

## Essential title page information

### Title:

New chloroplast microsatellites in *Helichrysum italicum* (Roth) G. Don, their characterization and application for the evaluation of genetic resources

### Author names and affiliations:

Matjaž Hladnik<sup>a</sup>, Alenka Baruca Arbeiter<sup>a</sup>, Félix Tomi<sup>b</sup>, Marc Gibernau<sup>b</sup>, Slavko Brana<sup>c</sup>, Dunja Bandelj<sup>a,\*</sup>

<sup>a</sup>University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technologies, Glagoljaška 8, 6000, Koper, Slovenia

<sup>b</sup>Laboratoire Sciences Pour l'Environnement, Université de Corse-CNRS, UMR 6134 SPE, Route des Sanguinaires, 20000 Ajaccio, France

<sup>c</sup>Istrian Botanical Society, Trgovačka 45, HR-52215 Vodnjan, Croatia

### E-mail addresses:

Matjaž Hladnik: [matjaz.hladnik@upr.si](mailto:matjaz.hladnik@upr.si), Alenka Baruca Arbeiter: [alenka.arbeiter@upr.si](mailto:alenka.arbeiter@upr.si), Félix Tomi: [tomi\\_f@univ-corse.fr](mailto:tomi_f@univ-corse.fr), Marc Gibernau: [gibernau\\_m@univ-corse.fr](mailto:gibernau_m@univ-corse.fr), Slavko Brana: [istra.botanica@gmail.com](mailto:istra.botanica@gmail.com)

**\*Corresponding author:** University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technologies, Glagoljaška 8, 6000 Koper, Slovenia

E-mail address: [dunja.bandelj@upr.si](mailto:dunja.bandelj@upr.si) (Dunja Bandelj)

## Title

New chloroplast microsatellites in *Helichrysum italicum* (Roth) G. Don, their characterization and application for the evaluation of genetic resources

## Highlights

1. Ten new polymorphic cpSSRs have been confirmed as reliable for genotyping.
2. The diversity of chloroplast haplotypes was detected within and between populations.
3. Seven of nine Kamenjak (Croatia) haplotypes are unique compared to those in Corsica (France).
4. New cpSSRs can differentiate *H. italicum*, *H. arenarium* and *H. litoreum* species.
5. Low polymorphism and mutations in flanking regions, limited some loci for analysis.

## Abstract

*Helichrysum italicum* ((Roth) G. Don is a famous Mediterranean medicinal plant that offers great potential in the field of health care due to the unique properties of its bioactive compounds in various extracts. The plant has only recently been introduced into agroecosystems, but the planting material is usually produced by vegetative propagation of naturally growing populations or by harvesting seeds and preparing seedlings. However, due to the great morphological, genetic and chemical diversity in natural populations, it is difficult to select plants with suitable characteristics for different industries. Although *H. italicum* plant material is important for various industries, only a very limited number of genomic SSR markers are available. In the present study, primers were designed for all 43 SSRs identified in the chloroplast genome of *H. italicum*. Mononucleotide SSR repeats were the most abundant (72.1%), followed by nine dinucleotide repeats (21%) and three trinucleotide repeats (6.9%). Populations from Cape Kamenjak (Istrian peninsula, Croatia) and the island of Corsica (France) were analysed with ten carefully selected cpSSR markers. From the originally selected set of 16 cpSSRs amplified in all samples, 6 cpSSR markers were removed due to low polymorphism, mutations in the flanking regions that could not be detected with allele length or size homoplasy. Of the 38 haplotypes detected, 32 were private to geographic origin. Cross-species transferability with *Helichrysum litoreum* Guss. and *Helichrysum arenarium* (L.) Moench was successfully tested and private alleles were found. In summary, newly developed cpSSR markers, alone or in combination with already published genomic SSRs, have great potential for use in plant breeding and for other purposes requiring genotype identification.

**Keywords:** immortelle, *Helichrysum italicum* (Roth) G. Don, chloroplast genome, microsatellite markers, alleles sequencing, genetic diversity, haplotypes

## Acronyms and abbreviations:

cpSSR: chloroplast simple sequence repeat

## 1. INTRODUCTION

The immortelle (*Helichrysum italicum* (Roth) G. Don), is an aromatic plant of the Asteraceae family, widely distributed in natural vegetation of the central Mediterranean Basin. This plant is appreciated for its essential oil and other plant extracts possessing high valued, biological active compounds used in pharmaceutical, cosmetic, perfumery and food industries. Several studies have reported and confirmed antimicrobial, anti-inflammatory, antioxidant, anti-viral, anticancer, insecticidal, anti-ageing and repellent activities (Antunes Viegas et al., 2014; Appendino et al., 2015; Kramberger et al., 2021; Maksimovic et al., 2017; Ninčević et al., 2019). These beneficial properties of immortelle extracts have led to the increased interest for its cultivation and domestication in some Mediterranean countries, and opened the need for a more accurate characterization of its genetic resources.

Taxonomy of *Helichrysum italicum* (Roth) G. Don is complicated and was revised several times in the past (Galbany-Casals et al., 2006, 2011). The most recent multidisciplinary revision of this species divided it into four subspecies (subsp. *italicum*, *tyrrhenicum*, *microphyllum* (Willd.), *siculum* (Jord. & Fourr.)) distributed through different regions of the Mediterranean Basin (Herrando-Moraira et al., 2016). Their morphological features and geographical distribution are well described by Galbany-Casals et al. (2006) and Herrando-Moraira et al. (2016). It should be noted that within populations from a small geographical area, a high diversity of individual plants has been observed (Tomičić et al., 2022), which is accompanied by a remarkable phenotypic plasticity that also leads to different chemotypes, mostly as a result of adaptation to diverse ecological factors in the habitats where it grows and also due to distinct genetics (Galbany-Casals et al., 2006; Bianchini et al., 2009). For example, the essential oil of Tyrrhenian immortelle is characterized by a high content of neryl esters, the most distinctive constituents valued in the perfumery industry and aromatherapy. Essential oils from Balkan and Adriatic coast, on the other hand, contain mainly pinenes and high curcumene content (reviewed by Ninčević et al. (2019)). Thus, the essential oil of Corsican origin is reaching a higher price on the market (Appendino et al., 2015). Therefore, it can be expected that some countries will decide to develop the quality certification Protected Geographical Indication (PGI) to prevent uncontrolled exportation of the plant from their native growing sites and to protect the quality/typicality of immortelle essential oil. These initiatives will probably contribute to the stricter control of the planting material offered on the market, and to a higher price of raw material.

Natural growing sites of some subspecies and species overlap, and their spontaneous hybridization has created new hybrids with intermediate morphological characters (Galbany-Casals et al., 2006). An example of this are hybrids between *Helichrysum litoreum* Guss. (avowed substitute for *Helichrysum angustifolium* (Lam.) DC (Aghababayan et al., 2009; Baruca Arbeiter et al., 2021; Hladnik et al., 2023)) and *H. italicum* subsp. *italicum*. These two species share natural habitats in west-central Italy, some islands in the Tyrrhenian sea and the Croatian Istra Peninsula (Galbany-Casals et al., 2006).

Plant genetic resources used for plantations are collected from natural populations. However, the morphological diversity and chemotype variation of *H. italicum* make it difficult to accurately identify plant species and determine the most promising plants for cultivation based on morphological traits and chemical analyses only. The observed differentiations of chemotypes are especially interesting for a diversified industrial use of the plant (Bianchini et al., 2009; Morone-Fortunato et al., 2010). In addition, the high price of immortelle essential oil on the market contributes to the intentional adulteration, and the gap between availability and demand for immortelle essential oil is often filled with plant material from other species of questionable chemotype and origin (Appendino

et al., 2015). Adulteration is the main problem for both the industry and consumers, as it jeopardizes the quality of immortelle products. Therefore, mechanisms for the identification and traceability of genetic resources on the market are necessary.

The characterization of plant genetic resources with molecular markers is therefore important not only for material traceability but also: a) to understand genetic diversity within and between populations from geographically distinct regions with defined ecological factors for the development of a systematic plan for field examination; b) for breeding programs, for the selection or development of cultivars with superior traits and also for controlled propagation of planting material with high performance for industrial use; c) to establish a solid basis for further studies to link genotypes to the chemical composition of extracts or metabolome diversity and consequently to pharmacology.

In comparison to other aromatic plants, the genomic research of immortelle is lagging far behind. Only recently have the first twenty-four nuclear microsatellite markers been developed from genomic sequences of the Adriatic immortelle (Baruca Arbeiter et al., 2021). However, the first genomic sequences, obtained with high throughput sequencing technology enabled the assembly of immortelle chloroplast genome (Hladnik et al., 2019) and provided the opportunity to develop new chloroplast-based markers.

Sequence analyses of chloroplast genomes have contributed significantly to the phylogenetic studies of several plant families and have revealed considerable variation within and between plant species (Daniell et al., 2016), providing a valuable source for genetic markers development. Traditionally, for authentication and identification of plants barcodes *matK* and *rbcL* from two chloroplast coding regions have been suggested as standard markers (CBOL Plant Working Group et al., 2009), which could be additionally supported with other barcodes, such as *trnH-psbA*, plastid intergenic spacer region and nuclear ribosomal ITS (Nazar et al., 2022). Chloroplast intergenic region *rpl32-trnL* was used for the systematics and phylogeographic study of the Mediterranean *Helichrysum pendulum* complex and several cpDNA haplotypes were shared by different species of this complex (Herrando-Moraira et al., 2017).

Chloroplast SSRs (cpSSRs) have been recognized as powerful markers due to their polymorphism and ability to detect intraspecific diversity, as well as the fact that they inherit the characteristics of the chloroplast genome, such as uniparental inheritance, non-recombinant nature and haploid state (Provan et al., 2001). The latter could provide information about the dispersal of seeds (if chloroplasts are inherited maternally) or pollen (if they are inherited paternally). In addition, cpSSRs are more effective indicators for population subdivision and differentiation compared to nuclear markers and the combination of cpSSRs with nuclear SSRs can further enrich various genetic studies (Ebert & Peakall, 2009).

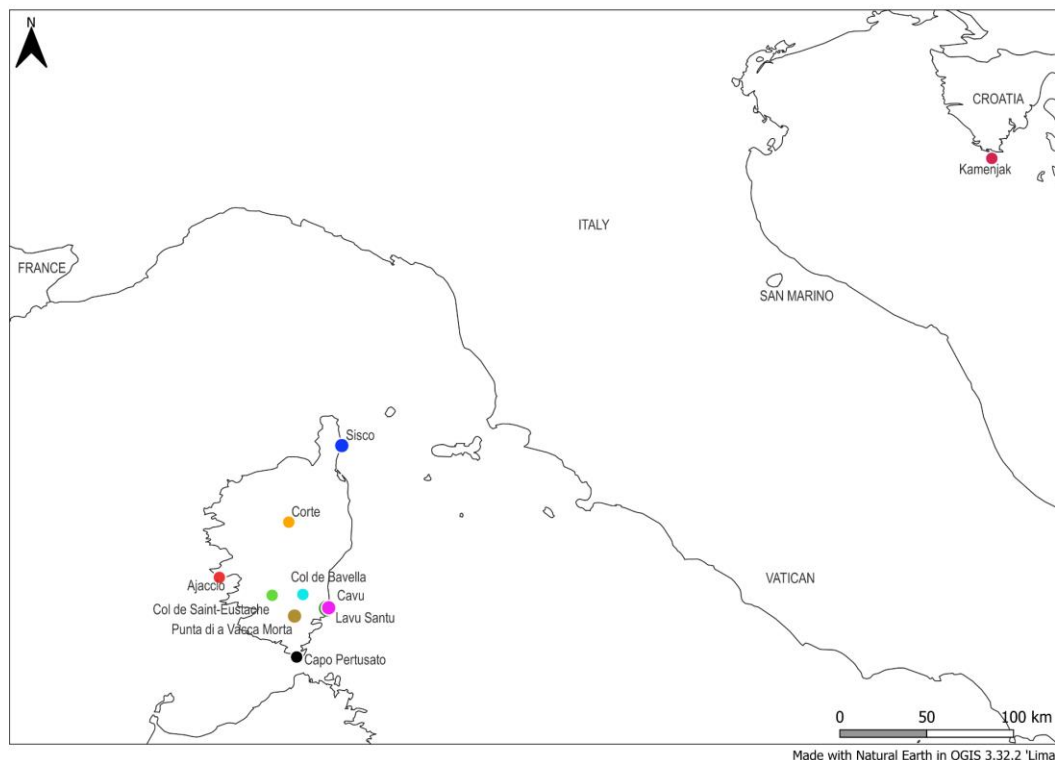
The present study focuses on the development of new cpSSR markers derived from the sequenced chloroplast genome of *H. italicum*. The newly developed markers were characterized in different immortelle populations from Corsica and the northeastern Adriatic coast, which represent important genetic resources for essential oil production and transfer of plants to agricultural cultivation. The transfer of wild populations from nature to the agroecosystem and further industrial utilization require the control of the origin of the planting material, which is why new cpSSRs represent an important molecular tool. The usefulness of cpSSRs for the unambiguous differentiation of three *Helichrysum* species (*H. italicum*, *H. litoreum*, *H. arenarium* (L.) Moench) was also tested.

