Centaurium littorale* (Turner) Gilmour contains two subspecies: *Centaurium littorale* (Turner) Gilmour subsp. *littorale* and *Centaurium littorale* subsp. *compressum* (Hayne) Kirschner, with *Centaurium uliginosum* (Waldst. & Kit.) Beck being a synonym for the latter one). They are both found in the continental areas of the central and east Europe [35] and, as we have shown earlier [32], both of them carry tetraploid genomes. Their interspecific hexaploid hybrids, named *C. pannonicum*, grow in stable populations in a few locations in the northern Serbia and were, however, found to be hexaploid [32]. With a hint that this might lead to speciation of this hybridogenic taxon, we aimed at analyzing the genetic basis of both parental taxa and the hybrid populations, implementing newly reported EST-SSR markers for the genus *Centaurium* [36] and analyzing the non-coding regions of the plastid DNA. As a contribution to perceiving the overall phenotypic picture of this interspecific hybridization case, which was started earlier by the phytochemical profiling [32], the divergence between the hybrid and the parental individuals is additionally investigated by a thorough analysis of the key morphological characteristics.

## 2. Results

### 2.1. Variation of Morphological Parameters

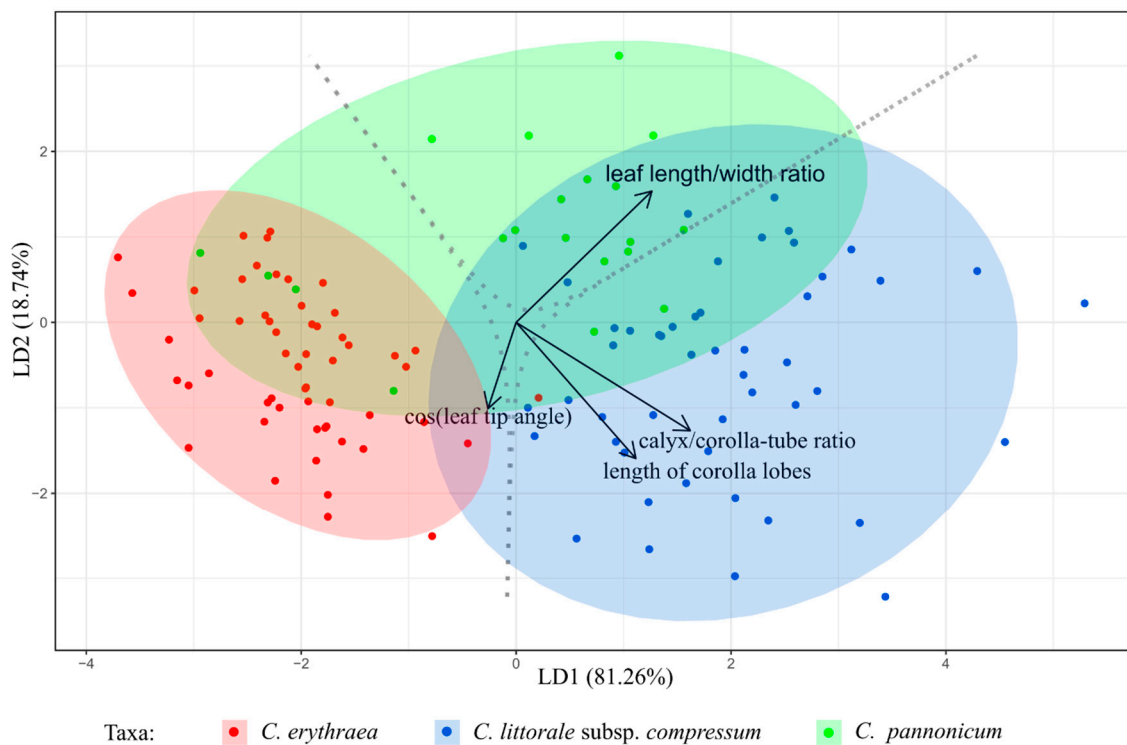
Principal component analysis (PCA) biplot based on five morphological characters of 127 individuals provided good discrimination between two parental species, particularly along the principal component 2 (PC2) (Figure 1). The first two dimensions contribute 68.6% to the total variation and are sufficient based on Horn’s parallel analysis. The cloud comprised of the hybrid individuals

is rather scattered, but is positioned between the two clouds belonging to the parental species. Sub-structuring of the tree taxa by the populations is not observed.



**Figure 1.** PCA biplot of three *Centaurium* taxa (*C. erythraea*, *C. littorale* subsp. *compressum* and *C. pannonicum*) based on the following characters: length/width ratio of floral stalk leaves, calyx/corolla-tube ratio, length of corolla lobes, and leaf tip angle, which was decomposed to the sine and cosine components. The scatterplot consists of 127 individuals in the plane of the first and the second principal coordinates. Feature loadings were scaled by multiplying with 5 so they could be more readily observed on the graph.

Linear discriminant analysis (LDA) showed that the selected morphological characters were very useful for taxa diversification (Figure 2), particularly between the hybrid individuals and *C. littorale* subsp. *compressum*, which was hardly discoverable by PCA. Based on the sequential backward search, the highest classification accuracy (91.5%) in repeated k-fold cross validation was achieved when LDA used four out of the five provided features (omitting  $\sin(\text{leaf tip angle})$ ). The top features in order are: length/width ratio of floral stalk leaves (LDA achieves 65.6% classification accuracy on average on the hold out instances in repeated k-fold cross validation with just this feature) + calyx/corolla-tube ratio (77.8% classification accuracy with the addition of this feature) + length of corolla lobes (89.5% classification accuracy with the addition of this feature) and +  $\cos(\text{leaf tip angle})$  achieving the observed 91.5% accuracy. The highest proportion of falsely classified individuals was from *C. pannonicum* which was predicted as *C. erythraea* on average in 3.2% of total cases and *C. littorale* subsp. *compressum* which was predicted to be *C. pannonicum* on average in 2.7% of total cases (Table S1).



**Figure 2.** Visualization of LDA of three *Centaurium* taxa (groups were defined as: *C. erythraea*, *C. littorale* subsp. *compressum* and *C. pannonicum*) based on the following characters: length/width ratio of floral stalk leaves, calyx/corolla-tube ratio, length of corolla lobes, and cosine of the leaf tip angle. For graph construction, the LDA model fitted on all the data was used. Model decision bounds are indicated by the dashed lines.

## 2.2. Genetic Diversity Estimation: Eight-Locus Microsatellite Analysis

Population-genetics parameters are summarized for nine populations in Table 1. The highest number of alleles per population (34) was observed for *C. erythraea* originating in Bački Vinogradi (*C.ery.BV*) with average 4.25 alleles per locus. The lowest number of alleles was recorded for two populations of *C. littorale* subsp. *compressum*, originating in Majdan and Palić (*C.l.c.MJ* and *C.l.c.PA*), both averaging 2.5 alleles per locus. Throughout the populations, 27.33 alleles per population were recorded on average.

**Table 1.** Population genetics parameters obtained by eight genic microsatellite loci: the total, average and effective number of alleles in populations and the values of the observed ( $H_O$ ) and expected heterozygosity ( $H_E$ ) of the populations are presented.

Populations	Number of Analyzed Individuals per Population	Total Number of Alleles in a Population (all 8 loci)	Average Number of Alleles for All Loci	Effective Number of Alleles in a Population	Average Expected Heterozygosity per Population ( $H_E$ )	Average Observed Heterozygosity per Population ( $H_O$ )
<i>C.ery.BE</i>	17	29	3.6250	1.1400	0.8772	0.5588
<i>C.ery.IV</i>	19	31	3.8750	1.1326	0.8829	0.5658
<i>C.ery.BV</i>	20	34	4.2500	1.1198	0.8930	0.5938
<i>C.pan.MJ</i>	7	31	3.8750	1.0089	0.9912	0.7969
<i>C.pan.PA</i>	22	33	4.1250	1.0114	0.9888	0.7337
<i>C.l.c.MJ</i>	13	20	2.5000	1.1303	0.8847	0.6635
<i>C.l.c.PA</i>	5	20	2.5000	1.1482	0.8709	0.5750
<i>C.l.c.HA</i>	15	24	3.0000	1.0995	0.9095	0.7750
<i>C.l.c.AS</i>	18	24	3.0000	1.0982	0.9105	0.8125
total	136					
average		27.33	3.4167	1.0988	0.9121	0.6750