## **2. MATERIAL AND METHODS**

### **2.1 Plant material and DNA extraction**

For the characterization of newly developed chloroplast SSR markers, populations of *H. italicum*, which can potentially serve as a source for the production of planting material, were sampled. *H. italicum* samples from Croatia were obtained from the southernmost cape of the Istrian peninsula in the northeastern Adriatic region, while other populations were sampled in France (Corsica) (Fig. 1, Table 1). For a cross-species transferability test, commercially available *H. arenarium* dry tea, and samples of *H. litoreum* obtained from certified seed (purchased as *H. italicum* seeds, but later identified as *H. litoreum*) were included (Table 1). Sample from Črišnjeva was previously used for genomic DNA and chloroplast sequencing (Baruca Arbeiter et al., 2021; Hladnik et al., 2019) and hereby served as a reference.

Leaves were taken from plants in natural habitat and immediately stored in a saturated NaCl-CTAB conservation buffer (Rogstad, 1992). To preserve the biodiversity and viability of natural populations and minimize potential stress or damage of the sampled individuals, a non-destructive method was used in this study, in which a small tip or a few leaves were carefully removed from each selected plant. The analysed plant material also included dried flower heads of commercially available *H. arenarium* tea (purchased from Flora Ltd., Rogatec, Slovenia). Total genomic DNA was isolated from fresh leaves and dried flowers using a CTAB-PVP protocol (Japelaghi et al., 2011) with modifications, previously described by Baruca et al. (2021). To amplify the majority of alleles on DNA bulks of *H. arenarium*, we performed four separate DNA extractions from randomly sampled dried flower heads. DNA concentration was quantified with Qubit™ v3.0 fluorometer and Qubit™ dsDNA BR Assay Kit (Thermo Fisher Scientific). DNA was diluted to 10 ng/μl for PCR reactions. DNA extracts of plants from all collection sites have been deposited in the genetic laboratory of the University of Primorska (Slovenia) under the accession numbers HIUP\_1-195 (*H. italicum* plants), HLUP\_1-8 (*H. litoreum* plants), and HAUP\_1-4 (*H. arenarium*) and in accordance with the agreement of the French Ministry for an Ecological and Solidary Transition (Ministère de la Transition écologique et de la Cohésion des territoires), all isolated DNAs are available to interested researchers.



**Fig. 1.** Map of *H. italicum* collection sites in Croatia (Istrian peninsula in the northeastern Adriatic region) and France (Corsica).

**Table 1**

List of *H. italicum* populations with number of samples per population and samples of other *Helichrysum* species (*H. litoreum* and *H. arenarium*), included in the analysis.

Sampling location	Species	Number of samples
Croatia, Črišnjeva	<i>H. italicum</i>	1
Croatia, Cape Kamenjak	<i>H. italicum</i>	15
France, Corsica, Capo Pertusato	<i>H. italicum</i>	19
France, Corsica, Cavu	<i>H. italicum</i>	20
France, Corsica, Col de Bavella	<i>H. italicum</i>	20
France, Corsica, Punta di a Vacca Morta	<i>H. italicum</i>	20
France, Corsica, Col de Saint-Eustache	<i>H. italicum</i>	20
France, Corsica, Ajaccio	<i>H. italicum</i>	20
France, Corsica, Corte	<i>H. italicum</i>	20
France, Corsica, Lavu Santu	<i>H. italicum</i>	20
France, Corsica, Sisco	<i>H. italicum</i>	20
Plants from purchased certified seeds, grown on experimental collection of the University of Primorska (Slovenia) (45°34'19" N 13°46'32" E)	<i>H. litoreum</i>	8
Slovenia, commercially available <i>H. arenarium</i> tea (Flora Ltd., Rogatec, Slovenia)	<i>H. arenarium</i>	4 bulk samples

## 2.2 Chloroplast SSR markers identification and primer development

Reference *H. italicum* chloroplast genome with GenBank accession number NC\_041458.1, previously submitted by Hladnik et al. (2019), served as a template for cpSSR mining. SSRs were identified using MISA v2.1 (Beier et al., 2017; Thiel et al., 2003). Misa source code was downloaded and ran on a local computer. Configuration file misa.ini was modified in order to search microsatellites with mono- to hexanucleotide motifs. The minimum numbers of repeat units for mononucleotides, dinucleotides and higher order repeats were 10, 5, and 5, respectively. SSRs with flanking regions 300 nt upstream and downstream were extracted from the reference genome. MISA was run again on the extracted sequences and results were piped to PERL scripts, updated by MISA developers to be used with primer3 versions 2 and newer (available at: [https://github.com/soltislab/transcriptome\\_microsats.git](https://github.com/soltislab/transcriptome_microsats.git)) and to primer3 v.2.6.1 to design primers (bioconda package) (Koressaar & Remm, 2007; Untergasser et al., 2012). Parameter PRIMER\_PRODUCT\_SIZE\_RANGE was set to 100-300 in the p3in file. Sequences with primers were deposited in the NCBI Nucleotide database with accession numbers from PP747094 to PP747109.

## 2.3 Chloroplast SSR markers preliminary testing and genotyping

Easy scoring signals in the electropherograms, corresponding to expected size PCR products were observed in preliminary testing. Therefore, 16 newly developed primer pairs (for di-, tri- and 4 mononucleotide cpSSRs) were used for genotyping of all 207 (195 *H. italicum*, 8 *H. litoreum*, and bulk samples of *H. arenarium*) samples. PCR reactions were performed using the economic method where forward primer was elongated with M13(-21) 18 bp sequence at 5' end for fluorescent labelling (Schuelke, 2000). PCR reaction was performed in a final volume of 12.5 µl containing: 1x supplied AllTaq PCR buffer, 1x supplied Q-Solution, 2 mM MgCl<sub>2</sub>, dNTP mix (0.2 mM of each dNTP), 1.25 U of AllTaq DNA polymerase (Qiagen), 0.2 mM reverse locus specific primer and M13(-21) tail

labelled with fluorochrome (6-FAM, VIC, NED or PET), and 0.1 mM elongated forward primer. Locus specific primers were synthesized by IDT-DNA and universal M13(-21) primer by Applied Biosystems (Thermo Fisher Scientific).

The amplification was performed on a SimpliAmp™ Thermal Cycler (Thermo Fisher Scientific). The conditions of the PCR amplification were as follows: an initial denaturation at 94 °C for 5 min, followed by 30 cycles of denaturation for 30 s at 94 °C, 45 s at the annealing temperature of 56 °C and the extension at 72 °C for 45 s. For the last 8 cycles annealing temperature was set to 53 °C. Final extension was performed at 72 °C for 10 min (Schuelke, 2000).

One microliter of PCR product mix, including amplicons with four different fluorescent dyes, was added to 10.7 µl Hi-Di™ Formamide (Thermo Fisher Scientific) and GeneScan™ 500 LIZ size standard (Thermo Fisher Scientific) and analysed on a SeqStudio™ Genetic Analyzer (Thermo Fisher Scientific). Electropherograms were scored with GeneMapper version 5 software (Thermo Fisher Scientific).

#### **2.4 Chloroplast SSR markers sequencing**

To determine the origin of polymorphism in cpSSR loci and to confirm the nucleotide level identity of alleles of the same length, most alleles from di- and trinucleotide cpSSR loci were sequenced twice in each population of *H. italicum* and *H. litoreum* samples. An attempt was also made to sequence alleles from mononucleotide cpSSR loci, but the sequencing of some alleles failed despite several attempts.

The DNA of selected samples was amplified as described in “SSR amplification testing and populations genotyping” section with the only two differences: locus specific primers were added in equal molar concentration (0.2 mM) and M13(-21) tail was replaced with water to the final volume of 12.5 µl per reaction. PCR product was cleaned with Exonuclease I (Thermo Fisher Scientific) and FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific) in a final volume of 7 µl: 5 µL of PCR product, 2 U of Exonuclease I, 0.5 U of FastAP Thermosensitive Alkaline Phosphatase, and 1.4 µL of 1× PCR buffer (Qiagen). The PCR cleanup reaction was incubated for 45 min at 37 °C and ended at 80 °C for 15 min. Sequencing was performed in both directions, with forward and reverse locus specific primers, for each allele. Sequencing reaction in a total volume of 10 µl contained: 3.5 µl cleaned PCR product, 0.2 µl locus specific primer (forward or reverse), 0.5 µl BigDye™ Terminator v3.1 Ready Reaction Mix (Thermo Fisher Scientific), 2 µl 5X Sequencing Buffer and 3.8 µl water (Thermo Fisher Scientific). Cycle sequencing was performed on a SimpliAmp™ Thermal Cycler (Thermo Fisher Scientific) as follows: initial 3 min denaturation at 96 °C, 50 cycles with 10 s at 96 °C, 10 s at 50 °C and 4 min at 60 °C, and final incubation at 72°C for 7 min. Sequencing reaction products were cleaned with EDTA and ethanol precipitation methods, resuspended in 12 µl Hi-Di™ Formamide (Thermo Fisher Scientific) and analysed on SeqStudio™ Genetic Analyzer (Thermo Fisher Scientific). Ab1 files were extracted from the Thermo Fisher Scientific cloud and analysed with CodonCode Aligner v6.0 (CodonCode Corporation) to align the forward and reverse sequences.

#### **2.5 Chloroplast SSR markers diversity parameters, haplotypes identification and population's analysis**

The number of alleles ( $N_a$ ), the number of effective alleles ( $N_e$ ), Shannon's information index ( $I$ ), and diversity ( $h$ ) were calculated for each locus using the GenAlEx v.6.5.1 computer software (Peakall & Smouse, 2012).

The population's genetic diversity parameters, such as the number of haplotypes ( $A$ ), number of private haplotypes ( $P$ ), effective number of haplotypes ( $N_e$ ), haplotypic richness ( $R_h$ ), genetic diversity ( $H_e$ ), the mean genetic distance between individuals ( $D^2_{sh}$ ), relative contribution ( $c$ ), gene diversity within each population ( $H_s$ ), diversity due to differentiation ( $D_{st}$ ), total gene diversity ( $H_t$ ), and proportion of the total genetic differentiation ( $F_{st}$ ) were calculated for ten *H. italicum* populations with software Haplotype Analysis v.1.05 (Eliades & Eliades, 2009).

Dendrogram with *H. italicum* and *H. litoreum* haplotypes was constructed with Bruvo's distance (Bruvo et al., 2004) and Ward's ("ward.D") clustering method using function `bruvo.dist` of R `poppr` package (Kamvar et al., 2014) and `hclust` function of `stats` package.

The Bayesian clustering method implemented by the software STRUCTURE (Pritchard et al., 2000) was used for estimating the number of populations based on chloroplast SSR data. Parameters used for the analysis were: length of Burnin Period 50k, number of MCMC Reps after Burnin 50k, K was set from 1K to 10K and finally 5 iterations were used. Results of STRUCTURE were subjected to the `pophelper` R package (Francis, 2017) for calculating the most likely number of genetic clusters by the Evanno method (Evanno et al., 2005). The output matrix for the best K-values were plotted using the R package `pophelperShiny` (Francis, 2017).