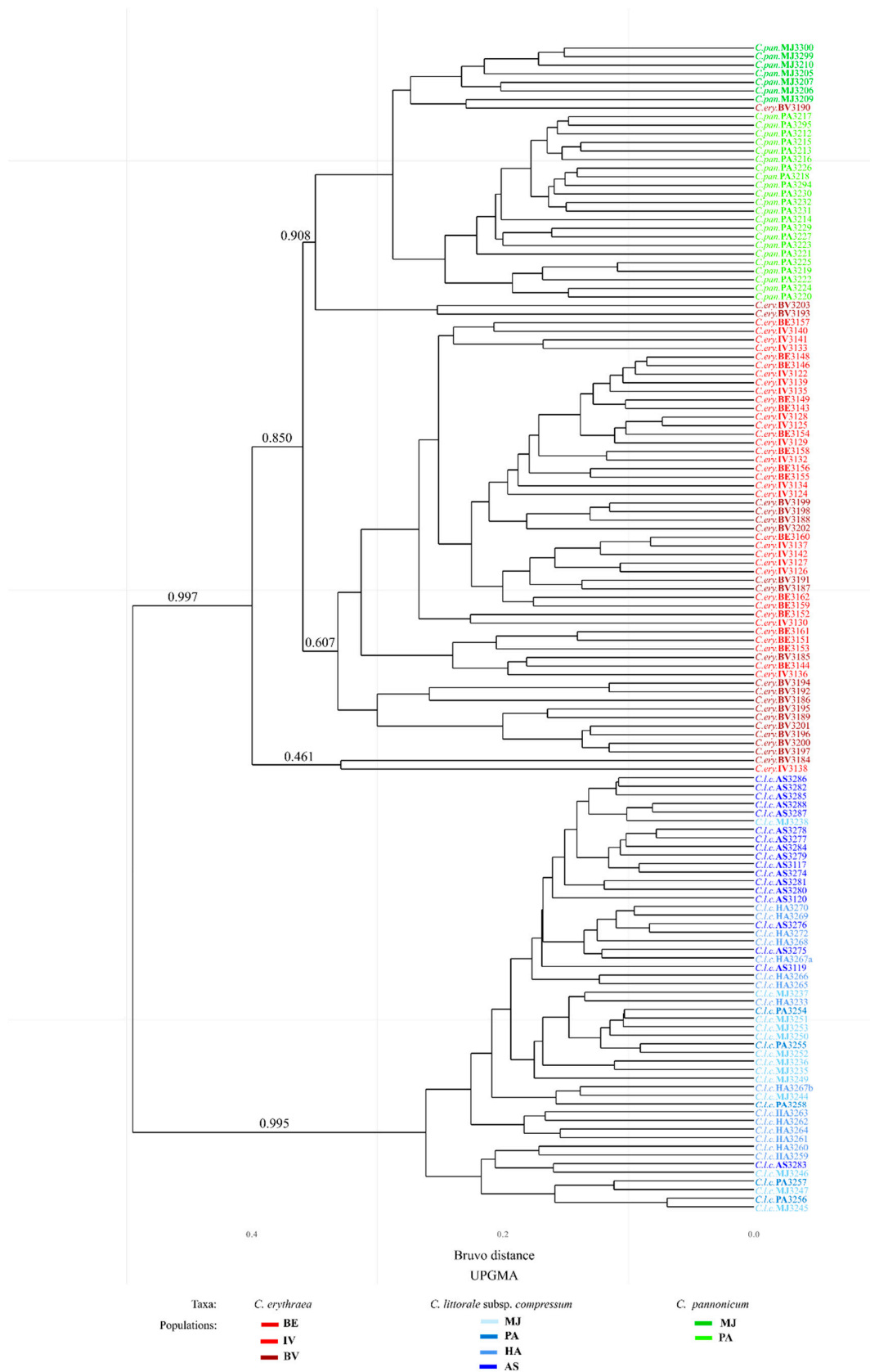
Expected heterozygosity ranged from 0.8709 in *C. littorale* subsp. *compressum* (**PA**) to 0.9912 in the *C. pannonicum* (**MJ**) with an average value of 0.9121 throughout the studied populations. Observed heterozygosity was evidently lower for all populations: from 0.5588 in the *C. erythraea* (**BE**) to 0.8125 in the *C. littorale* subsp. *compressum* (**AS**), with an average of 0.6750 (Table 1).

The  $F_{ST}$  and Jost's D indices, which quantify genetic differences between populations, are presented in Table 2. The lowest distances were recorded between three *C. littorale* subsp. *compressum* populations originating in Subotica Sands: Majdan (**MJ**), Palić (**PA**) and Hajdukovo (**HA**), followed by two *C. erythraea* populations from Fruška Gora Mountain (**BE** and **IV**). According to these results, the three populations of *C. littorale* subsp. *compressum* differ at the most from two *C. erythraea* populations from Fruška Gora Mountain and slightly less from the neighboring *C. erythraea* population **BV**. Each of the two *C. pannonicum* populations had considerably lower  $F_{ST}$  values vs. all the three populations of *C. erythraea* than vs. each of *C. littorale* subsp. *compressum*. The same trends can be observed within the Jost's D values.

**Table 2.** Estimated genetic differentiation of populations based on  $F_{ST}$  and Jost's D indices.

$F_{ST}$	Jost's D	<i>C.ery.IV</i>	<i>C.ery.BE</i>	<i>C.ery.BV</i>	<i>C.pan.MJ</i>	<i>C.pan.PA</i>	<i>C.l.c.HA</i>	<i>C.l.c.MJ</i>	<i>C.l.c.PA</i>	<i>C.l.c.AS</i>
<i>C.ery.IV</i>			0.015	0.055	0.106	0.113	0.200	0.241	0.173	0.210
<i>C.ery.BE</i>		0.019		0.063	0.111	0.108	0.200	0.240	0.180	0.208
<i>C.ery.BV</i>		0.126	0.141		0.072	0.096	0.144	0.174	0.114	0.158
<i>C.pan.MJ</i>		0.282	0.270	0.236		0.031	0.173	0.210	0.170	0.184
<i>C.pan.PA</i>		0.305	0.300	0.312	0.124		0.154	0.162	0.090	0.179
<i>C.l.c.HA</i>		0.526	0.497	0.433	0.553	0.514		0.033	0.024	0.013
<i>C.l.c.MJ</i>		0.553	0.509	0.437	0.547	0.501	0.054		0.007	0.032
<i>C.l.c.PA</i>		0.533	0.491	0.415	0.507	0.467	0.039	−0.021		0.030
<i>C.l.c.AS</i>		0.552	0.519	0.466	0.609	0.573	0.019	0.065	0.077	

UPGMA dendrogram constructed from Bruvo genetic distances, which were obtained from the allele frequencies at eight EST-SSR loci (Figure 3), consists of two main clusters which split in Bruvo distance at the value of about 0.5. The first cluster, which is supported by a high (0.995) bootstrap value, is comprised solely of the individuals belonging to *C. littorale* subsp. *compressum*. The cluster is quite coherent, and four populations of this taxon are spread more or less evenly, with no population differentiation visible except that the individuals from the population **AS** tend to cluster together. The second cluster, also having a strong (0.997) bootstrap support, contains two other taxa. A substantial genetic variation of *C. erythraea* is visible in this cluster. Two individuals (*C.ery.BV* 3184 and *C.ery.IV* 3138), originating in different populations, are the first to segregate at the Bruvo distance of 0.4. Most of the remaining individuals of the same taxon are further clustering together in a fairly heterogeneous sub-cluster without population diversification. However, the other sub-cluster of the second cluster, comprising mainly hybrid individuals, also contains three "interlopers" from *C. erythraea* (Figure 3). Two of them differentiate earlier from the sub-cluster, with a negligible distance from the *C. erythraea* sub-cluster. The third individual (*C.ery.BV* 3190) turned out to be misclassified at the beginning of the study (see below). In any case, the mentioned sub-cluster is made of the hybrid *C. pannonicum* individuals with obvious differentiation between its two populations.

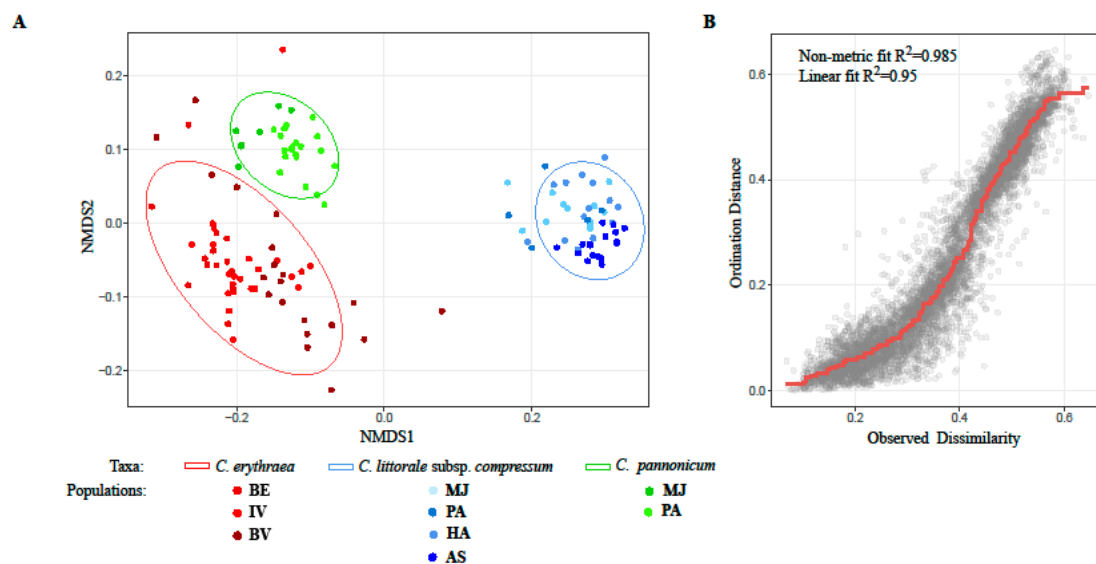


**Table 3.** *Centaureum erythraea*, *C. littorale* subsp. *compressum* and *C. pannonicum* accessions used in the study.

Taxon	Locality	Population Accession Code	Geographic Latitude and Longitude	Estimated Ploidy Level *	No. of Studied Individuals		
					EST-SSR	Morphometrics	Plastid DNA
<i>C. erythraea</i>	Beočin	<i>C.ery.BE</i>	45°10'35.24" N 19°43'16.32" E	4x	17	17	2
<i>C. erythraea</i>	Iriški Venac	<i>C.ery.IV</i>	45°09'27.65" N 19°51'43.64" E	4x	19	19	3
<i>C. erythraea</i>	Bački Vinogradi	<i>C.ery.BV</i>	46°07'29.19" N 19°50'56.02" E	4x	20	20	10
<i>C. littorale</i> subsp. <i>compressum</i>	Majdan	<i>C.l.c.MJ</i>	46°09'29.34" N 19°36'20.53" E	4x	13	13	4
<i>C. littorale</i> subsp. <i>compressum</i>	Palić	<i>C.l.c.PA</i>	46°03'08.04" N 19°46'34.89" E	4x	5	4	3
<i>C. littorale</i> subsp. <i>compressum</i>	Hajdukovo	<i>C.l.c.HA</i>	46°06'52.27" N 19°49'53.58" E	4x	15	15	4
<i>C. littorale</i> subsp. <i>compressum</i>	Ásotthalom	<i>C.l.c.AS</i>	46°11'04.50" N 19°50'47.62" E	4x	18	18	1
<i>C. pannonicum</i>	Majdan	<i>C.pan.MJ</i>	46°09'29.34" N 19°36'20.53" E	6x	7	5	5
<i>C. pannonicum</i>	Palić	<i>C.pan.PA</i>	46°03'08.04" N 19°46'34.89" E	6x	22	16	10
<i>C. pulchellum</i>	Palić	<i>C.pul.PA</i>	46°03'08.04" N 19°46'34.89" E	NA	/	/	1
in total					136	127	43

\* Based on the results published in Banjanac et al. [32].