We also constructed an evolutionary network with haplotypes using the median-joining (MJ) method (Bandelt et al., 1999) implemented in the software Network 10.2.0.0. Table with cpSSR alleles was converted to a binary matrix. As recommended by the program authors MP calculation was run on MJ network calculations.

## 2.6 Cross-species transferability

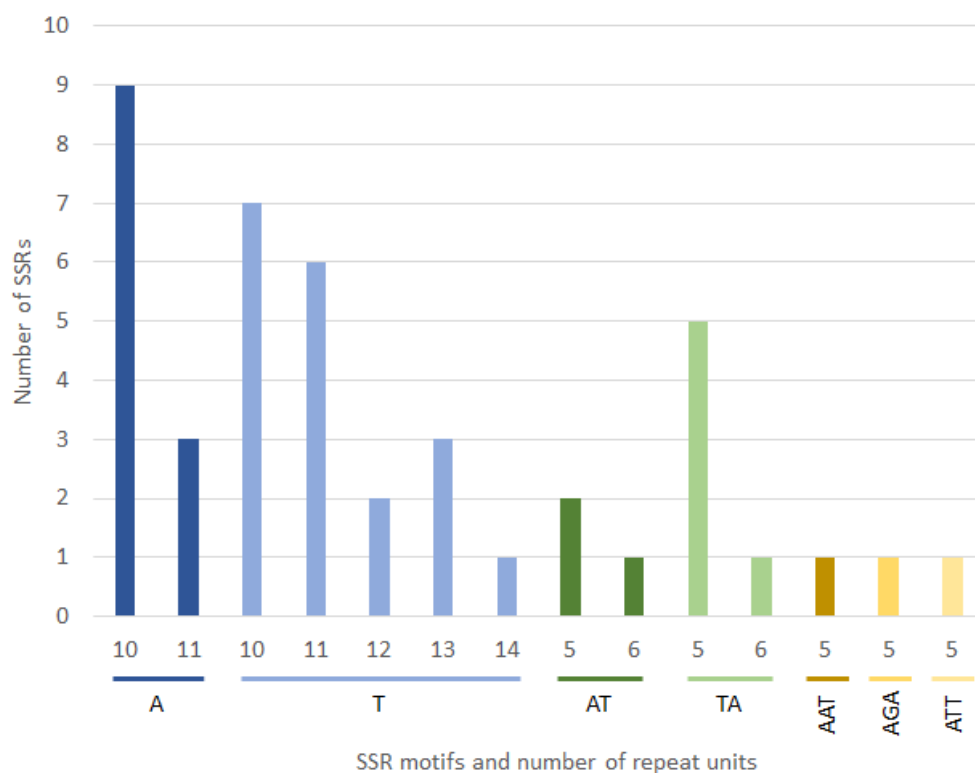
The ability of the developed microsatellite markers to cross-amplify was tested in two *Helichrysum* species: *H. litoreum* and *H. arenarium*. All 12 di- and trinucleotide and 4 mononucleotide cpSSRs were amplified in both species using the same PCR protocol as described above for *H. italicum* samples. Private alleles were identified with GeAlEx for *H. litoreum* and *H. italicum*. Additionally, private haplotypes were determined for *H. litoreum* with Haplotype Analysis v.1.05. For *H. arenarium* it was not relevant to identify haplotypes because bulk samples were used.

## 3. RESULTS

### 3.1 Chloroplast microsatellites characterization

In the present study, the complete chloroplast genome sequence of *H. italicum* (Hladnik et al., 2019) was examined with the MISA tool for microsatellites identification. In total 43 SSRs were identified. These SSRs comprised 31 mononucleotide repeats (72.1%), nine dinucleotides (21.0%), and three trinucleotide repeats (6.9%) (Fig. 2). Notably, there were no tetranucleotide, pentanucleotide and hexanucleotide repeats. Mononucleotides consisted exclusively of the nucleotides A or T. The most frequent mononucleotide motif was T, which accounted for 44.2% (of all identified SSRs), followed by the A (27.9%) motif. Among dinucleotides the TA motif (14.0%) was the most abundant, followed by AT (7.0%) motif. From the SSRs distribution perspective, 35 SSRs (81.4%) were observed in the LSC (long single copy) region, seven SSRs (16.3%) in the SSC (small single copy) region and only one SSR (2.3%) in the IRA (inverted repeat A) region. In addition, the results showed that most SSR repeats were distributed in the intergenic regions (28 SSRs; 65.1%) as they always contain more polymorphisms, followed by the protein-coding and non-coding RNA genes (15 SSRs; 34.9%), including *trnK-UUU*, *rpoB*, *rpoC1*, *rpoC2*, *ycf1*, and *ycf3* genes, while no SSR repeat was observed in rRNA genes.





**Fig. 2.** Overall abundance of various SSR repeat types and number of repeat units in *H. italicum* chloroplast genome.

### 3.2 Primer designing and marker validation

Primers were designed for all SSRs. Test amplification was performed with four samples from different *H. italicum* populations (Kamenjak, Capo Pertusato, Cavu and Col de Bavella) to increase the probability for detecting different alleles. DNA from an individual *H. italicum* plant (Črišnjeva, Croatia) previously collected for high-throughput sequencing (Baruca Arbeiter et al., 2021) and chloroplast genome assembly (Hladnik et al., 2019), was used as a reference sample for PCR amplification of all tested loci. Finally, 16 SSRs (all 12 di- and trinucleotide repeats, and the four mononucleotide repeats with/presenting the highest polymorphism) were selected for further analysis of the ten *H. italicum* populations (194 samples) obtained from Cape Kamenjak (Istrian peninsula, Croatia) and the island of Corsica (France). The final selected set of SSRs with primers and amplicon lengths, observed in sequenced *H. italicum* plant, is listed in Table 2 whereas data related to the rest of the mononucleotide SSRs is detailed in Table A.1.

**Table 2**

Characteristics of developed 16 cpSSR markers as observed in all *H. italicum* samples with corresponding GenBank accession numbers.

Locus	Repeat motif	Primer sequences (5' -3')	Location/ region	SSR Position	Allele size range (bp)	GenBank accession number
<i>cpHiUP-01</i>	(ATT) <sub>5</sub>	F: *TTCATTGCACACGGCTTTCC R: TGATTGGCCAATCATGAATGAT	<i>tRNK-UUU</i> intron	2063 - 2077	276 - 289	PP747094
<i>cpHiUP-02</i>	(TA) <sub>5</sub>	F: *AGGTAGTCCTTTGTGGCTGC R: TGAGACTCGAATTCGCCGCTG	<i>RpoC1</i>	18669 - 18678	271	PP747095
<i>cpHiUP-03</i>	(AT) <sub>5</sub>	F: *GCCCCGTAAGGAGTTGTGGA R: TTGACAAGCCCAACCCCAAT	<i>RpoC2</i>	19665 - 19674	266	PP747096
<i>cpHiUP-04</i>	(AT) <sub>5</sub>	F: *ACCCGCCCAAAATGAAGTCA R: AGGCAAGGAGGGGAAGGATA	intergenic	36185 - 36194	253 - 293	PP747101
<i>cpHiUP-05</i>	(TA) <sub>6</sub>	F: *TACTTCTGGTTCCGGCGAAC R: TAGATCCGAACACTTGCCCC	intergenic	41628 - 41639	227 - 229	PP747102
<i>cpHiUP-06</i>	(AT) <sub>6</sub>	F: *ATTGCAAATCACGCGATCGG R: TCCCTTCTAAGCGTTCTATTTC	intergenic	42072 - 42083	240 - 253	PP747103
<i>cpHiUP-07</i>	(TA) <sub>5</sub>	F: *ACGCGACATAAAGACTCCTTCT R: TCCTCTTTTCGAAAAGCCA	intergenic	46438 - 46447	291 - 297	PP747104
<i>cpHiUP-08</i>	(TA) <sub>5</sub>	F: *AGATCGAAGCGAGTACCTGC R: GGCAAACGCCTACGAAAAGA	intergenic	67942 - 67951	238 - 244	PP747105
<i>cpHiUP-09</i>	(TA) <sub>5</sub>	F: *CTGAGACGACCCAGAAAAGCA R: TTCATCATCCGGCTCGAGC	<i>petD</i> intron	77234 - 77243	260 - 264	PP747106
<i>cpHiUP-10</i>	(AGA) <sub>5</sub>	F: *AGAAACTGAACAACCTCGACA R: TCGTAATGGCCTATTCCATCGG	<i>Ycf1</i>	109291 - 109305	207 - 213	PP747107
<i>cpHiUP-11</i>	(AAT) <sub>5</sub>	F: *TGGACAAATTATTCACAGGCCG R: TTGTAGAGTGATTCTGCGCA	<i>Ycf1</i>	111479 - 111493	274	PP747108
<i>cpHiUP-12</i>	(TA) <sub>5</sub>	F: *GAGCAGCGTGTCTACCGATT R: TCGCGAGAATTCTCTAGTTGCA	intergenic	123336 - 123345	195 - 218	PP747109
<i>cpHiUP-13</i>	(T) <sub>13</sub>	F: *AAGGGCTGATTGCGGATGA R: TCCTTGCTTGCAACCCCTTCT	intergenic	26467 - 26479	108 - 123	PP747097
<i>cpHiUP-14</i>	(T) <sub>11</sub>	F: *TTGGGGGCGATGAAACAAC R: ACGAGCAATGCCATCACCTA	intergenic	27974 - 27984	210 - 213	PP747098
<i>cpHiUP-15</i>	(A) <sub>10</sub>	F: *TGGACGCCTTTCATTGTGAT R: ACCCTAGCCTTCCAAGCTA	intergenic	29779 - 29788	273 - 292	PP747099
<i>cpHiUP-16</i>	(T) <sub>10</sub>	F: *AGTCACTTTGGTTCCCG R: AAAGAAAGGGAAGGGGCTCG CAT	intergenic	34914 - 34923	133 - 140	PP747100

\*Primers elongated for the universal M13 (-21) sequence (5'-TGTAACGACGGCCAGT-3') at the 5' ends.

All 16 microsatellite markers amplified successfully across ten *H. italicum* populations and allele size ranges at all loci were comparable with expected allele sizes. However, *cpHiUP-02*, *cpHiUP-03* and *cpHiUP-11* were monomorphic. Although sequencing of the *cpHiUP-03* alleles revealed the presence of a SNP in a few samples (Fig. A.1). All three mentioned loci were excluded from the population analysis. Additionally, two alleles with the same length but with different nucleotide sequences were identified in the locus *cpHiUP-04*, making this locus unsuitable for fragment analysis. The SNPs were observed also in some *cpHiUP-09* alleles with the same length, and low discrimination ability of *H. italicum* samples of locus *cpHiUP-10* (only one sample had allele 207 whereas in all other samples allele 213 was observed) were also the reasons for excluding these two cpSSR loci. Finally, out of the 12 di- and trinucleotide SSR loci, six loci (*cpHiUP-01*, *cpHiUP-05*, *cpHiUP-06*, *cpHiUP-07*, *cpHiUP-08* and *cpHiUP-12*) were selected for further analysis. Contrary to di- and trinucleotide SSRs, all four mononucleotide SSRs were kept in the final dataset for population analysis.

Source of length polymorphism was due to a different number of SSR motif replications in all finally selected di- and trinucleotide SSRs, but some alleles showed mutations also in flanking regions. Allele *cpHiUP-05* 252 showed duplication of the 18 bp region present within the TA motif (Fig. A.1). Locus *cpHiUP-07* showed highest variability. Alleles 291 and 293 of *cpHiUP-07* locus were different because of different number of SSR repeat motif, whereas 285, 287 and 297 alleles were different because of insertion of two thymines in the mononucleotide region (287), duplication of ATTAAT motif (297) and duplication of TCTAT motif (297). All three alleles had deletion of ATAAATAT sequence compared to 291 and 293 alleles. Similarly, *cpHiUP-12* 218 allele had duplicated 15 bp motif (TTATATTC AATCGTC), whereas allele 187 had deletion of AAATAGAA motif in a flanking region (two such motifs were present in other alleles). Due to difficulties in sequencing loci with mononucleotides, the source of polymorphism could not be detected for all alleles despite several

attempts. However, based on the successfully sequenced alleles and the observation that most alleles differ in length by 1 bp (with some exceptions), all four mononucleotide cpSSRs were included in the genotyping analysis.

In total, 38 alleles were detected in 10 chloroplast SSR loci among the 194 samples. The number of amplified alleles ( $N_a$ ) at each locus ranged from two (cpHiUP-05, cpHiUP-08) to seven (cpHiUP-15) with an average of 3.8 alleles. The effective number of alleles ( $N_e$ ) ranged from 1.03 (cpHiUP-08) to 2.05 (cpHiUP-14) with an average of 1.49. Shannon's information index (I) varied between 0.080 (cpHiUP-08) and 0.974 (cpHiUP-15), with an average value of 0.516, whereas observed diversity by locus (h) was between 0.030 (cpHiUP-08) to 0.513 (cpHiUP-14), with an average value of 0.271 (Table 3).

**Table 3**

Genetic diversity parameters for the 10 polymorphic chloroplast SSR markers based on genotyping results of ten *H. italicum* populations.

Locus	$N_a$	$N_e$	I	h
cpHiUP-01	3	1.098	0.204	0.089
cpHiUP-05	2	1.097	0.188	0.088
cpHiUP-06	3	1.087	0.195	0.080
cpHiUP-07	3	1.975	0.772	0.494
cpHiUP-08	2	1.031	0.080	0.030
cpHiUP-12	3	1.075	0.170	0.070
cpHiUP-13	6	1.762	0.863	0.432
cpHiUP-14	4	2.053	0.913	0.513
cpHiUP-15	7	1.883	0.974	0.469
cpHiUP-16	5	1.813	0.804	0.448

$N_a$  - number of alleles,  $N_e$  - number of effective alleles, I - Shannon's information index, h - diversity

### 3.3 Genetic structure of *H. italicum* populations

The genetic diversity parameters of the studied *H. italicum* populations are summarised in Table 4. The highest number of alleles (from all ten cpSSR loci) was obtained in the Kamenjak population (25) and the lowest in Sisco (14), representing 66 % and 37 % of all 38 alleles, respectively. Of the ten cpSSR markers analysed, seven produced 10 private alleles for seven populations (Table 5). The highest number of private alleles (three) was characteristic for cpSSR marker cpHiUP-15 and only one private allele was detected at seven loci (cpHiUP-01, cpHiUP-06, cpHiUP-07, cpHiUP-08, cpHiUP-12, cpHiUP-13, and cpHiUP-16). Two of the private alleles were observed for Kamenjak, Pertusato and Corte, and one private allele for Col de Bavella, Punta di a Vaca Morta, Col de Saint-Eustache and Ajaccio. No private alleles were detected for the populations Cavu, Lavu Santu and Sisco. The Shannon's information index across populations ranged from 0.169 (Sisco) to 0.665 (Punta di a Vaca Morta) with an average value of 0.309.

**Table 4**

Intrapopulation analysis of *H. italicum* genetic resources based on 10 chloroplast SSR markers.

Population	N	No. of alleles/pop.	No. of private alleles/pop.	I	A	P	N <sub>e</sub>	R <sub>h</sub>	H <sub>e</sub>	D <sup>2</sup> <sub>sh</sub>
Kamenjak	15	25	2	0.571	9	7	7.258	8.000	0.924	13.914
Pertusato	19	15	2	0.213	6	5	3.374	4.333	0.743	1.553
Cavu	20	15	/	0.239	5	3	4.444	3.938	0.816	0.524
Col de Bavella	20	19	1	0.268	7	3	4.000	4.998	0.789	7.319
Punta di a Vaca Morta	20	23	1	0.665	7	5	3.704	5.144	0.768	86.404
Col de Saint-Eustache	20	17	1	0.229	6	3	3.390	4.241	0.742	1.947
Ajaccio	20	15	1	0.280	4	1	2.899	2.939	0.689	30.806
Corte	20	22	2	0.457	7	4	4.878	5.421	0.837	9.962
Lavu Santu	20	10	/	0.000	1	0	1.000	0.000	0.000	0.000
Sisco	20	14	/	0.169	2	1	1.342	0.991	0.268	0.671
<b>Mean</b>	/	/	/	<b>0.309</b>	<b>5.400</b>	<b>3.200</b>	<b>3.629</b>	<b>4.000</b>	<b>0.658</b>	<b>15.310</b>

N - sample size in each *H. italicum* population, No. of alleles/pop. - number of alleles per population, No. of private alleles/pop. - number of private alleles per population, I - Shannon's information index, A - number of different haplotypes, P - number of private haplotypes, N<sub>e</sub> - effective number of haplotypes, R<sub>h</sub> - haplotypic richness, H<sub>e</sub> - genetic diversity, D<sup>2</sup><sub>sh</sub> - mean genetic distance between individuals

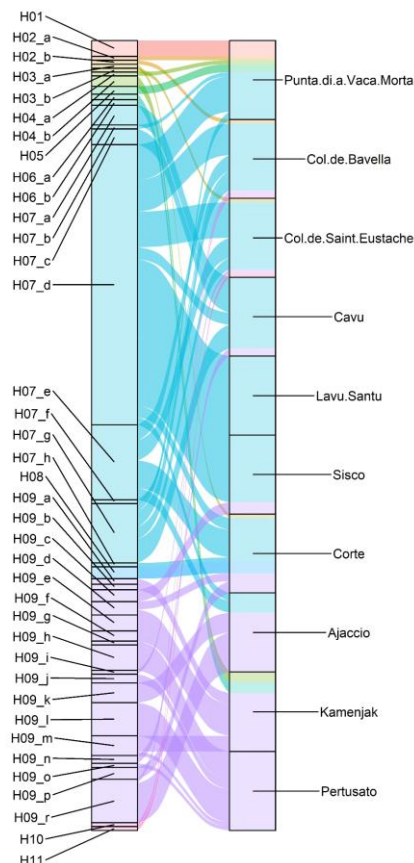
**Table 5**

List of private alleles (locus name, allele length in bp, and frequencies) by *H. italicum* populations.

Population	Locus	Allele	Frequency
Kamenjak	P-06	240	0.133
Kamenjak	PM-10	140	0.067
Pertusato	PM-09	279	0.053
Pertusato	PM-09	292	0.053
Col de Bavella	P-12	218	0.050
Punta di a Vaca Morta	P-07	297	0.300
Col de Saint-Eustache	P-01	289	0.050
Ajaccio	PM-09	273	0.250
Corte	P-08	238	0.150
Corte	PM-07	112	0.050

Based on the observed genotyping data for all *H. italicum* samples (and eight *H. litoreum* samples that served as outgroup species), the haplotypes were determined and their distribution across populations. Eleven haplotypes were identified when only 6 di- and trinucleotide SSRs were taken into account (and additional 6 haplotypes were determined for *H. litoreum*), whereas 38 haplotypes were identified in *H. italicum* samples with four additional mononucleotide SSRs. To visualize the

differences between haplotypes based on 6 di- and trinucleotide SSRs and 4 additional mononucleotide SSRs, the main groups of haplotypes were labeled as H01 to H17, while further divisions based on mononucleotide SSRs were labeled as sub-haplotypes (e. g. H02\_a, H02\_b, etc.) (Figures 3 and A.2.). Haplotypes H07 and H09 had the largest number of sub-haplotypes (8 and 17, respectively), while haplotypes H01, H05, H08, H10 and H11 were not further divided in sub-haplotypes. The *H. litoreum* samples showed species-specific haplotypes (H12, H13, H14, H15, H16, and H17). Two haplotypes, H07 (in blue) and H09 (in violet), were the most highly represented across the sampled locations: Ajaccio, Pertusato and Kamenjak predominantly featured the H09 sub-haplotypes, while H07 sub-haplotypes dominated in all other populations (Figs. 3, A.2.).



**Fig. 3:** Distribution of haplotypes and their sub-haplotypes across populations, all from Corsica except Kamenjak (Croatia). Haplotypes, as determined with di- and trinucleotide SSRs, are marked with different colors, whereas sub-haplotypes are indicated by a horizontal line within each haplotype.

The number of haplotypes per population ( $A$ ) ranged from one to nine, and the effective number of haplotypes ( $N_e$ ) was between 1.00 and 7.26. When all populations were analysed together, the mean  $A$  and mean  $N_e$  were 5.40 and 3.63, respectively. With the exception of the Lavu Santu population, in which only one haplotype was found, the values for haplotypic richness ( $R_h$ ) ranged from 0.99 (Sisco) to 8.00 (Kamenjak), with a mean value of 4.00. Similarly, genetic diversity values ( $H_e$ ) were between 0.268 (Sisco) and 0.924 (Kamenjak), and most populations had a genetic diversity value above 0.700. The mean genetic distance ( $D^2_{sh}$ ) between individuals varied more widely, ranging from 0.524 (Cavu) up to 86.404 (Punta di a Vaca Morta), with a mean value of 15.310 (Table 4).

Out of 38 haplotypes, 32 private haplotypes ( $P$ ) were found in nine out of ten *H. italicum* populations (Table A.2). The highest number of private haplotypes was found in Kamenjak (seven), followed by the populations of Pertusato (five), Punta di a Vaca Morta (five) and Corte (four), while three private haplotypes were found in Cavu, Col de Bavella, Col de Saint-Eustache, and only one in Ajaccio and Sisco (Table 4). The frequency of the private haplotypes was mostly between 0.05 and 0.25, with the exception of the two haplotypes (H09\_h and H09\_l) in the Pertusato population, which had a frequency of 0.32 and 0.42, respectively.

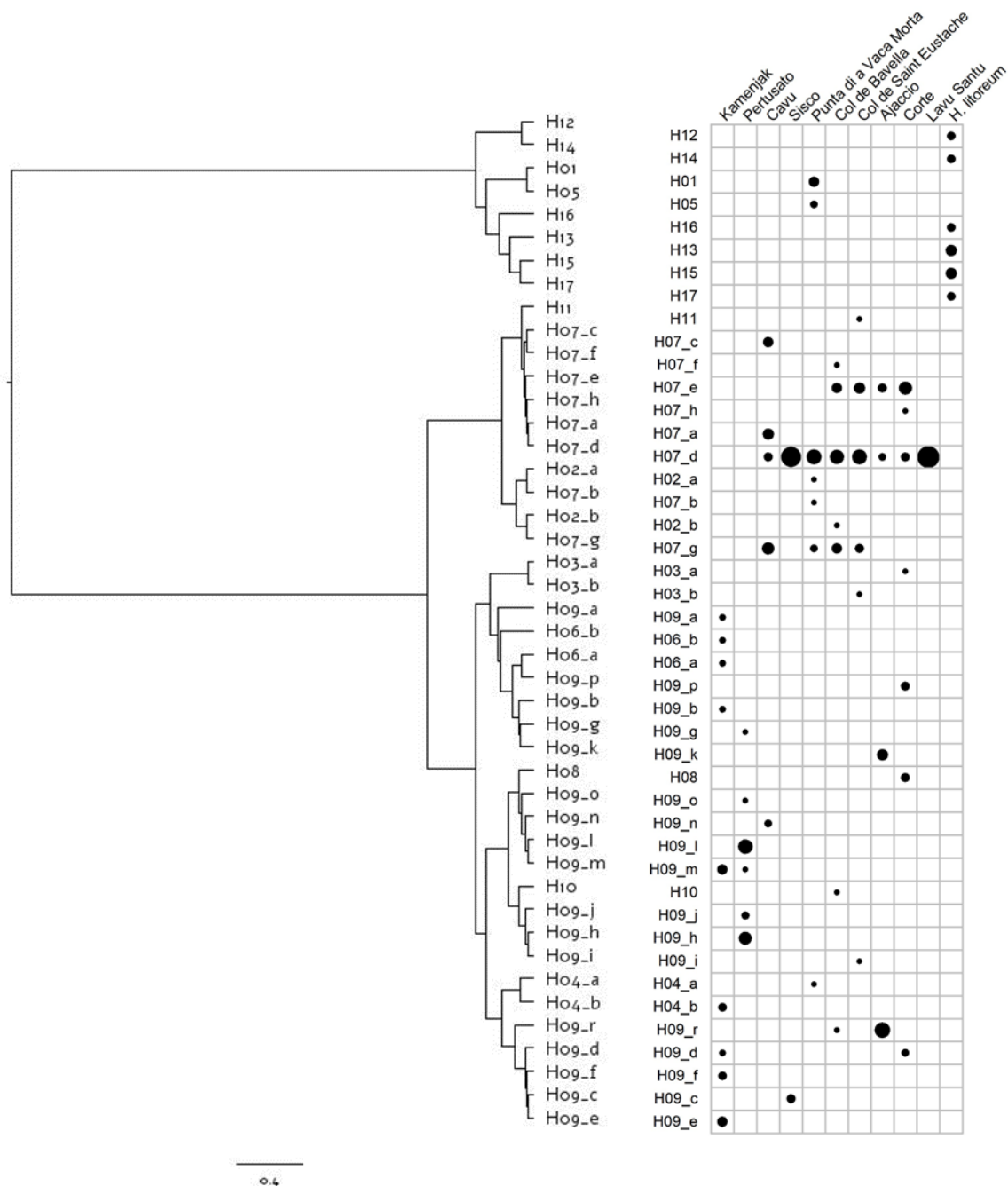
Gene diversity within populations ( $H_s$ ) ranged from 0.026 in the Sisco population to 0.082 in the Corte population (Table 6). The values of  $D_{st}$  (diversity due to differentiation) were lower than  $H_s$ , ranging from 0.014 to 0.034, indicating that most of the genetic diversity was found within populations (Table 6). In order to assess the level of genetic differentiation among the *H. italicum* populations, the  $F_{st}$  values (proportion of the total genetic differentiation) between populations were calculated. The highest difference was found between the populations of Lavu Santu and Pertusato ( $F_{st}=0.185$ ; data not shown) and the lowest difference between Sisco and Lavu Santu ( $F_{st}=0.010$ ; data not shown). As expected, the proportion of total genetic differentiation within populations was higher than between populations. In addition, the relatively low overall value of  $F_{st}$  (0.266) indicates that a low to marginal proportion of diversity (26.6%) is observed between populations (Table 6).