Non-metric multidimensional scaling (Figure 4) was used to visualize the results obtained from the Bruvo genetic distances among 136 individuals. It showed an evident differentiation of the *C. littorale* subsp. *compressum* group of individuals from the two other taxa, upholding the results obtained by unsupervised UPGMA. Likewise, the second group is comprised of *C. erythraea* and *C. pannonicum* genotypes, the latter one forming its own subgroup.



**Figure 4.** Nonmetric multidimensional scaling (NMDS) based on the Bruvo distances between individuals. (A)—NMDS scatterplot; populations are indicated by different colors as shown in the legend. Each taxon is encircled by a 95% confidence ellipse: red—*C. erythraea*, green—*C. pannonicum*, blue—*C. littorale* subsp. *compressum*. (B)—Shepard diagram (goodness-of-fit scatterplot) of observed Bruvo distances and NMDS transformed distances. Stress = 0.12.

### 2.3. Variation in *trnL-F* Plastid Region

The length of the unaligned amplified *trnL-F* region of 43 individuals included in research ranged between 504 and 514 bp. Within this research, it was not possible to read the sequences with primer “c”. The final alignment was done with 43 sequences (Table 3), which were aligned against the 16 *trnL-F* intergenic sequences retrieved from GenBank (Table 4).

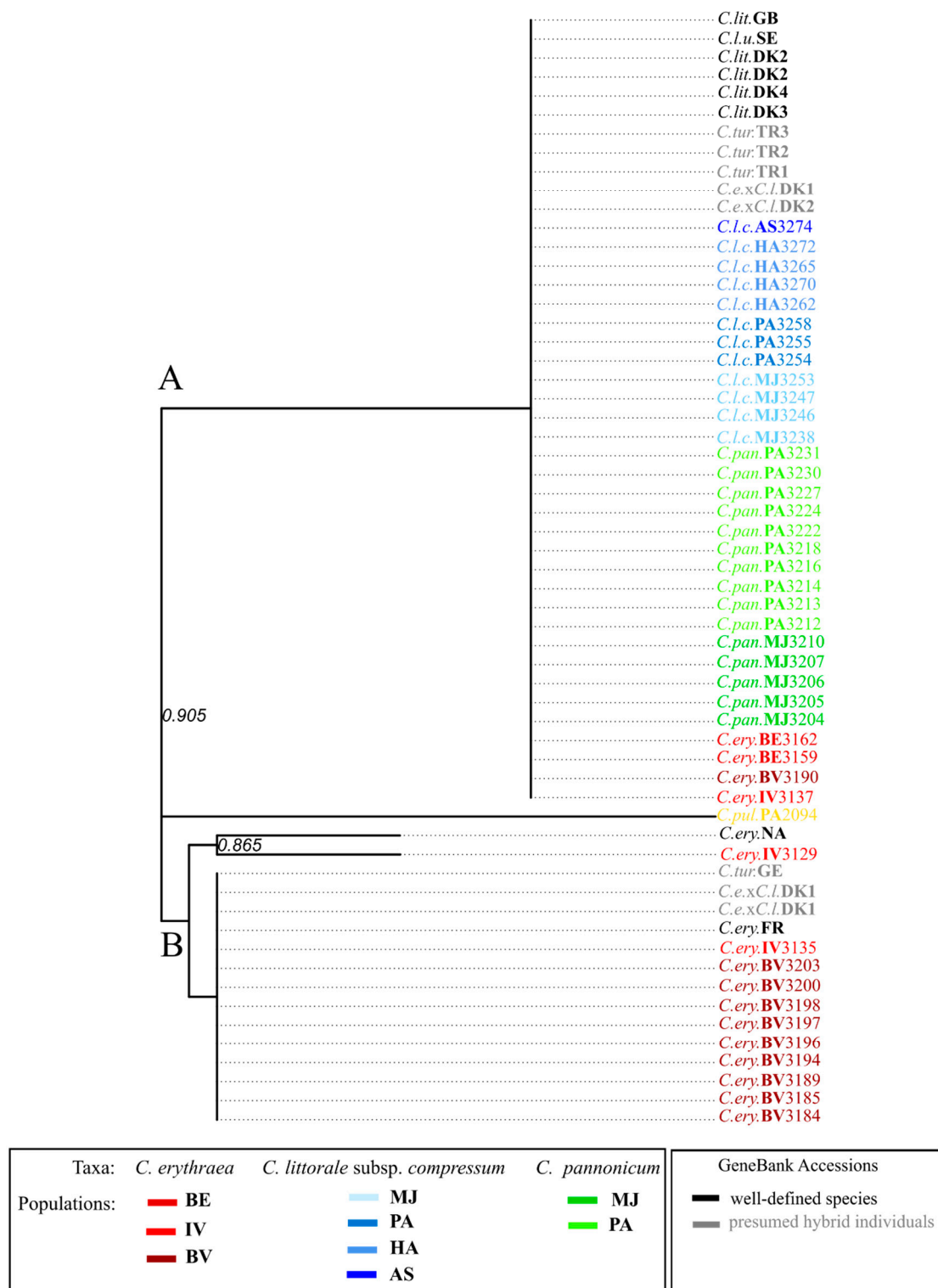
**Table 4.** Origin of plant material, reference information and GenBank accessions for DNA sequences used for comparison of *trnL-F* haplotypes.

Taxon	GenBank Accession No.	Accession Code Used in Manuscript	References	Ploidy Level	Country, Region, Locality
<i>Centaurium erythraea</i>	AY315697	<i>C.ery.NA</i>	Mansion, G. and Zeltner, L.; Unpublished	NA	NA
<i>Centaurium erythraea</i>	AY879973	<i>C.ery.FR</i>	[8]	4x	France, Gard, Espiguette
<i>C. erythraea</i> × <i>C. littorale</i>	AY879975	<i>C.e.×C.l.DK1</i>	[8]	4x	Denmark, Jutland, Sand Milen
<i>C. erythraea</i> × <i>C. littorale</i>	AY879976	<i>C.e.×C.l.DK2</i>	[8]	4x	Denmark, Jutland, Skallingen
<i>C. erythraea</i> × <i>C. littorale</i>	AY879977	<i>C.e.×C.l.DK1</i>	[8]	4x	Denmark, Jutland, Sand Milen
<i>C. erythraea</i> × <i>C. littorale</i>	AY879978	<i>C.e.×C.l.DK1</i>	[8]	4x	Denmark, Jutland, Sand Milen
<i>Centaurium littorale</i>	AY879979	<i>C.lit.DK3</i>	[8]	4x	Denmark, Jutland, Grenen
<i>Centaurium littorale</i>	AY879980	<i>C.lit.DK4</i>	[8]	4x	Danemark, Lakolk, Romo Island
<i>Centaurium littorale</i>	AY879981	<i>C.lit.DK2</i>	[8]	4x	Denmark, Jutland, Skallingen
<i>Centaurium littorale</i>	AY879982	<i>C.lit.DK2</i>	[8]	4x	Denmark, Jutland, Skallingen
<i>Centaurium littorale</i>	AY251756	<i>C.lit.GB</i>	[34]	4x	England, Holy Island
<i>Centaurium turcicum</i>	AY880009	<i>C.tur.GE</i>	[8]	6x	Georgia, Agara
<i>Centaurium turcicum</i>	AY880010	<i>C.tur.TR1</i>	[8]	6x	Turkey, Phrygie, Pamukkale
<i>Centaurium turcicum</i>	AY880011	<i>C.tur.TR2</i>	[8]	6x	Turkey, Phrygie, Beysehir
<i>Centaurium turcicum</i>	AY880012	<i>C.tur.TR3</i>	[8]	6x	Turkey, Pamphylie, Aksu Isparta
<i>C. littorale</i> subsp. <i>uliginosum</i>	AY880013	<i>C.l.u.SE</i>	[8]	4x	Sweden, Gotland, Herrvik

A maximum-likelihood phylogenetic tree of *trnL-F* dataset (Figure 5) revealed two haplotypes in the whole set of sequences. Haplotype A possessed all the individuals characterized as *C. pannonicum*, *C. littorale* subsp. *compressum*, *C. littorale* and *C. uliginosum* (the synonym for *C. littorale* subsp. *compressum*). This haplotype was harbored by 5 more individuals characterized as hexaploid *C. turcicum* and tetraploid *C. erythraea* × *C. littorale*. Haplotype A also appeared in four individuals of *C. erythraea*, two of which were from **BE** locality, and one from each **IV** and **BV**.

Haplotype B owned all the remaining sequences belonging to *C. erythraea* individuals as well as another two sequences belonging to the tetraploid hybrid individuals of *C. erythraea* × *C. littorale* and one sequence from a possible hexaploid hybrid taxon, *C. turcicum*.