**Table 6**

Interpopulation analysis of *H. italicum* genetic resources based on 10 chloroplast SSR markers.

	N	c	$H_s$	$D_{st}$	$H_t$	$F_{st}$
Kamenjak	15	0.077	0.067	0.019	0.086	0.224
Pertusato	19	0.098	0.069	0.030	0.099	0.306
Cavu	20	0.103	0.080	0.022	0.102	0.217
Col de Bavella	20	0.103	0.077	0.014	0.091	0.149
Punta di a Vaca Morta	20	0.103	0.075	0.015	0.091	0.169
Col de Saint-Eustache	20	0.103	0.073	0.014	0.087	0.163
Ajaccio	20	0.103	0.068	0.029	0.096	0.298
Corte	20	0.103	0.082	0.020	0.102	0.200
Lavu Santu	20	0.103	0.000	0.034	0.034	1.000
Sisco	20	0.103	0.026	0.026	0.052	0.496
			<b><math>H_s</math></b>	<b><math>D_{st}</math></b>	<b><math>H_t</math></b>	<b><math>F_{st}</math></b>
<b>Total</b>	<b>194</b>	<b>1.000</b>	<b>0.617</b>	<b>0.223</b>	<b>0.840</b>	<b>0.266</b>

N - sample size in each population, c - relative contribution,  $H_s$  - gene diversity within each population,  $D_{st}$  - diversity due to differentiation,  $H_t$  - total gene diversity,  $F_{st}$  - proportion of the total genetic differentiation

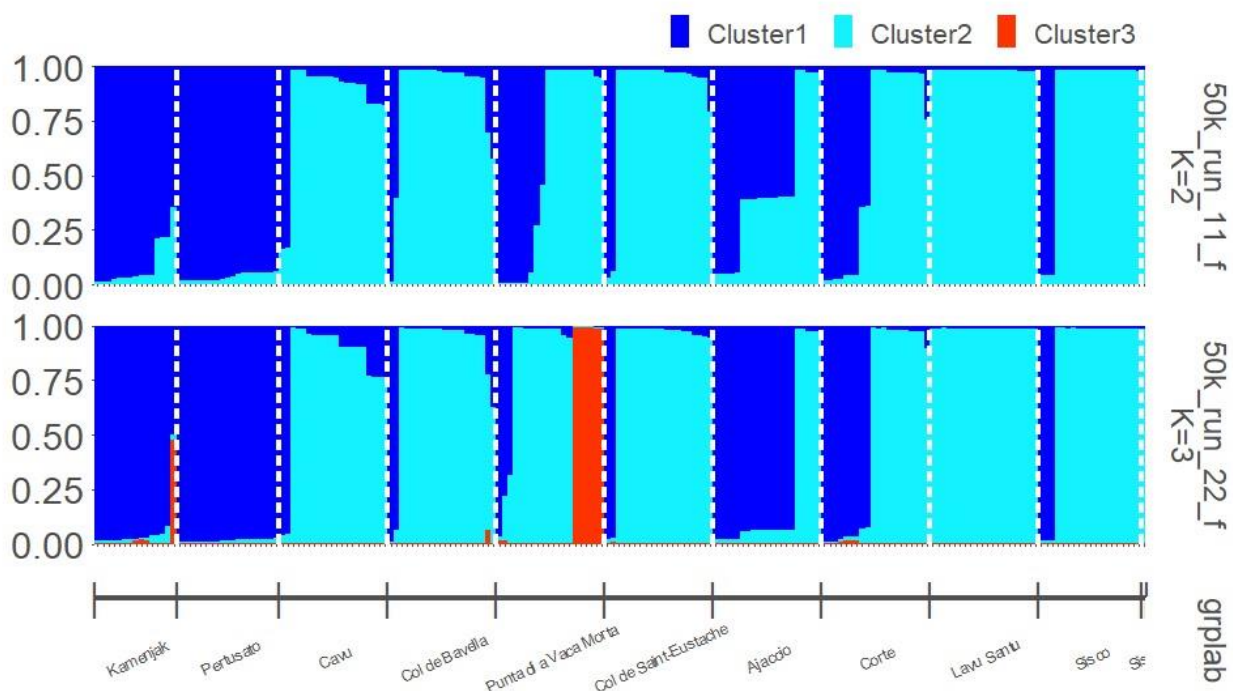


**Fig. 4.** Left: Dendrogram with *H. italicum* and *H. litoreum* (sub-)haplotypes based on Bruvo's distance and Ward's linkage method for the hierarchical clustering. Right: frequency distribution of (sub-)haplotypes across the studied populations/species.

The samples from Kamenjak (Croatia) showed the following sub-haplotypes: H04\_b, H06\_a and H06\_b and six H09 sub-haplotypes (Figures 3, 4 and A.2.). The Pertusato population had only H09 sub-haplotypes and one of its sample/individual/plant shared the same haplotype with samples from Kamenjak. The population in which H09 sub-haplotypes were predominant was also from Ajaccio (75%). In other Corsican populations, H07 sub-haplotypes were dominating. Three haplotypes had a population distribution similar to H07 and were grouped in the same cluster, namely H02\_a, H02\_b and H11 (Fig. 4). On the other hand, eight haplotypes (H03\_a, H03\_b, H04\_a, H04\_b, H06\_a, H06\_b, H08 and H10) were clustered with the H09 sub-haplotypes. *H. litoreum* formed a separate group together with six samples from the Corsican population of Punta di a Vacca Morta. It is interesting to

note that the only haplotype (H07\_d) detected in the population of Lavu Santu was also the most widespread haplotype, present in eight populations. Haplotype H07\_g was only observed in populations from south of Corsica (Punta di a Vacca Morta, Cavu, Col de Bavella and Col de Saint Eustache). Similarly, haplotype H07\_e was mostly observed in southern Corsican populations (Col de Bavella, Col de Saint Eustache, Ajaccio) with the exception of the population from Corte (centre-north Corsica). Haplotypes, other than H07 and H09 (H02, H03, H04, H05, H06, H08, H09, H10 and H11) were identified only in a single or few samples. Although the majority of Corsican populations were most frequently constituted of H07 sub-haplotypes, H09 sub-haplotypes were also identified, although in minority, in almost all populations.

Geographic population structure was evaluated with the Bayesian clustering method and the most likely number of genetic clusters was determined by the Evanno method (Evanno et al., 2002). When all *H. italicum* populations were analyzed, the modal value of the  $\Delta K$  distribution indicated that the true K value was two (Fig. 5;  $K = 2$ ). Out of the 10 populations studied, six (Kamenjak, Pertusato, Cavu, Col de Bavella, Col de Saint Eustache, and Lavu Santu) appeared to be very homogeneous, as almost all individuals were assigned to one cluster, while the other four (Punta di a Vacca Morta, Ajaccio, Corte and, to a lesser extent, Sisco) appeared to be more heterogeneous, as individuals were assigned to two different clusters. Although the Evanno method indicated  $K = 2$  as the most significant K value, at  $K = 3$  (Fig. 5) we found that six samples from Punta di a Vacca Morta were defined as a third cluster, and these samples were represented by haplotypes H01 and H05. This result is comparable to evolutionary haplotype networks showing the genetic similarity between the haplotypes of Punta di a Vacca Morta and the species *H. litoreum* (Fig. A.3). Haplotype network also demonstrated distribution of haplotypes, similar to distribution in dendrogram (Fig. 4).

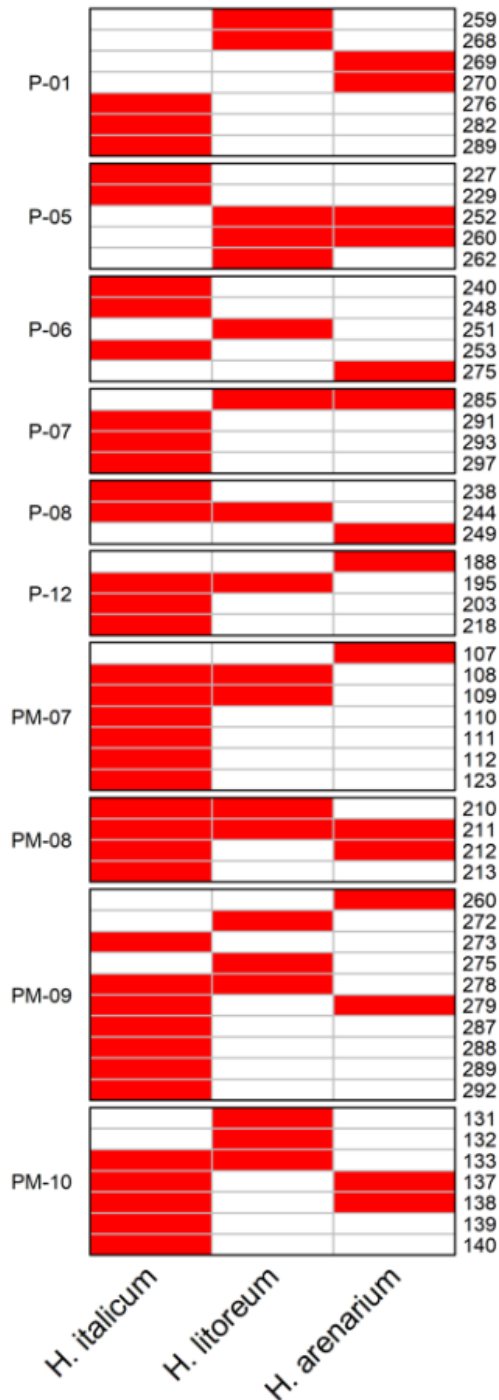


**Fig. 5:** Results of the Bayesian model-based clustering STRUCTURE analysis of 194 individuals of *H. italicum*. Results were plotted when  $K = 2$  and  $K = 3$  based on 10 cpSSR loci.



### 3.4 Cross-species transferability

The new set of cpSSR markers developed from *H. italicum* was used to verify cross-species transferability with *H. litoreum* and *H. arenarium*. All 10 finally selected SSR markers showed successful amplification. Out of the 56 amplified alleles, seven alleles were identified as private for *H. litoreum*, six alleles for *H. arenarium*, and 26 for *H. italicum* (Fig. 6). Only one allele (211) is common to the three studied species of *Helichrysum*, seven alleles are in common between *H. italicum* and *H. litoreum*, four between *H. italicum* and *H. arenarium*, and three between *H. litoreum* and *H. arenarium*.