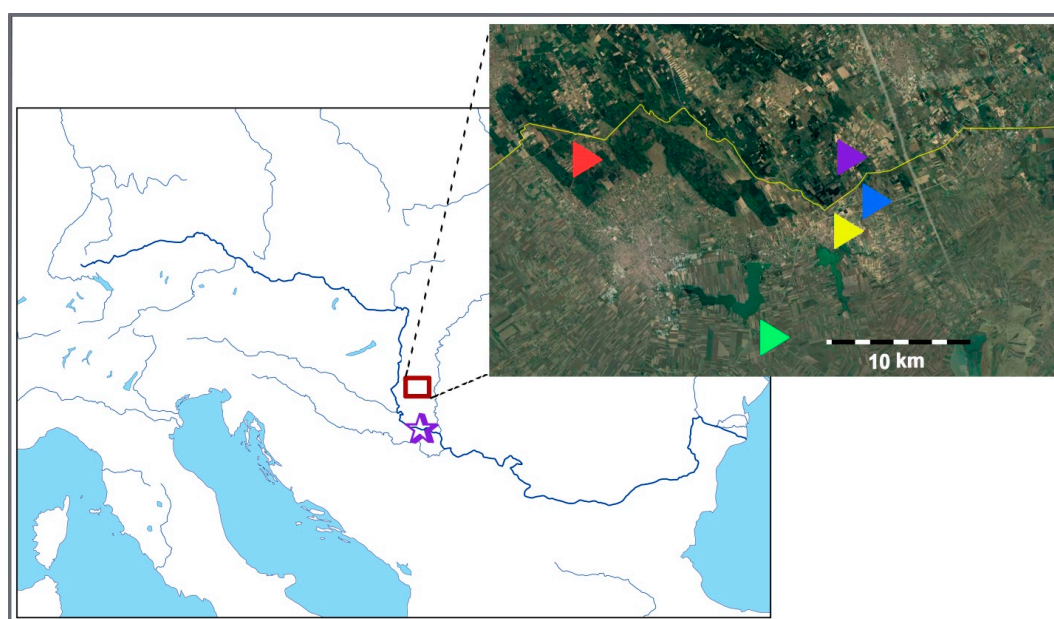


**Figure 5.** Maximum likelihood phylogenetic tree based on *trnL-F* region of the studied *Centaurium* taxa with *C. pulchellum* as outgroup. Support cluster values, estimated using the Shimodaira-Hasegawa test, are indicated. Sequence tags retrieved from the GenBank are written in black and gray letters while the initials of the country of origin are follows: **GB**-Great Britain, **SE**-Sweden, **TR**-Turkey, **GE**-Georgia, **FR**-France, **DK**-Denmark. For labels please refer to Tables 3 and 4.

### 3. Discussion

#### 3.1. About the Sample Set

The focus of this research was placed on the interspecific hybridization between two tetraploid species belonging to the genus *Centaureum*, *C. erythraea* and *C. littorale* subsp. *compressum*. Hybrid individuals having a hexaploid genome were only recently noticed in two localities, namely Majdan—MJ and Palić—PA. The aims and the directions of the presented research have been conceived on the basis of several facts and observations. Firstly, literature data indicate great phylogenetic relatedness of *C. erythraea* and *C. littorale* [8,34,37]. Secondly, examples of hybridization of certain subspecies of these two species have been previously documented and confirmed by various methods [8,11–13]. The fact that these authors have reported hybridization between *C. erythraea* and *C. littorale* in geographically remote localities from our study site led us to presume that the hybrids might arise after independent hybridization events, which certainly deserves our attention. Finally, a research on interspecific hybridization would possibly raise public awareness on importance to further protect Subotica Sands and Nature Park Palić, particularly the locations MJ and PA, since the ubiquitous arable land of these plain landscapes, roads, and other obstacles derived from human activity represent barriers for species distribution and cause strong fragmentation of the habitat (as can be seen on Figure 6). Although disadvantageous for the survival of the species, such relatively small number of clearly restricted fragmented populations provides a good basis to study cross-species hybridization incidence in natural conditions within the genus *Centaureum*.



**Figure 6.** Map of south-east Europe with a detailed terrain map of the study area of Subotica Sands. The colored triangles on the extracted Google map\* represent localities of existing populations in the area: red—Majdan (MJ); green—Palić (PA); yellow—Hajdukovo (HA); blue—Bački Vinogradi (BV); purple—Ásothalom (AS). The purple star on the basic map represents approximate position of Fruška Gora Mt. populations of *C. erythraea*—Iriški venac (IV) and Beočin (BE). \* Satellite image downloaded from Google Earth software (Google Earth Pro, v. 7.3.2.5776, recording date: March 31st, 2019, Subotica, view height: 32.33 km, DigitalGlobe 2019, download date: 6 May 2019).

This study included all four existing populations of *C. littorale* subsp. *compressum* and two of *C. pannonicum* found in the specified area. On the other hand, due to the deficiency of nearby populations of *C. erythraea*, in addition to neighboring BV, two populations located about 200 km south, on the Fruška Gora Mountain were included in this research. Plants of *C. littorale* subsp.

*compressum* and *C. pannonicum* were found growing in sympatry on the localities **MJ** and **PA** (Figures 6 and 7) but the second species involved in hybridization, *C. erythraea*, was not found there. This fact raises the following question: since growing in sympatry represents the main prerequisite for the interspecific hybridization to occur, particularly for species which are pollinated by short-distance insects [38], why is *C. erythraea* no longer found in these locations? Literature data indicate that *C. erythraea* and *C. littorale* subsp. *compressum* exhibit slightly different habitat preferences: while *C. erythraea* is more often found on dry limestone soils, and rarely in loamy, sandy and wetlands, *C. littorale* subsp. *compressum* has a preference for settled and humid habitats [39]. Also, in the study of interspecific hybridization of *C. erythraea* subsp. *erythraea* and *C. littorale* subsp. *littorale* on the territories of Great Britain, Germany and Denmark, one of the author's conclusions was that these taxa have shown different habitat preferences [13]. According to this study, the most probable assumption is that the two species step into hybridization when the habitat is disturbed, mostly by human activity, which potentiates their contact. Localities **MJ** and **PA** are certainly habitats created by human activity and it can be reasonably assumed that the emergence of a hybridogenic taxon at these sites was the result of anthropogenic changes. Once established, the hybridogenic taxon can push out one or both parental species destroying the evidence of its own hybrid origin [4,10]. Bearing that in mind, we may hypothesize that this situation has happened in the case of interspecific hybridization between *C. erythraea* and *C. littorale* subsp. *compressum*. *C. erythraea*, as a species that prefers dryer habitats, may have been suppressed by the other two taxa and finally perished from the mutual habitats. Changes in the habitat caused by human activity can greatly alter ecological conditions and thus favor some groups of organisms that were not competitive before the disturbance [1]. Our study points to a stunning postulation that the activities of people, completely unconsciously, by altering life conditions in habitats, may have led to the creation of a new hybridogenic taxon. Although the fate of this taxon in nature may be uncertain, the results of this research point out to the need to protect the unique natural goods of Majdan and Palić, which represent places where it can be possible to follow evolutionary processes, reproductive isolation, and speciation in situ.



**Figure 7.** Localities where hexaploid plants characterized as *C. pannonicum* were found. (a–c)—Majdan (**MJ**); (d)—Palić (**PA**).

### 3.2. Morphometrics in the Detection of Hybrids

A well-known fact is that plant morphological characteristics may be highly susceptible to environmental factors, edaphic conditions, seasons, plant infections, pests, etc. This is especially true for plants belonging to the genus *Centaureum* in which high clinal variation in gross morphology, often caused by ecological conditions, is frequently observed [7,13,16,24,34]. However, meticulous selection of parameters to consider for a study, which would be only negligibly influenced by environmental pressures, is of critical importance to obtain an unbiased morphometric comparison. The second challenge of this part of the study was to select morphological traits which would have enough resolution power to differentiate between two closely related species and the hybridogenic taxon. For instance, as proposed by Ubsdell [11], we have studied pollen grains' dimensions of the three taxa, determined to find the differentiation among them. However, we recorded variance that was several times greater within the taxa than among them and had to discard this parameter as uninformative.

As can be seen from Figures 1 and 2, as well from Figure S1, five selected morphological traits showed different potential to support differentiation between the taxa. For some characteristics, basic statistics indicated that they were under higher environmental pressures than others. This specifically refers to length/width ratio of floral stalk leaves. It can be observed (Figure S2) that the two populations of *C. pannonicum* considerably differ among themselves regarding this trait, which can be explained by the different habitats the plants grew in. Plants from the population **MJ** were found in an open habitat with only a few other plant species found in the same phytocenosis (Figure 7a–c). They grew up to 30 cm in height and have developed smaller leaves due to high insolation. Contrastingly, plants from the population **PA** were found in an irrigation canal in a semi-shaded habitat where the plant community consisted of at least twenty different, densely growing species (personal observation, Figure 7d). In this habitat, *C. pannonicum* developed several times longer leaves on plants up to 80 cm tall (Figure S3). Leaf width, however, was mildly affected by such an environment, which resulted in more lanceolate leaves (Figure S4). Therefore, while highly appreciated to differentiate two parental species, length/width ratio of floral stalk leaves showed weak relevance in delimitation of interspecific hybrids, as proposed by Ubsdell [11], although this trait, along with leaf tip angle, falls among the indispensable basic morphological key parameters for use in the field studies for the determination of the potential hybrid plants (personal observation). As the results of such practice applied in the field, in the whole study we failed to correctly determine only one plant (*C. ery.***BV** 3190).

Data obtained by measuring other parameters, such as calyx/corolla-tube ratio and length of corolla lobes, were highly concordant with previous studies [11]. The calyx/corolla-tube ratio delimitates all the three taxa quite well, while the length of corolla lobes gives a good basis for discrimination of sympatric taxa *C. littorale* subsp. *compressum* and *C. pannonicum* (Figure S1).

In the PCA biplot based on the five morphometric traits (Figure 1) we can notice partial overlap of the hybrid individuals' cloud with the clouds of both of the parental species. This could explain why this hybridogenic taxon has remained unnoticed until few years ago, as well as why sporadic misidentifications occur in the field during the collection of plant material while opening possibilities for disambiguation in taxonomy of the genus [24,34].

Although we were guided by previous morphological analyses [11] in the initial selection of morphological traits to be measured, it was not possible to give a diagnostic key which would avoid overlapping in the character measures of the hybrid and either of the parental species. It may be necessary to emphasize once again that the two parental species are very closely related and such a situation regarding the lack of differences in the morphological characteristics is highly anticipated.

More thorough studies, which would include additional morphological and anatomical characters, might give a clearer delimitation of *C. pannonicum* from *C. erythraea* and *C. littorale* subsp. *compressum*.

### 3.3. Genetic Variation

To estimate genetic variation in natural populations of both parental species and the interspecific hybrid, the EST-SSR multilocus genotyping was performed and reported for the first time for the populations within the genus *Centaureium*. For this purpose, we employed newly designed genic microsatellite markers [36].

The specificity of the microsatellite markers enabled us to analyze plant material collected directly from nature, without amplifying foreign (e.g., fungal or insect) DNA, which may give false variations when non-specific markers are used. In the previous study [36], we confirmed the transferability of the eight genic microsatellite markers, initially developed for *C. erythraea*, to seven other taxa, and, among them, the two studied here. Still, since EST-SSR markers are basically developed from *C. erythraea* transcriptome, it is not surprising that they showed higher variability in this species (Table 1) in comparison to the other two studied taxa [40,41].