**Fig. 6:** Comparison of the 56 amplified alleles across three *Helichrysum* species (*H. italicum*, *H. litoreum* and *H. arenarium*). Red color in the cells indicates the successful amplified alleles, indicated on the right side.

#### 4. DISCUSSION

The plant of *H. italicum* is one of the most remarkable and iconic medicinal plants of the Mediterranean area, with high morphological, genetic and chemical polymorphism in its natural populations. Nowadays, the essential oil of *H. italicum* is highly appreciated in the glamorous perfume and cosmetics industry for its anti-ageing properties and offers great potential in the field of health care as pharmaceuticals, functional food ingredients and dietary supplements due to the unique properties of its bioactive compounds in various extracts. As a result, commercial cultivation of *H. italicum* is increasing in many countries (France-Corsica, Portugal, Italy, Slovenia, Croatia, Serbia, Bosnia-Herzegovina, Albania, Bulgaria) (Erbaş et al., 2023), but no varieties with superior traits have been registered yet. The plant is only in the early stages of domestication and cultivation in the agroecosystem is based on unverified planting material. Planting material is usually produced by vegetative propagation of randomly selected mother plants or by seeds collected in nature, which indicates a questionable quality of the raw plant material for medicinal, pharmaceutical and cosmetic purposes. The management of planting material is therefore a huge problem, and the lack of control and certification has led to economic losses in some regions (Baruca Arbeiter et al., 2021), mainly due to production of essential oil with inappropriate chemical composition or low quantities of biologically important molecules requested from industry. Therefore, there is an urgent need for the development of molecular tools that allow the identification of local genotypes (to protect the typicality and geographical origin of local products) and the development of new varieties with higher quality properties for the needs of the target industry, as well as strict control of the authenticity and traceability of planting material on the market.

The chloroplast genomes of most angiosperms have a highly conserved sequence and structure, but they can also differ within species due to accumulated mutations or rare genomic rearrangements, making the chloroplast suitable for various genetic analyses. SSRs can become mutational hotspots and are therefore one of the most commonly used molecular markers. The chloroplast SSRs contain rich information on genetic variation and have been successfully used in various crop species (Cheng et al., 2005; Islam et al., 2021) and in medicinal/aromatic plants (Guo et al., 2022; Liu et al., 2023; Shi et al., 2022) to study their genetic diversity, origin and population genetic structures.

In total, 43 chloroplast microsatellite regions containing mono-, di-, and trinucleotide repeats were detected in this study. Comparable number and distribution of SSR repeats have been observed in chloroplast genomes of 12 Asteraceae species (Liu et al., 2023). Mononucleotide repeats (A/T) were the most highly represented, as it has been observed in the chloroplast genome of other higher plants (Liu et al., 2023; Lukas & Novak, 2013; Zhang et al., 2023). Contrary to our study, tetra-, penta-, and hexanucleotide repeats have also been reported in other medicinal plants from the Asteraceae family, such as *Carpesium* species (Shi et al., 2022), *Achillea millefolium* (Liu et al., 2023) and *Sinosenecio* species (Peng et al., 2022). Also, in accordance with previous reports, most chloroplast SSRs were present in the intergenic regions (Vieira et al., 2016).

Primers were successfully designed for all 43 SSRs, but a subset of 16 SSRs (including all di- and trinucleotide loci and 4 mononucleotides), selected on the basis of preliminary testing, was used to test their ability to discriminate plants within and between selected *H. italicum* populations. Based on the genotyping results, four cpSSRs (cpHiUP-02, cpHiUP-03, cpHiUP-10 and cpHiUP-11) were removed as they were either monomorphic or showed a very low level of polymorphism, while at locus cpHiUP-09 two alleles of the same length but with SNP in the flanking region were observed (the same was observed for cpHiUP-03). The locus cpHiUP-04 was excluded due to size homoplasy (alleles with different indels but with the same final length). Indels and base substitutions in the flanking regions, which can also lead to size homoplasy, were frequently reported (Ebert & Peakall, 2009; Hale et al., 2004; Peakall et al., 1998). Insertions, consisting of duplications, of several bp regions, as observed in our study, have also been previously reported in the chloroplast genome of *Citrus* species (Carbonell-Caballero et al., 2015). This sequence complexity could be a potential limitation in the use of cpSSR markers for estimating phylogenetic relationships in plants (Provan et al., 2001; Hale et al., 2003). The results of the present study confirm that sequencing of alleles of the

same size in different populations or species should be performed to correctly select only the most appropriate cpSSR loci and thus reduce the presence of size homoplasy in the dataset.

Finally, 10 cpSSR markers, including six di- and trinucleotides (cpHiUP-01, cpHiUP-05, cpHiUP-06, cpHiUP-07, cpHiUP-08, cpHiUP-12) and four mononucleotide cpSSR loci (cpHiUP-13, cpHiUP-14, cpHiUP-15, cpHiUP-16), were selected as informative and suitable for the study of plant material. A higher number of alleles per locus was observed for mononucleotide cpSSR loci than for di- and trinucleotide cpSSR loci, which is consistent with the study by Tian et al. (2023) who found that mononucleotide repeats contribute more to genetic variation than other SSRs. The average values for  $N_a$  (3.8) and  $N_e$  (1.49) were slightly lower compared to the values obtained from a genotyping of 145 *Ribes* genotypes, but these were obtained from material from different species and 19 cpSSR makers were used (Lācis et al., 2022). Much higher values for both parameters ( $N_a = 8.5$ ,  $N_e = 4.46$ ) were obtained in the study, in which 28 *H. italicum* samples from the northeastern Adriatic region were tested with genomic SSRs (Baruca Arbeiter et al., 2021). These observations support the fact that cpSSRs are less polymorphic than genomic SSRs due to the lower mutation rate, but the substitution rates observed in SSR loci are still higher than elsewhere in the chloroplast genome (Provan et al., 2001).

The genetic diversity and differentiation of *H. italicum* populations from different natural habitats in Corsica (France) and Cape Kamenjak (Istrian peninsula in the northeastern Adriatic, Croatia) were investigated with 10 newly developed cpSSR markers. The genetic diversity within the 10 populations was estimated using the Shannon information index, which provides information on the genetic variability and richness of the populations studied. In particular, the populations of Kamenjak, Punta di a Vaca Morta and Corte had the highest Shannon index values ( $I = 0.475 - 0.571$ ). All other populations had lower Shannon index values ( $I = 0.169 - 0.268$ ). The overall genetic variability found in this study ( $I = 0.309$ ) is comparable to the 18 *H. italicum* populations from the eastern Adriatic coast ( $I = 0.355$ ) analysed with AFLP markers by Ninčević et al. (2021) and by Marini et al. (2023) in Tuscan Archipelago *Helichrysum* taxa ( $I = 0.213$ ).

The populations of *H. italicum*, with the exception of the populations of Lavu Santu, Sisco and Ajaccio (with the lowest number of haplotypes), are characterized by a high genetic diversity, which can be also observed with other statistical parameters ( $A$ ,  $P$ ,  $R_h$ ,  $H_e$ ). The highest values were found in the Kamenjak population, which refers to the southernmost cape of the Istrian peninsula and represents the only population of the northeastern Adriatic region in this study. The presence of rocky pastures on the Istrian peninsula, where heliophilous plants prevail, significantly enrich the plant and landscape diversity (Ljubičić et al., 2020), could be an explanation for the high genetic diversity. Among the Corsican *H. italicum* populations, the lowest values of genetic diversity parameters were detected in Sisco and Lavu Santu. One possible explanation for these results is that Lavu Santu is a completely isolated population originating from the beach of the same name at the mouth of the Cavu River, while Sisco represents the northernmost population of Cap Corse, which is also isolated. The other six Corsican populations (Pertrusato, Cavu, Col de Bavella, Punta di a Vaca Morta, Col de Saint-Eustache, Corte) were characterized by a high genetic diversity comparable to the Kamenjak population. A comparison of populations from the north (Haute-Corse) and the south of Corsica (Corse-du-Sud) revealed some differences in haplotype distribution. Haplotypes H9\_i, H9\_j, H9\_h and H10 were only identified in populations from southern Corsica (Pertrusato, Col de Bavella, Col de Saint-Eustache), while haplotypes H01 and H05 were only found in the population of Punta di a Vaca Morta. Plants in this location were morphologically very different and we suspect the natural mixture/hybridization between two subspecies *H. italicum* spp. *italicum* and *H. italicum* spp. *thyrrenicum*. In order to gain a better insight into the genetic structure of *H. italicum* in Corsica, further populations from northern Corsica should be included in future studies and also a valid reference of *H. italicum* spp. *thyrrenicum* to control for any genetic introgression.

Two haplotypes (H07 and H09, as defined with di- and trinucleotide SSRs) were most common in Corsica and Cape Kamenjak. The Corsican populations of Ajaccio and Pertusato and the Cape Kamenjak had predominantly the H09 sub-haplotypes, while the H07 sub-haplotypes were the most prevalent at all other sites in Corsica. The result of the Ajaccio population is quite surprising, as in

all other Corsican populations, with the exception of Pertusato, the H07 sub-haplotypes dominated. A possible explanation for this could be that considering the anthropogenic context of the sampled population, genetic material was introduced from other locations, additional samples should be analysed to confirm this hypothesis.

The presence of two clusters was confirmed by STRUCTURE analysis and clustering analysis (using Bruvo distance and Ward's linkage method). The admixture of genetic material from two gene pools was observed in Ajaccio, Corte, Punta di a Vacca Morta and Sisco. The diversity due to differentiation ( $D_{st}$ ) was lower than the gene diversity within each population ( $H_s = 0.617$ ), suggesting that most of the genetic variability was found within populations. Our results are in agreement with the studies of Ninčević et al. (2021) and Marini et al. (2023), who analysed *H. italicum* populations from the eastern Adriatic region and the Tuscan Archipelago, respectively, and observed higher variability at the level of individual populations. Considering the fact that *H. italicum* is an outcrossing, entomophilous and anemochorous species (Galbany-Casals et al., 2011), and seeds dispersed by air flows could play an important role in its distribution and contribute to gene flow within and between populations. On the other hand, the high number of private haplotypes observed in the populations from Cape Kamenjak and Corsica could reflect selection of specific alleles due to adaptations to ecological and environmental challenges.

All primer pairs were able to amplify the DNA of all tested *H. litoreum* and *H. arenarium* samples, which means that the transferability rates of the new *H. italicum* cpSSR markers to other *Helichrysum* species are 100%. Furthermore, genetic analysis with cpSSRs confirmed a clear differentiation between *H. italicum*, *H. litoreum* and *H. arenarium*, and unique alleles were observed in each species. The identification of H01 and H05 haplotypes in *H. italicum* showing similarity to *H. litoreum* haplotypes indicates the usefulness of the newly developed cpSSR markers also for the possible detection of hybrids.

## 5. Conclusions

Based on the considerable amount of genetic variation detected within *H. italicum* populations and the high private haplotype richness, it can be summarized that the newly developed cpSSR markers have significant potential for identifying plant material for breeding, as well as for other purposes, such as planting material traceability and natural resource conservation. Similarly, private alleles and haplotypes belonging to different *Helichrysum* species confirm the usefulness of new cpSSRs for species identification, which will aid in identifying off-type species during the selection process.

**Funding:** This work was financially supported by the Slovenian Research Agency, the research program “Conservation biology, from molecules to ecosystem” [P1-0386]; the applied project, “Production, isolation and formulation of health beneficial substances from *Helichrysum italicum* for application in cosmetic industry” [L2-4430], founded by the Slovenian Research Agency and Afrodita Ltd.; and the bilateral cooperation project within the framework of the PROTEUS programme: Genetic & chemical diversity of Slovenian & Corsican immortelle [BI-FR/23-24-PROTEUS-003], founded by the Slovenian Research Agency (MVZI) and by the French Institute of Slovenia, the French Ministries of Europe and Foreign Affairs (MEAE), and of Higher Education and Research (MESR).