The most prominent genetic variability within the *C. erythraea* populations was recorded for Bački Vinogradi (BV). This might happen because BV locality may be from time to time populated with hybrid individuals and also with *C. littorale* subsp. *compressum*, a few of them spotted in 2018, which certainly promotes the gene flow between the taxa. This may be the main reason for misidentification of the individual *C.ery.BV* 3190, which turned out to group with *C. pannonicum* (Figure 3). Plastid DNA marker analysis supports this conclusion (please see below). However, our previous research on populations' ploidy [32] did not reveal hexaploid individuals in this locality. Does this indicate continuous gene flow between different tetraploid taxa? Probably yes. Similar examples have been reported earlier [8,13,24]. This information can give very good support to the hypothesis that hybridization between *C. erythraea* and *C. littorale* subsp. *compressum* is very common and may or may not include polyploidization. The mentioned individual, *C.ery.BV* 3190, is also positioned somewhere on the edge of the *C. erythraea* cloud in the PCA scatterplot based on morphometrics (Figure 1, marked with a rectangle around the corresponding point). LDA based on morphometrics data did not include this individual.

Within the *Gentianaceae* family, sequences of the non-coding *trnL* (UAA) intron and *trnL-trnF* spacer region from the plastid DNA have been previously analyzed in order to estimate phylogenetic relationships [34,42–45], or to detect evidence of reticulation [8,24]. The primer combination used for the amplification of this plastid non-coding region (“c” and “f” primers reported by Taberlet et al. [46]) was not used in similar studies. Such a combination was used here in order to provide longer amplicons which would consist of two joint plastid regions, *trnL* (UAA) intron and *trnL-trnF* spacer region. Nevertheless, within this research, it was not possible to read the sequences with primer “c”, probably due to the existence of polyA region (about 17–20 bp), which was observed on the forward part of the sequences. PolyA regions are unstable, causing polymerase slips during extension, and thus making sequencing impossible [47]. Because of this inconvenience, the sequences obtained in this study were aligned only according to intergenic spacer *trnL-trnF* sequences of closely related taxa and potential hybrid individuals retrieved from GenBank. Although the resulting matrix of the aligned sequences consisted of just 380 characters, the phylogenetic analysis showed two distinctive haplotypes present among all analyzed individuals. The same haplotypes and affiliations were recorded with 43 longer sequences (504–514 bp), which were obtained by sequencing the amplicons in only one direction (results are not shown but are available upon request).

Plastid DNA is maternally inherited in the genus *Centaureium* [10,48] and the analysis of plastid haplotypes was done in order to reveal the maternal progenitor of the hybridogenic taxon *C. pannonicum*. Plastid introgression after intrageneric interspecific hybridization is a well-documented event [49,50]. The analysis of the plastid haplotypes reported here, clearly indicates that the most likely plastid donor (and therefore the maternal species) to the hybridogenic taxon *C. pannonicum* were some population(s) of *C. littorale* subsp. *compressum* from the same or nearby areas, since all analyzed individuals of *C. pannonicum*, as well as all individuals of *C. littorale* subsp. *compressum*, have been shown to possess the same plastid haplotype. Three individuals of *C. erythraea* share the same haplotype but these

individuals belong to distant populations (**BE** and **IV**; Figure 6). An interesting fact is that one individual of *C. erythraea* from Bački Vinogradi (*C.ery.BV* 3190) also showed a haplotype characteristic for hybridogenic taxon and *C. littorale* subsp. *compressum*. However, for this individual, with very high certainty, we can infer it was incorrectly determined. As already mentioned, the analysis of microsatellite markers also grouped this individual with *C. pannonicum*. In general, two haplotypes can be noticed within the sample set made up from the studied individuals and previously reported GenBank sequences, but in some cases, it was not possible to link haplotypes to a particular species. It can be rather said that different haplotypes could be linked to different populations. This claim can be supported by the repeatedly underlined (and studied) fact that interspecific hybridization within the genus *Centaurium* occurs very commonly [7,10,13,16,51,52], leading to reticulate evolution of the plastid haplotypes. Because of its haploid nature and mode of inheritance, plastid genome is more sensitive to genetic drift effects, which may result in fixation of population-characteristic haplotypes.

When talking about the direction of interspecific hybridization, the choice of which species would act as the pollen donor may be influenced by various ecological factors (e.g., the number of individuals of a particular species in the habitat, pollinators' preferences), as well as the biological characteristics of two species which hybridize (e.g., pollination mechanisms, flowering period, different habitat preferences). In his extensive studies on hybridization between *C. erythraea* subsp. *erythraea* and *C. littorale* subsp. *littorale*, Ubsdell [11–13] has come to a conclusion that the pollen donor had to be *C. littorale* subsp. *littorale* individuals due to slightly different flowering period and difference in maturity of stigmas and pollen grains between the two taxa. On the other hand, when we “switched off” the ecological factors and the species-specific biology by performing artificial hybridization under in vitro conditions [9], seeds were formed only in cases when the pollen donor was *C. erythraea*.

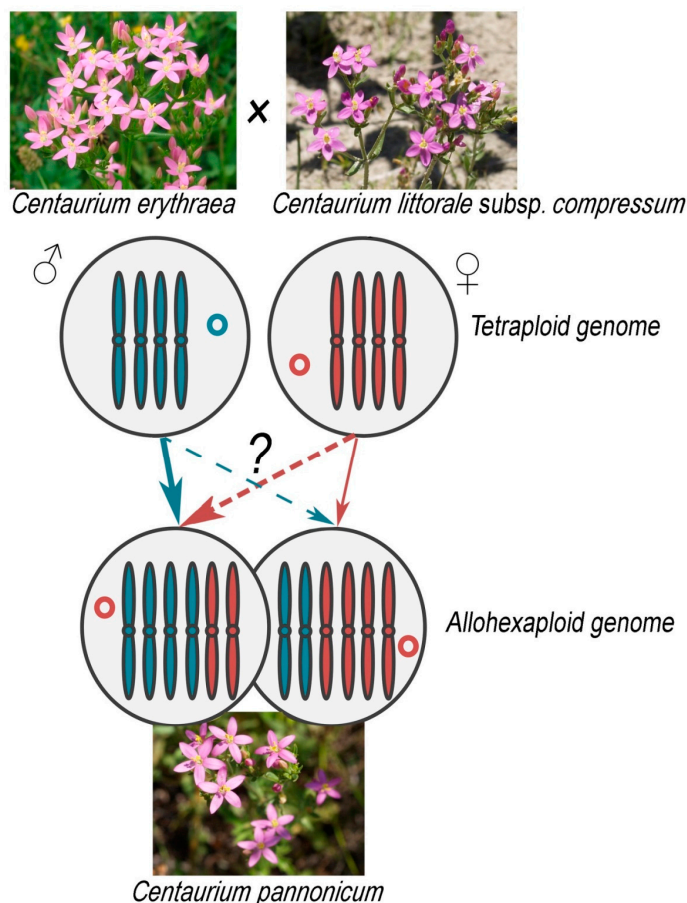
Among the populations studied here, the analysis with microsatellite molecular markers showed closer phylogenetic relationship of hybridogenic taxon *C. pannonicum* with *C. erythraea* than with *C. littorale* subsp. *compressum*. On the contrary, plastid *trnL-F* marker haplotyping pointed at *C. littorale* subsp. *compressum* to be the maternal species of *C. pannonicum*. The similar situation of discordant data obtained with differently inherited markers is often found among plant species [8,53]. Additional screening with differently inherited markers, possibly with other DNA barcoding markers, could provide more information and enrich our perspective on this amazing phenomenon.

#### 3.4. Origin of *Centaurium pannonicum*

As previously mentioned, the basic aim of this study was to set up the genetic and morphological frame for understanding and characterizing a specific case of the interspecific hybridization within the genus *Centaurium*. Our earlier study [32] had already established a solid basis to fathom this evolutionary mechanism when it comes to examine chemodiversity, as well as the difference in the ploidy levels of parental and hybridogenic taxa. The present study, which involved both morphometrics and different types of molecular markers, along with the previous research, strongly supports the hypothesis of allopolyploid origin for *C. pannonicum* at two locations (**PA** and **MJ**) and determines with high confidence its tetraploid progenitors (Scheme 1).

While the EST-SSR genotyping indicates closer genetic relationship of *C. erythraea* and *C. pannonicum*, sharing the same plastid marker haplotype by all the individuals of *C. pannonicum* and *C. littorale* subsp. *compressum* undoubtedly nominee *C. littorale* subsp. *compressum* as the other taxon that has participated in this interspecific hybridization process, the same being supported by the results of the phytochemical profiling [32]. Moreover, the results of the analysis of *trnL-F* region of the plastid DNA led us to propose the tetraploid taxon *C. littorale* subsp. *compressum* as the plastid donor (maternal progenitor) of the hexaploid hybrid *C. pannonicum*, and consequently, the tetraploid taxon *C. erythraea* as the pollen donor (Scheme 1). Nevertheless, it still remains to answer the question as to which of the two species involved in the process of interspecific hybridization contributed to the hexaploid *C. pannonicum* genome with four chromosomal sets and which with two. The closer relationship with *C. erythraea* revealed by EST-SSR genotyping implies that this parental species might contribute with

four chromosomal sets to the hexaploid hybrid genome (bold arrows in Scheme 1). This issue may open up the opportunities for future research which would include cytogenetic methods of counting, identification and differential staining of chromosomes, FISH (Fluorescent In Situ Hybridization) and GISH (Genomic In Situ Hybridization) methods that could return the final solution about the share of parental genomes in hybridogenic taxon [54,55].



**Scheme 1.** Proposed model for the formation of allohexaploid hybrids between tetraploid taxa *C. littorale* subsp. *compressum* and *C. erythraea*. Dashed arrows represent reduced, while full-line arrows represent non-reduced gametes that participate in the interspecific hybridization event. Plastid DNA is presented as either red or blue circle.

## 4. Materials and Methods

### 4.1. Sampling Sites

Plant materials were collected at the localities presented in Table 3 during June 2015 in the Subotica Sands, around Lake Palić and Fruška Gora Mt. (Vojvodina, Serbia) (Figures 6 and 7). Subotica Sands are protected area with sustainable use of natural resources (category VI according to IUCN), Nature Park Palić is classified as protected landscape (IUCN category V) while National Park Fruška Gora is classified in category II according to IUCN. Plants were identified in the field by the authors (B. Šiler, T. Banjanac and D. Mišić)-based either on available taxonomic keys [24] in the case of *C. erythraea* and *C. littorale* subsp. *compressum* or solely on morphological characters of habitus in the case of individuals classified as *C. pannonicum*. Their taxonomical status was investigated earlier by both flow cytometry and phytochemical profiling [32]. Ploidy levels of the collected plants have been previously published [32] and are given in Table 3. Corresponding voucher specimens were deposited in the herbarium of the University of Belgrade, Serbia (BEOU, Thiers, 2019; voucher numbers: 17271–17279).

## 4.2. Morphometric Analysis

### 4.2.1. Plant Material

A set of 127 individuals belonging to the three *Centaureium* taxa was included in the morphometric analyses. Immediately after plant harvesting, three randomly chosen flowers and the first three leaves above the rosette per plant were collected and preserved in a mixture of 35% ethanol and 20% glycerol (v:v = 3:1) for the comparative study of morphological characters (Figure S5). Samples were kept at room temperature until use.

### 4.2.2. Analysis of Morphological Traits

After careful consideration of the morphological characters suggested by Ubsdell [11] and the basic statistical analysis of the morphological parameters (Figure S1), six leaf and flower quantitative traits were selected, measured using Digimizer Image Analysis software (MedCalc Software, Belgium) and presented as the following characteristics: length/width ratio of floral stalk leaves, calyx/corolla-tube ratio, length of corolla lobes, and leaf tip angle which was decomposed to the sine and cosine components. Length/width ratio of floral stalk leaves was obtained by dividing the total length of a leaf by its greatest width. For each plant, the first three floral stalk leaves above the basal rosette were measured and a mean value was obtained. Calyx/corolla-tube ratio was obtained by dividing the total length of the calyx from the base to the tip of the teeth by the length of the corolla-tube. Length of corolla lobes was measured from the top of the corolla-tube to the tip of the corolla lobes. Three flowers per plant were scored and mean values were obtained (Table S2).

### 4.2.3. Statistical Analysis of the Morphometric Data

Morphometric data were analyzed with both principal component analysis (PCA) and linear discriminant analysis (LDA) using the R programming language packages 'stats', 'MASS' [56] and 'caret' [57]. PCA was performed on zero centered and unit scaled data. To evaluate the number of components that should be retained in PCA, Horn's parallel analysis [58] was performed using the R package 'paran' [59]. To assess the usability of the quantitative characters in predictive classification, LDA was assessed with five times repeated five-fold cross-validation using prediction accuracy on the hold out instances as metric. Furthermore, LDA was performed with a presumption of equal prior probability of classes. Feature selection for LDA was performed using a wrapper approach by sequential backwards search as implemented in 'caret' package.

## 4.3. EST-SSR Genotyping

### 4.3.1. Plant Material for the DNA Extraction

For the purposes of DNA extraction, young leaves were collected from the upper half of the flower stem and immediately stored in plastic zip-bags containing silica gel. Samples were kept in darkness, at the room temperature until use. In total, 136 individual plants were included in the ESR-SSR genotyping (Table 3).

### 4.3.2. DNA Isolation and EST-SSR Amplification

Total genomic DNA was isolated from dried leaf samples (20 mg leaf tissue) using a modified CTAB method [60]. DNA concentration and the purity of the isolates were assessed by the spectrophotometric absorbance at 260, 280, and 230 nm (Agilent 8453, Agilent Technologies, Waldbronn, Germany).

Primer pairs (Invitrogen, Thermo Fisher Scientific, Carlsbad, CA, USA) for amplification of eight EST-SSR loci (M4, M5, M7, M8, M10, M12, M13 and M17), previously published by Banjanac et al. [36], were employed to study the genetic variation within the sample set. PCR amplifications were conducted according to the same study as well. Fragment analysis of PCR products was performed using the Lab-on-a-Chip technology (DNA 1000 LabChip) on the Agilent Bioanalyzer 2100 system



(Agilent Technologies, Santa Clara, CA, USA). PCR products' peaks on electropherograms were aligned using Agilent 2100 Expert software (Agilent Technologies, Santa Clara, CA, USA) as described in Banjanac et al. [36]. Alleles were scored manually.

#### 4.3.3. Genetic Diversity of Populations Based on the EST-SSR Variation

Data obtained by eight EST-SSR molecular markers of both tetraploid and hexaploid plants were first analyzed using the package 'polysat' [61] in R programming language [62]. Distances between genotypes were calculated by the method described by Bruvo et al. [63] using the 'meandistance.matrix2' function in 'polysat' with 'Bruvo2.distance' distance metrics (equation 6 in [63]). The basic settings within the 'polysat' package included presumptions such as: polysomatic way of inheritance for all the populations, corresponding ploidy of populations (tetraploid/hexaploid, according to Banjanac et al. [32]), selfing rate of 0.56 (according to Brys and Jacquemyn [38]) and the length of microsatellite repeats for each locus (see in [36]). The genetic variability was estimated using the following parameters: the total number of alleles in a population, the average number of alleles per locus per population, the observed ( $H_O$ ) and the expected heterozygosity ( $H_E$ ), and the effective number of alleles in a population based on the expected heterozygosity.  $H_O$  of a population was obtained by counting heterozygous individuals, while  $H_E$  was calculated on the basis of allele frequencies generated by 'polysat' using formula:

$$H_E = 1 - \sum p_i^4 \text{ for tetraploid populations and}$$

$$H_E = 1 - \sum p_i^6 \text{ for hexaploid populations,}$$

where  $p$  is the frequency of the  $i^{\text{th}}$  allele of a given locus.

Given genetic data, allele frequencies by population were calculated using the 'simpleFreq' function in 'polysat' package for R, assuming polysomic inheritance. In the following step, using the data frame of allele frequencies and population sizes, we calculated a matrix of pairwise  $F_{ST}$  and Jost's  $D$  values using the function 'calcPopDiff' of the same R package.

#### 4.3.4. Visualization of the Genetic Distances

The obtained Bruvo genetic distances among the individuals were visualized using two-dimensional nonmetric multidimensional scaling (NMDS) which was performed employing the package 'vegan' for R [64].

Further representation of relationships within the sample set included hierarchical cluster analysis (HCA), using UPGMA method. The dendrogram was constructed on the basis of the Bruvo genetic distances between individuals, which were obtained from the allele frequencies at eight EST-SSR loci, calculated by 'polysat' package. Assessment of the clusterwise stability of the clustering was performed by bootstrapping using the R package 'fpc' [65].

### 4.4. *trnL-F* Haplotype Analysis

#### 4.4.1. Sample Set

A subset of 43 DNA isolates (Table 3) was used to assess the variation in the plastid genome through *trnL-F* haplotyping. Analysis of the variation in plastid *trnL-F* region also included 16 GenBank accessions of the taxa of interest previously reported by other authors (Table 4).

#### 4.4.2. Plastid DNA Amplification and Marker Sequencing

Two non-coding regions of plastid DNA, *trnL* UAA intron and the *trnL-trnF* intergenic spacer were amplified in the same reaction using primers reported by Taberlet et al. [46]: "f" (ATT TGA ACT GGT GAC ACG AG) and "c" (CGA AAT CGG TAG ACG CTA CG). PCR amplifications were carried out using Eppendorf Mastercycler nexus gradient thermal cycler (Eppendorf AG, Hamburg, Germany)

in a final volume of 25  $\mu$ l. Each reaction contained 100 ng of the template DNA, DreamTaq Green PCR Master Mix (according to the manufacturer's instructions; Thermo Fisher Scientific, Waltham, MA, USA) and 0.5  $\mu$ M of both forward and reverse primers (Invitrogen<sup>®</sup>, Thermo Fisher Scientific, Carlsbad, CA, USA). The PCR reaction program used for the amplification of this combined plastid non-coding regions was as follows: 94 °C for 10 min; 35 cycles of 94 °C for 1 min, 55 °C for 45 s, 72 °C for 1.5 min, and 72 °C for 10 min as a final extension step.

Amplified PCR products were purified using off-the-shelf absorption, washing and elution DNA buffers, and EconoSpin DNA spin columns according to the manufacturer's instructions (<http://www.epochlifescience.com/>). They were run on a 1% agarose gel stained with ethidium bromide, in order to evaluate the quality and quantity of the amplified templates.

Direct sequencing was performed on an ABI 3130 Genetic Analyzer (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, USA), using BigDye Terminator v. 3.1 Cycle Sequencing Kit (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, USA) for cycle sequencing reactions following the manufacturer's instructions. All samples were sequenced with "f" primer. The obtained results were visualized and analyzed using the Sequencing Analysis Software 6 (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, USA) and BioEdit (Ibis Biosciences, Carlsbad, CA, USA). Sequences alignment was done with MUSCLE program (<https://www.ebi.ac.uk/Tools/msa/muscle/>). Selection of conserved blocks from the alignment was performed using Gblocks v0.91b [66] with the following options: minimum number of sequences for a conserved position was set to 30, minimum number of sequences for a flank position was set to 50, maximum number of contiguous non-conserved positions was set to 8, minimum length of a block was set to 10, while the allowed gap positions were set to half.

The resulting alignment was visualized with FastTree v2.1.10 [67] using GTR model (generalized time-reversible model [68]) of evolution. Phylogenetic tree was rooted by *Centaureum pulchellum*. Local support values were estimated using the Shimodaira-Hasegawa test on the three alternate topologies around the specific split [67]. A total of 43 sequences were generated and submitted to GenBank (accession numbers MN104899–MN104941).