## CRedit authorship contribution statement

**Matjaž Hladnik:** Methodology, Formal analysis, Investigation, Writing—original draft preparation, Writing—review and editing, Visualization. **Alenka Baruca Arbeiter:** Formal analysis, Investigation, Resources, Writing—original draft preparation, Writing—review and editing, Funding acquisition. **Félix Tomi:** Methodology, Resources, Writing—original draft preparation, Writing—review and editing. **Marc Gibernau:** Methodology, Resources, Writing—original draft preparation, Writing—review and editing, Funding acquisition. **Slavko Brana:** Resources, Writing—review and

editing. **Dunja Bandelj**: Conceptualization, Methodology, Resources, Writing—original draft preparation, Writing—review and editing, Supervision, Project administration, Funding acquisition.

## Declaration of Competing Interest

The authors declare no conflicts of interest.

## Acknowledgments

The authors acknowledge the French Ministry for an Ecological and Solidary Transition (Ministère de la Transition écologique et de la Cohésion des territoires) for the declaration (NOR: TREL2002508S / 338) for access to genetic material of *Helichrysum italicum* from the island of Corsica for sharing the benefits arising from their utilization.

## REFERENCES

- Aghababayan, M., Greuter, W., Mazzola, P., & Raimondo, F. M. (2009). On the taxonomy and nomenclature of *Gnaphalium angustifolium* Lam. And *Helichrysum litoreum* Guss. (Compositae). *Bocconea*, *23*, 157–163.
- Antunes Viegas, D., Palmeira-De-Oliveira, A., Salgueiro, L., Martinez-De-Oliveira, J., & Palmeira-De-Oliveira, R. (2014). *Helichrysum italicum*: From traditional use to scientific data. *Journal of Ethnopharmacology*, *151*(1), 54–65. <https://doi.org/10.1016/j.jep.2013.11.005>
- Appendino, G., Tagliatela-Scafati, O., Minassi, A., Pollastro, F., Ballero, M., Maxia, A., & Sanna, C. (2015). *Helichrysum italicum*: The Sleeping Giant of Mediterranean Herbal Medicine. *HerbalGram*, *105*, 34–45.
- Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*(1), 37–48. <https://doi.org/10.1093/oxfordjournals.molbev.a026036>
- Baruca Arbeiter, A., Hladnik, M., Jakše, J., & Bandelj, D. (2021). First set of microsatellite markers for immortelle (*Helichrysum italicum* (Roth) G. Don): A step towards the selection of the most promising genotypes for cultivation. *Industrial Crops and Products*, *162*, 113298. <https://doi.org/10.1016/j.indcrop.2021.113298>
- Beier, S., Thiel, T., Münch, T., Scholz, U., & Mascher, M. (2017). MISA-web: A web server for microsatellite prediction. *Bioinformatics*, *33*(16), 2583–2585. <https://doi.org/10.1093/bioinformatics/btx198>
- Bianchini, A., Santoni, F., Paolini, J., Bernardini, A.-F., Mouillot, D. & Costa J. (2009). Partitioning the relative contributions of inorganic plant composition and soil characteristics to the quality of *Helichrysum italicum* subsp. *italicum* (Roth) G. Don fil. Essential oil. *Chemistry & Biodiversity*, *6*, 1014–1033. <https://doi.org/10.1002/cbdv.200800328>
- Bruvo, R., Michiels, N. K., D'souza, T. G., & Schulenburg, H. (2004). A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Molecular Ecology*, *13*(7), 2101–2106. <https://doi.org/10.1111/j.1365-294X.2004.02209.x>
- Carbonell-Caballero, J., Alonso, R., Ibañez, V., Terol, J., Talon, M., & Dopazo, J. (2015). A Phylogenetic Analysis of 34 Chloroplast Genomes Elucidates the Relationships between Wild and Domestic Species within the Genus *Citrus*. *Molecular Biology and Evolution*, *32*(8), 2015–2035. <https://doi.org/10.1093/molbev/msv082>
- CBOL Plant Working Group, Hollingsworth, P. M., Forrest, L. L., Spouge, J. L., Hajibabaei, M., Ratnasingham, S., van der Bank, M., Chase, M. W., Cowan, R. S., Erickson, D. L., Fazekas, A. J., Graham, S. W., James, K. E., Kim, K.-J., Kress, W. J., Schneider, H., van AlphenStahl, J., Barrett, S. C. H., van den Berg, C., ... Little, D. P. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences*, *106*(31), 12794–12797. <https://doi.org/10.1073/pnas.0905845106>
- Cheng, Y., De Vicente, M. C., Meng, H., Guo, W., Tao, N., & Deng, X. (2005). A set of primers for

- analyzing chloroplast DNA diversity in *Citrus* and related genera. *Tree Physiology*, 25(6), 661–672. <https://doi.org/10.1093/treephys/25.6.661>
- Daniell, H., Lin, C. S., Yu, M., & Chang, W. J. (2016). Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol*, 17(1), 134. <https://doi.org/10.1186/s13059-016-1004-2>
- Ebert, D., & Peakall, R. (2009). Chloroplast simple sequence repeats (cpSSRs): Technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. *Molecular Ecology Resources*, 9(3), 673–690.
- Eliades, N.-G., & Eliades, D. G. (2009). *HAPLOTYPE ANALYSIS: Software for analysis of haplotype data*. [Computer software]. Forest Genetics and Forest Tree Breeding, Georg-August University Goettingen, Germany. <http://www.uni-goettingen.de/en/134935.html>
- Erbaş, S., Erdoğan, Ü., & Mutlucan, M. (2023). The Scent Compounds of Immortelle Ecotypes (*Helichrysum italicum* (Roth) G. Don.) grown in Türkiye and Its New Products (Absolute and Concrete). *South African Journal of Botany*, 158, 301–311. <https://doi.org/10.1016/j.sajb.2023.05.029>
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, 14(8), 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Francis, R. M. (2017). pophelper: An R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, 17(1), 27–32. <https://doi.org/10.1111/1755-0998.12509>
- Galbany-Casals, M., Blanco-Moreno, J. M., Garcia-Jacas, N., Breitwieser, I., & Smissen, R. D. (2011). Genetic variation in Mediterranean *Helichrysum italicum* (Asteraceae; Gnaphalieae): Do disjunct populations of subsp. *Microphyllum* have a common origin? *Plant Biology*, 13(4), 678–687. <https://doi.org/10.1111/j.1438-8677.2010.00411.x>
- Galbany-Casals, M., Sáez, L., & Benedí, C. (2006). A taxonomic revision of *Helichrysum* sect. *Stoechadina* (Asteraceae, Gnaphalieae). *Canadian Journal of Botany*, 84(8), 1203–1232. <https://doi.org/10.1139/B06-082>
- Guo, Q., Guo, L., Li, Y., Yang, H., Hu, X., Song, C., & Hou, X. (2022). Development and Characterization of Microsatellite Markers Based on the Chloroplast Genome of Tree Peony. *Genes*, 13(9), Article 9. <https://doi.org/10.3390/genes13091543>
- Hale, M. L., Borland, A. M., Gustafsson, M. H. G., & Wolff, K. (2004). Causes of Size Homoplasy Among Chloroplast Microsatellites in Closely Related *Clusia* Species. *Journal of Molecular Evolution*, 58(2), 182–190. <https://doi.org/10.1007/s00239-003-2540-4>
- Herrando-Moraira, S., Blanco-Moreno, J. M., Sáez, L., & Galbany-Casals, M. (2016). Re-evaluation of the *Helichrysum italicum* complex (Compositae: Gnaphalieae): A new species from majorca (Balearic Islands). *Collectanea Botanica*, 35(September), e009. <https://doi.org/10.3989/collectbot.2016.v35.009>
- Herrando-Moraira, S., Carnicero, P., Blanco-Moreno, J. M., Sáez, L., Véla, E., Vilatersana, R., & Galbany-Casals, M. (2017). Systematics and phylogeography of the mediterranean ~~*helichrysum*~~ *Helichrysum pendulum* complex (Compositae) inferred from nuclear and chloroplast DNA and morphometric analyses. *Taxon*, 66(4), 909–933. <https://doi.org/10.12705/664.7>
- Hladnik, M., Baruca Arbeiter, A., & Bandelj, D. (2023). Sequence Characterization of ITS Regions of Immortelle *Helichrysum italicum* (Roth) G. Don from the East Adriatic Coast. *Genes*, 14(2), Article 2. <https://doi.org/10.3390/genes14020480>
- Hladnik, M., Baruca Arbeiter, A., Knap, T., Jakše, J., & Bandelj, D. (2019). The complete chloroplast genome of *Helichrysum italicum* (Roth) G. Don (Asteraceae). *Mitochondrial DNA Part B*, 4(1), 1036–1037. <https://doi.org/10.1080/23802359.2019.1580156>
- Islam, M. A., Shivaraj, S. M., Kumar, V., Phad, D. S., Sonah, H., Tripathi, S. B., & Deshmukh, R. K. (2021). Development of chloroplast microsatellite markers in *Capsicum*: Insight into evolution of Bhut Jolokia - a clad of ghost chilli landraces. *INDIAN JOURNAL OF*

*GENETICS AND PLANT BREEDING*, 81(01), Article 01.

<https://doi.org/10.31742/IJGPB.81.1.10>

- Japelaghi, R. H., Haddad, R., & Garoosi, G.-A. (2011). Rapid and Efficient Isolation of High Quality Nucleic Acids from Plant Tissues Rich in Polyphenols and Polysaccharides. *Molecular Biotechnology*, 49(2), 129–137. <https://doi.org/10.1007/s12033-011-9384-8>
- Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281. <https://doi.org/10.7717/peerj.281>
- Koressaar, T., & Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics*, 23(10), 1289–1291. <https://doi.org/10.1093/bioinformatics/btm091>
- Kramberger, K., Kenig, S., Jenko Pražnikar, Z., Kočevar Glavač, N., & Barlič-Maganja, D. (2021). A Review and Evaluation of the Data Supporting Internal Use of *Helichrysum italicum*. *Plants*, 10(8), 1738. <https://doi.org/10.3390/plants10081738>
- Lācis, G., Kārklīņa, K., Bartulsons, T., Stalažs, A., Jundzis, M., Baļķe, I., Ruņģis, D., & Strautiņa, S. (2022). Genetic structure of a *Ribes* genetic resource collection: Inter- and intra- specific diversity revealed by chloroplast DNA simple sequence repeats (cpSSRs). *Scientia Horticulturae*, 304, 111285. <https://doi.org/10.1016/j.scienta.2022.111285>
- Liu, F., Movahedi, A., Yang, W., Xu, D., & Jiang, C. (2023). The complete plastid genome and characteristics analysis of *Achillea millefolium*. *Functional & Integrative Genomics*, 23(2), 192. <https://doi.org/10.1007/s10142-023-01121-0>
- Ljubičić, I., Paulik, H., & Bogdanović, S. (2020). Habitat mapping of Protected Landscape of Donji Kamenjak, Istria (Croatia). *Journal of Central European Agriculture*, 21(3), 676–685. <https://doi.org/10.5513/JCEA01/21.3.2697>
- Lukas, B., & Novak, J. (2013). The complete chloroplast genome of *Origanum vulgare* L. (Lamiaceae). *Gene*, 528(2), 163–169. <https://doi.org/10.1016/j.gene.2013.07.026>
- Maksimovic, S., Tadic, V., Skala, D., & Zizovic, I. (2017). Separation of phytochemicals from *Helichrysum italicum*: An analysis of different isolation techniques and biological activity of prepared extracts. *Phytochemistry*, 138, 9–28. <https://doi.org/10.1016/j.phytochem.2017.01.001>
- Marini, L., Bini, L., Gori, M., Biricolti, S., Galbany-Casals, M., Foggi, B., Palchetti, E., & Bruschi, P. (2023). Genetic and morphological assessment of *Helichrysum* Mill. from Tuscan Archipelago (Italy). *Scientia Horticulturae*, 321, 112360. <https://doi.org/10.1016/j.scienta.2023.112360>
- Morone-Fortunato, I., Montemurro, C., Ruta, C., Perrini, R., Sabetta, W., Blanco, A., Lorusso, E., & Avato, P. (2010). Essential oils, genetic relationships and in vitro establishment of *Helichrysum italicum* (Roth) G. Don ssp. *italicum* from wild Mediterranean germplasm. *Industrial Crops and Products*, 32(3), 639–649. <https://doi.org/10.1016/j.indcrop.2010.07.023>
- Nazar, N., Howard, C., Slater, A., & Sgamma, T. (2022). Challenges in Medicinal and Aromatic Plants DNA Barcoding—Lessons from the Lamiaceae. *Plants*, 11(1), 137. <https://doi.org/10.3390/plants11010137>
- Ninčević, T., Grdiša, M., Šatović, Z., & Jug-Dujaković, M. (2019). *Helichrysum italicum* (Roth) G. Don: Taxonomy, biological activity, biochemical and genetic diversity. *Industrial Crops and Products*, 138(February), 111487. <https://doi.org/10.1016/j.indcrop.2019.111487>
- Ninčević, T., Jug-Dujaković, M., Grdiša, M., Liber, Z., Varga, F., Pljevljakušić, D., & Šatović, Z. (2021). Population structure and adaptive variation of *Helichrysum italicum* (Roth) G. Don along eastern Adriatic temperature and precipitation gradient. *Scientific Reports*, 11(1), 1–16. <https://doi.org/10.1038/s41598-021-03548-6>
- Peakall, R., Gilmore, S., Keys, W., Morgante, M., & Rafalski, A. (1998). Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: Implications for the transferability of SSRs in plants. *Molecular Biology and Evolution*, 15(10), 1275–1287.