## 5. Conclusions

Plenty of evidence reveals the importance of both polyploidization and interspecific hybridization as evolutionary mechanisms responsible for speciation within the genus *Centaureum* [8,10,34,51]. We hope that the results of the presented research will contribute to the clarification of the phylogenetic relationships within the genus but also to the understanding of the processes of polyploidization and interspecific hybridization and their roles in comprehending the concept of plant species. In the presented case, we are witnessing the generation of evolutionary novelty and the promotion of expansion of the allopolyploid hybridogenic taxon, possibly endowed with new or improved adaptive characteristics, which might increase its evolutionary potential. This is extremely important if we consider rapid climate changes, but also severe anthropogenic intrusion in most of the known *C. pannonicum* habitats. In the end, but certainly not less important, is the unavoidable need to design and suggest a conservation strategy for new hybridogenic taxon *C. pannonicum* by evaluating its genetic diversity and the evolutionary potential of parental species in the zones where the hybrids may arise, as well as to define the current status of the habitats and to recognize possible threats. As a means of ex situ conservation strategy, we have established a seed collection representing the genetic diversity of *C. pannonicum* in its natural populations, which will enable the long-term preservation of its germplasm. On the other hand, as a means of in situ conservation, we may propose raising the level of protection of the sites and the prohibition of further habitat disturbance by humans.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2223-7747/8/7/224/s1>, Table S1: Confusion matrix depicting the performance of the LDA model in taxon discrimination. The model was constructed using the following morphometric features: length/width ratio of floral stalk leaves, calyx/corolla-tube ratio, length of corolla lobes, and cosine of the leaf tip angle. Performance was estimated based on the hold out set predictions in 10 times repeated 5-fold cross validation (entries are presented as percent average cell counts across resamples), Table S2: Raw morphometric data. Figure S1: Box-plot diagrams depicting distributions of

measurements of morphological traits in three taxa of interest. Measurements are in mm except for leaf veins number (discontinuous values 3 or 5) and leaf tip angle which are in degrees of angle. Figure S2: Box-plot diagrams depicting distributions of length/width ratio of floral stalk leaves of *C. littorale* subsp. *compressum* and *C. pannonicum* on two localities (**MJ** and **PA**). Measurements are in mm, Figure S3: Box-plot diagrams depicting distributions of length of floral stalk leaves of *C. littorale* subsp. *compressum* and *C. pannonicum* on two localities (**MJ** and **PA**). Measurements are in mm. Figure S4: Box-plot diagrams depicting distributions of width of floral stalk leaves of *C. littorale* subsp. *compressum* and *C. pannonicum* on two localities (**MJ** and **PA**). Measurements are in mm. Figure S5: Representative photographs of analyzed leaves and flowers belonging to the three studied taxa.

**Author Contributions:** Conceptualization, B.Š., T.B. and D.M.; Methodology, T.B., S.Đ., M.J. and M.D.; Validation, T.B., M.S., B.Š. and M.D.; Investigation, T.B., J.N.Ž., S.Đ. and B.Š.; Formal Analysis, T.B., S.Đ. and M.D.; Data Curation, T.B., B.Š., M.J. and S.Đ.; Writing—Original Draft Preparation, T.B. and B.Š.; Writing—Review & Editing, D.M., S.Đ., M.S., J.N.Ž., M.D. and M.J.; Visualization, T.B., M.D. and S.Đ.; Supervision, B.Š. and D.M.; Funding Acquisition, D.M.

**Funding:** This research was funded by the Serbian Ministry of Education, Science and Technological Development, grant numbers OI173024 and OI173030. The APC was funded by B.Š.

**Acknowledgments:** The authors are grateful to Dmtar Lakušić, University of Belgrade, for their professional advices.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Soltis, P.S.; Soltis, D.E. The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* **2009**, *60*, 561–588. [[CrossRef](#)] [[PubMed](#)]
- Rieseberg, L.H.; Willis, J.H. Plant speciation. *Science* **2007**, *317*, 910–914. [[CrossRef](#)] [[PubMed](#)]
- Barker, M.S.; Arrigo, N.; Baniaga, A.E.; Li, Z.; Levin, D.A. On the relative abundance of autopolyploids and allopolyploids. *New Phytol.* **2016**, *210*, 391–398. [[CrossRef](#)] [[PubMed](#)]
- Mallet, J. Hybrid speciation. *Nature* **2007**, *446*, 279–283. [[CrossRef](#)] [[PubMed](#)]
- Qiao, X.; Li, Q.; Yin, H.; Qi, K.; Li, L.; Wang, R.; Zhang, S.; Paterson, A.H. Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biol.* **2019**, *20*, 38. [[CrossRef](#)] [[PubMed](#)]
- Alix, K.; Gérard, P.R.; Schwarzacher, T.; Heslop-Harrison, J.S. Polyploidy and interspecific hybridization: Partners for adaptation; speciation and evolution in plants. *Ann. Bot.* **2017**, *120*, 183–194. [[CrossRef](#)] [[PubMed](#)]
- Zeltner, L. Recherches de biosystématique sur les genres *Blackstonia* Huds. et *Centaurium* Hill (Gentianaceae). *Bull. Soc. Neuch. Sci. Nat.* **1970**, *93*, 1–164.
- Mansion, G.; Zeltner, L.; Bretagnolle, F. Phylogenetic Patterns and Polyploid Evolution within the Mediterranean Genus *Centaurium* (Gentianaceae–Chironieae). *Taxon* **2005**, *54*, 931–950. [[CrossRef](#)]
- Banjanac, T.; Šiler, B.; Skorić, M.; Ghalawenji, N.; Milutinović, M.; Božić, D.; Mišić, D. Interspecific in vitro hybridization in genus *Centaurium* and identification of hybrids via flow cytometry, RAPD, and secondary metabolite profiles. *Turk. J. Bot.* **2014**, *38*, 68–79. [[CrossRef](#)]
- Guggisberg, A.; Bretagnolle, F.; Mansion, G. Allopolyploid Origin of the Mediterranean Endemic, *Centaurium bianoris* (Gentianaceae), Inferred by Molecular Markers. *Syst. Bot.* **2006**, *31*, 368–379. [[CrossRef](#)]
- Ubsdell, R.A.E. Studies on variation and evolution in *Centaurium erythraea* Rafn and *C. littorale* (D. Turner) Gilmour in the British Isles 1. Taxonomy and biometrical studies. *Watsonia* **1976**, *11*, 7–31.
- Ubsdell, R.A.E. Studies on variation and evolution in *Centaurium erythraea* Rafn and *C. littorale* (D. Turner) Gilmour in the British Isles 2. Cytology. *Watsonia* **1976**, *11*, 33–43.
- Ubsdell, R.A.E. Studies on variation and evolution in *Centaurium erythraea* Rafn and *C. littorale* (D. Turner) Gilmour in the British Isles 3. Breeding systems, floral biology and general discussion. *Watsonia* **1979**, *12*, 225–232.
- Grisebach, A. *Genera et Species Gentianearum Adjectis Observationibus Quibusdam Phytogeographicis*; J. G. Cotta: Stuttgart, Germany, 1839.
- Gilg, E. Gentianaceae. In *Die Natürlichen Pflanzenfamilien*; Engler, A., Prantls, K., Eds.; Verlag von Wilhelm Engelmann: Leipzig, Germany, 1895; pp. 50–108.
- Melderis, A. Genetical and taxonomical studies in the genus *Erythraea* Rich. *Acta Horti. Bot. Univ. Latv.* **1931**, *6*, 123–256.

17. Robyns, A. Essai d'étude systématique et écologique des *Centaurium* de Belgique. *Bull. Jard. Bot. l'Etat Brux.* **1954**, *24*, 349–398. [CrossRef]
18. Hegi, G. Gentianaceae. In *Illustrierte Flora von Mittel-Europa*; Carl Hanser Verlag: Munich, Germany, 1966; pp. 1953–2047.
19. Euro+Med Plantbase Project. Available online: <http://ww2.bgbm.org/euroPlusMed/query.asp> (accessed on 8 May 2019).
20. Melderis, A. Taxonomic studies on the European species of the genus *Centaurium* Hill. *Bot. J. Linn. Soc.* **1972**, *65*, 224–250.
21. Pringle, J.S. Proposal to reject the name *Centaurium minus* Moench (Gentianaceae). *Taxon* **1987**, *36*, 482–483. [CrossRef]
22. Pringle, J.S. The identity and nomenclature of the Pacific North American species *Zeltnera muhlenbergii* (Gentianaceae) and its distinction from *Centaurium tenuiflorum* and other species with which it has been confused. *Madroño* **2010**, *57*, 184–202. [CrossRef]
23. Anderson, E. Hybridization of the habitat. *Evolution* **1948**, *2*, 1–9. [CrossRef]
24. Mansion, G. A new classification of the polyphyletic genus *Centaurium* Hill (Chironiinae; Gentianaceae): Description of the New World endemic *Zeltnera*, and reinstatement of *Gyrandra* Griseb. and *Schenkia* Griseb. *Taxon* **2004**, *53*, 719–740. [CrossRef]
25. Mészáros, S.; De Laet, J.; Goethals, V.; Smets, E.; Nilsson, S. Cladistics of Gentianaceae: A morphological approach. In *Gentianaceae—Systematics and Natural History*; Struwe, L., Albert, V., Eds.; Cambridge University Press: London, UK, 2002; pp. 310–376.
26. Nilsson, S. Gentianaceae: A review of palynology. In *Gentianaceae—Systematics and Natural History*; Struwe, L., Albert, V., Eds.; Cambridge University Press: London, UK, 2002; pp. 377–497.
27. Jensen, S.R.; Schripsema, J. Chemotaxonomy and pharmacology of Gentianaceae. In *Gentianaceae—Systematics and Natural History*; Struwe, L., Albert, V., Eds.; Cambridge University Press: London, UK, 2002; pp. 573–631.
28. Šiler, B.; Mišić, D.; Nestorović, J.; Banjanac, T.; Glamočlija, J.; Soković, M.; Ćirić, A. Antibacterial and Antifungal Screening of *Centaurium pulchellum* Crude Extracts and Main Secoiridoid Compounds. *Nat. Prod. Commun.* **2010**, *5*, 1525–1530. [CrossRef] [PubMed]
29. Šiler, B.; Avramov, S.; Banjanac, T.; Cvetković, J.; Nestorović Živković, J.; Patenković, A.; Mišić, D. Secoiridoid glycosides as a marker system in chemical variability estimation and chemotype assignment of *Centaurium erythraea* Rafn from the Balkan Peninsula. *Ind. Crops Prod.* **2012**, *40*, 336–344. [CrossRef]
30. Šiler, B. Phytochemical and Molecular Characterization of Centaury (*Centaurium erythraea* Rafn) Populations from the Balkan Peninsula. Ph.D. Thesis, University of Belgrade, Belgrade, Serbia, 5 June 2012.
31. Šiler, B.; Živković, S.; Banjanac, T.; Cvetković, J.; Nestorović Živković, J.; Ćirić, A.; Soković, M.; Mišić, D. Centauries as underestimated food additives: Antioxidant and antimicrobial potential. *Food Chem.* **2014**, *147*, 367–376. [CrossRef] [PubMed]
32. Banjanac, T.; Dragičević, M.; Šiler, B.; Gašić, U.; Bohanec, B.; Nestorović Živković, J.; Trifunović, S.; Mišić, D. Chemodiversity of two closely related tetraploid *Centaurium* species and their hexaploid hybrid: Metabolomic search for high-resolution taxonomic classifiers. *Phytochemistry* **2017**, *140*, 27–44. [CrossRef] [PubMed]
33. Patwardhan, A.; Ray, S.; Roy, A. Molecular Markers in Phylogenetic Studies—A Review. *J. Phylogenet. Evol. Biol.* **2014**, *2*, 131. [CrossRef]
34. Mansion, G.; Struwe, L. Generic delimitation and phylogenetic relationships within the subtribe Chironiinae (Chironieae: Gentianaceae), with special reference to *Centaurium*: Evidence from nrDNA and cpDNA sequences. *Mol. Phylogenet. Evol.* **2004**, *32*, 951–977. [CrossRef] [PubMed]
35. Marhold, K. Gentianaceae. In *Euro+ Med Plantbase—The Information Resource for Euro-Mediterranean Plant Diversity*; 2011; Available online: <http://ww2.bgbm.org/EuroPlusMed/PTaxonDetail.asp?NameCache=Gentianaceae&PTRefFk=7200000> (accessed on 13 July 2019).
36. Banjanac, T.; Skorić, M.; Belamarić, M.; Nestorović Živković, J.; Mišić, D.; Jelić, M.; Dmitrović, S.; Šiler, B. Pursuit for EST microsatellites in a tetraploid model from de novo transcriptome sequencing. *Genetika* **2018**, *50*, 687–703. [CrossRef]
37. Mansion, G. Twenty-first Century Centauries: An Updated Review on *Centaurium* Hill and Allies (Gentianaceae). In *The Gentianaceae—Volume 1: Characterization and Ecology*; Rybczyński, J.J., Davey, M.R., Mikula, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2014; pp. 231–250.

38. Brys, R.; Jacquemyn, H. Variation in the functioning of autonomous self-pollination, pollinator services and floral traits in three *Centaurium* species. *Ann. Bot.* **2011**, *107*, 917–925. [[CrossRef](#)]
39. Jovanović-Dunjić, R. Fam. *Gentianaceae*. In *Flora SR Srbije*; Srpska Akademija Nauka i Umetnosti: Belgrade, Yugoslavia, 1973; pp. 403–432.
40. Dawson, D.A.; Horsburgh, G.J.; Küpper, C.; Stewart, I.R.K.; Ball, A.D.; Durrant, K.L.; Hansson, B.; Bacon, I.; Bird, S.; Klein, A.; et al. New methods to identify conserved microsatellite loci and develop primer sets of high cross-species utility—As demonstrated for birds. *Mol. Ecol. Resour.* **2010**, *10*, 475–494. [[CrossRef](#)]
41. Dawson, D.A.; Ball, A.D.; Spurgin, L.G.; Martín-Gálvez, D.; Stewart, I.R.K.; Horsburgh, G.J.; Potter, J.; Molina-Morales, M.; Bicknell, A.W.J.; Preston, S.A.J.; et al. High-utility conserved avian microsatellite markers enable parentage and population studies across a wide range of species. *BMC Genom.* **2013**, *14*, 176. [[CrossRef](#)]
42. Chassot, P.; Nemomissa, S.; Yuan, Y.-M.; Küpfer, P. High paraphyly of *Swertia* L. (*Gentianaceae*) in the *Gentianella*-lineage as revealed by nuclear and chloroplast DNA sequence variation. *Plant Syst. Evol.* **2001**, *229*, 1–21. [[CrossRef](#)]
43. Gielly, L.; Taberlet, P. A phylogeny of the European gentians inferred from chloroplast *trnL* (UAA) intron sequences. *Bot. J. Linn. Soc.* **1996**, *120*, 57–75. [[CrossRef](#)]
44. Struwe, L.; Thiv, M.; Kadereit, J.W.; Pepper, A.S.R.; Motley, T.J.; White, P.J.; Rova, J.H.E.; Potgieter, K.; Albert, V.A. *Saccifolium* (*Saccifoliaceae*), an endemic of Sierra de la Neblina on the Brazilian-Venezuelan border, is related to a temperate-alpine lineage of *Gentianaceae*. *Harv. Pap. Bot.* **1998**, *3*, 199–214.
45. Yuan, Y.-M.; Wohlhauser, S.; Möller, M.; Chassot, P.; Mansion, G.; Grant, J.; Küpfer, P.; Klackenberg, J. Monophyly and relationships of the tribe Exaceae (*Gentianaceae*) inferred from nuclear ribosomal and chloroplast DNA sequences. *Mol. Phylogenet. Evol.* **2003**, *28*, 500–517. [[CrossRef](#)]
46. Taberlet, P.; Gielly, L.; Pautou, G.; Bouvet, J. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* **1991**, *17*, 1105–1109. [[CrossRef](#)] [[PubMed](#)]
47. Kieleczawa, J. Fundamentals of sequencing of difficult templates—An overview. *J. Biomol. Tech.* **2006**, *17*, 207–217. [[PubMed](#)]
48. Corriveau, J.L.; Coleman, A.W. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *Am. J. Bot.* **1988**, *75*, 1443–1458. [[CrossRef](#)]
49. Tsitrone, A.; Kirkpatrick, M.; Levin, D.A. A model for chloroplast capture. *Evolution* **2003**, *57*, 1776–1782. [[CrossRef](#)]
50. Daniell, H.; Lin, C.S.; Yu, M.; Chang, W.J. Chloroplast genomes: Diversity; evolution; and applications in genetic engineering. *Genome Biol.* **2016**, *17*, 134. [[CrossRef](#)]
51. Çiçek, M.; Yaprak, A.E.; Alan, A.R. Morphometric and flow cytometric evaluations of a putative natural hybrid of *Centaurium* (*Gentianaceae*) from Turkey. *Phytotaxa* **2015**, *204*, 22–32. [[CrossRef](#)]
52. Zeltner, L. Recherches sur le *Centaurium bianoris* (Sennen) Sennen. *Rev. Biol. Ecol. Médit.* **1978**, *5*, 51–58.
53. Soltis, D.E.; Kuzoff, R.K. Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (*Saxifragaceae*). *Evolution* **1995**, *49*, 727–742. [[CrossRef](#)] [[PubMed](#)]
54. Bennett, S.T.; Kenton, A.Y.; Bennett, M.D. Genomic *in situ* hybridization reveals the allopolyploid nature of *Milium montianum* (*Gramineae*). *Chromosoma* **1992**, *101*, 420–424. [[CrossRef](#)]
55. Hegarty, M.J.; Hiscock, S.J. Hybrid speciation in plants: New insights from molecular studies. *New Phytol.* **2005**, *165*, 411–423. [[CrossRef](#)] [[PubMed](#)]
56. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S*, 4th ed.; Springer: New York, NY, USA, 2002.
57. Kuhn, M.; Wing, J.; Weston, S.; Williams, C.A.; Keefer, C.; Engelhardt, A.; Cooper, T.; Mayer, Z.; Kenkel, B.; Benesty, M.; et al. ‘caret’: Classification and Regression Training. 2018. Available online: <https://cran.r-project.org/web/packages/caret/caret.pdf> (accessed on 13 July 2019).
58. Horn, J.L. A rationale and test for the number of factors in factor analysis. *Psychometrika* **1965**, *30*, 179–185. [[CrossRef](#)] [[PubMed](#)]
59. Dinno, A. ‘paran’, Horn’s Test of Principal Components/Factors. R Project. 2018. Available online: <https://cran.r-project.org/web/packages/paran/paran.pdf> (accessed on 13 July 2019).
60. Doyle, J.J.; Doyle, J.L. Isolation of plant DNA from leaf tissue. *Focus* **1990**, *12*, 13–15.
61. Clark, L.V.; Jasieniuk, M. POLYSAT: An R package for polyploid microsatellite analysis. *Mol. Ecol. Resour.* **2011**, *11*, 562–566. [[CrossRef](#)] [[PubMed](#)]

62. R Core Team: The R Project for Statistical Computing. 2018. Available online: <https://www.r-project.org/> (accessed on 13 July 2019).
63. Bruvo, R.; Michiels, N.K.; D'Souza, T.G.; Schulenburg, H. A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Mol. Ecol.* **2004**, *13*, 2101–2106. [[CrossRef](#)]
64. Oksanen, J.; Blanchet, F.G.; Friendly, M.; Kindt, R.; Legendre, P.; McGlenn, D.; Minchin, P.R.; O'Hara, R.B.; Simpson, G.L.; Solymos, P.; et al. 'vegan': Community Ecology Package. 2019. Available online: <https://cran.r-project.org/web/packages/vegan/vegan.pdf> (accessed on 13 July 2019).
65. Hennig, C. 'fpc': Flexible Procedures for Clustering. 2019. Available online: <https://cran.r-project.org/web/packages/fpc/fpc.pdf> (accessed on 13 July 2019).
66. Talavera, G.; Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **2007**, *56*, 564–577. [[CrossRef](#)]
67. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLoS ONE* **2010**, *5*, e9490. [[CrossRef](#)]
68. Tavaré, S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect. Math. Life Sci.* **1986**, *17*, 57–86.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).