<https://doi.org/10.1093/oxfordjournals.molbev.a025856>

- Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—An update. *Bioinformatics*, 28(19), 2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>
- Peng, J.-Y., Zhang, X.-S., Zhang, D.-G., Wang, Y., Deng, T., Huang, X.-H., Kuang, T.-H., & Zhou, Q. (2022). Newly reported chloroplast genome of *Sinosenecio albonervius* Y. Liu & Q. E. Yang and comparative analyses with other *Sinosenecio* species. *BMC Genomics*, 23(1), 639. <https://doi.org/10.1186/s12864-022-08872-3>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959.
- Provan, J., Powell, W., & Hollingsworth, P. M. (2001). Chloroplast microsatellites: New tools for studies in plant ecology and evolution. *Trends in Ecology & Evolution*, 16(3), 142–147. [https://doi.org/10.1016/S0169-5347\(00\)02097-8](https://doi.org/10.1016/S0169-5347(00)02097-8)
- Rogstad, S. H. (1992). Saturated NaCl-CTAB Solution as a Means of Field Preservation of Leaves for DNA Analyses. *Taxon*, 41(4), 701–708. <https://doi.org/10.2307/1222395>
- Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments A poor man ' s approach to genotyping for research and high-throughput diagnostics . 18(February), 1–2.
- Shi, X., Xu, W., Wan, M., Sun, Q., Chen, Q., Zhao, C., Sun, K., & Shu, Y. (2022). Comparative analysis of chloroplast genomes of three medicinal *Carpesium* species: Genome structures and phylogenetic relationships. *PloS One*, 17(8), e0272563. <https://doi.org/10.1371/journal.pone.0272563>
- Thiel, T., Michalek, W., Varshney, R., & Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics*, 106(3), 411–422. <https://doi.org/10.1007/s00122-002-1031-0>
- Tian, Y., Liu, X., Xu, Y., Yu, B., Wang, L., & Qu, X. (2023). Comparative and phylogenetic analysis of *Asparagus meiocladus* Levl. And *Asparagus munitus* Wang et S. C. Chen plastomes and utility of plastomes mutational hotspots. *Scientific Reports*, 13, 15622. <https://doi.org/10.1038/s41598-023-42945-x>
- Tomičić, M., Magdić, M., Peršić, M., Zec Vojinović, M., & Dudaš, S. (2022). Morphometric characteristics of *Helichrysum italicum* (Roth.) G. Don from northwestern coast of Istria. *Zbornik Veleučilišta u Rijeci*, 10(1), 453–466. <https://doi.org/10.31784/zvr.10.1.25>
- Tong, Y., Wu, C.-Y., & Gao, L.-Z. (2013). Characterization of chloroplast microsatellite loci from whole chloroplast genome of *Camellia taliensis* and their utilization for evaluating genetic diversity of *Camellia reticulata* (Theaceae). *Biochemical Systematics and Ecology*, 50, 207–211. <https://doi.org/10.1016/j.bse.2013.04.003>
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—New capabilities and interfaces. *Nucleic Acids Research*, 40(15), e115. <https://doi.org/10.1093/nar/gks596>
- Vieira, M. L. C., Santini, L., Diniz, A. L., & Munhoz, C. D. F. (2016). Microsatellite markers: What they mean and why they are so useful. *Genetics and Molecular Biology*, 39(3), 312–328. <https://doi.org/10.1590/1678-4685-GMB-2016-0027>
- Zhang, D., Tu, J., Ding, X., Guan, W., Gong, L., Qiu, X., Huang, Z., & Su, H. (2023). Analysis of the chloroplast genome and phylogenetic evolution of *Bidens pilosa*. *BMC Genomics*, 24(1), 113. <https://doi.org/10.1186/s12864-023-09195-7>



## Appendix A Supplementary data

**Table A.1** Characteristics of the developed 27 mononucleotide cpSSR markers tested on four samples from different *H. italicum* populations.

Locus	Repeat motif	Primer sequences (5' -3')	Location/region	SSR position	Expected size (bp)	Allele size range (bp)
<i>cpHiUP-17</i>	(T)11	F: ACAATTTCGGTTCTGGGGTT R: GCCGAAAAAATTCCACGGAC	8992-9265	intergenic	274	203-274
<i>cpHiUP-18</i>	(A)10	F: GTCCGTGGAAGTTTTTCGGC R: CCCGGATTGAACTGGGGAA	9246-9528	intergenic	283	278-283
<i>cpHiUP-19</i>	(T)12	F: TAATGCGCCAGCCAAAATGG R: GGTATTGTGGCCATTGCGTC	12675-12790	intergenic	116	112-115
<i>cpHiUP-20</i>	(A)10	F: GGGTTCGTGTGAGTAGGCAA R: ATACCGGATCGCCCTCTACA	13264-13467	RpoB	204	206
<i>cpHiUP-21</i>	(T)10	F: TCAGAAAGAGCTCGGGACTT R: GACCAATCCACCTACCCGAC	20775-21031	RpoC2	257	256
<i>cpHiUP-22</i>	(T)10	F: CAAAGAGGGGTTTCGGGACAA R: AAGCGGCAGAAATCAGTGGA	25801-26080	intergenic	280	279
<i>cpHiUP-23</i>	(A)11	F: TTCAAGACCGGGGCTTTCAA R: ACATGGAGTCGGGGGAGAAT	35106-35332	intergenic	227	218-225
<i>cpHiUP-24</i>	(T)11	F: GATCCTACACCGCTGCTCAA R: AGGGCTTTCTACATATGCATCGT	42887-43169	ycf3 intron	283	283-290
<i>cpHiUP-25</i>	(A)10	F: CCCAATCACGAATCATGCGC R: GGGATCATTTCGCGGAGA	44102-44356	ycf3 intron	255	252-253
<i>cpHiUP-26</i>	(T)11	F: ACATGATAAATCGATGGCTTTTGC R: TGGTCATCGGTTTCGATTCCG	46501-46642	intergenic	142	141-142
<i>cpHiUP-27</i>	(A)11	F: TCGACATGCATAGGAATTCAAGA R: TCTCTTGCACTCATTTCCTCA	46701-46998	intergenic	298	297-299
<i>cpHiUP-28</i>	(T)10	F: ATTTTGTTTTGTAGCCTGCTGT R: TATGGCGCAACCCACTTCTT	55129-55306	intergenic	178	179
<i>cpHiUP-29</i>	(T)10	F: CCGGCAATCGCAATAGCTTC R: TTGGGAGATTTTCGGCGGAA	59764-59941	intergenic	178	176-177
<i>cpHiUP-30</i>	(T)13	F: GGGCTAGAAACGAGGATCAGG R: ACAACAATGTATGCATTAGTGATTCCA	64976-65268	intergenic	293	293-301

<i>cpHiUP-31</i>	(A)11	F: TGAGTTGGGCAATATTGGGGA R: AGTACCGGTGCCACAGAAAG	66944-67124	intergenic	181	182
<i>cpHiUP-32</i>	(A)10	F: ACCCATAAAGGTTTGCCCGT R: TGCATAATGCCCGTACGGTT	70114-70367	clpP intron	254	252-253
<i>cpHiUP-33</i>	(A)10	F: TCGCTATCAACCTCTTGCC R: GGAACACAACTCAGGTTTCGT	71117-71341	clpP intron	225	223-228
<i>cpHiUP-34</i>	(T)11	F: GGGTAAAGATCCGCCCGATT R: TGAAGTATCCAGGCTCCGTT	71538-71803	clpP intron	266	265-266
<i>cpHiUP-35</i>	(T)14	F: GGAACGGTCGCTTCAAAGTG R: TGGACCTTTGAATAACAGCCA	77873-78154	intergenic	282	277-283
<i>cpHiUP-36</i>	(T)13	F: TGCTACTCGAACCCACGCTTT R: AGTTTCATTAGCTCCCGAGGT	80541-80832	intergenic	292	292-294
<i>cpHiUP-37</i>	(T)11	F: GAGGCATGTTTGGAAGTGC R: GCGGGTGCAGGTTTCTTTTT	81020-81262	intergenic	243	241-244
<i>cpHiUP-38</i>	(T)10	F: AACGTGTCACAGCTTCCTCC R: GCTAGGTAAGCGCCCTGTAG	84322-84579	intergenic	258	257-258
<i>cpHiUP-39</i>	(A)10	F: TCGATTGGCTAATGCTGAGACT R: CCTTCCACGATACATCCCGT	109795-109921	Ycf1	127	127
<i>cpHiUP-40</i>	(A)10	F: TCCCACAAAAATGGAGAAATGGAC R: AGAGGCCTTCTTCTTAGTTTCAGA	112083-112306	Ycf1	224	223
<i>cpHiUP-41</i>	(T)12	F: TGGCATGAAACAACACGCAG R: TTTTGGGGCACAGTTTTGG	120178-120281	intergenic	104	101
<i>cpHiUP-42</i>	(A)10	F: AATTCTTGCAACTAGAGAATTCTCGC R: ACAAGTGTCAATTTGGAGTAGTCA	123381-123556	intergenic	176	176-178
<i>cpHiUP-43</i>	(T)10	F: ATTAGCAGCCGTTTCCAGCT R: CGAACTCCGGGCGAATATGA	136302-136572	intergenic	271	272

a) P-01



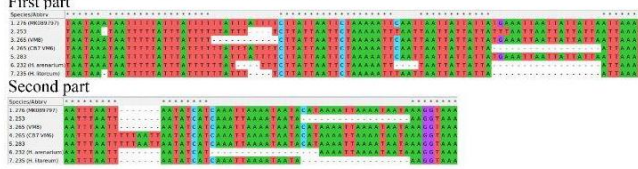
All sequenced alleles with the same length were confirmed to be identical except in sample VM8, where one SNP 2048 A/G was observed.

b) P-03 (excluded from the analysis)



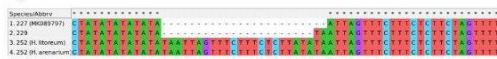
Some of the sequenced samples had 19487 C/T mutation: CB2, CB20, CB7, CN10, CN20, VM6, VM7 AND VM13.

c) P-04 (excluded from the analysis)



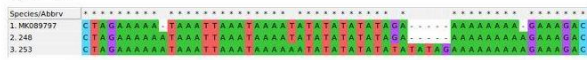
Two alleles with the same length (265) had different nucleotide sequence.

d) P-05



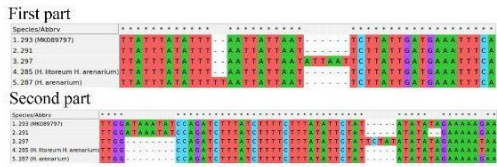
Duplication of TAAATAGTTCCTCTTAA motif in allele 252.

e) P-06



Reference sequence (MK089797) is shorter for two bases since in the mononucleotide region wrong number of As was detected with Ion Torrent technology or error occurred during chloroplast genome assembly.

e) P-07



f) P-08



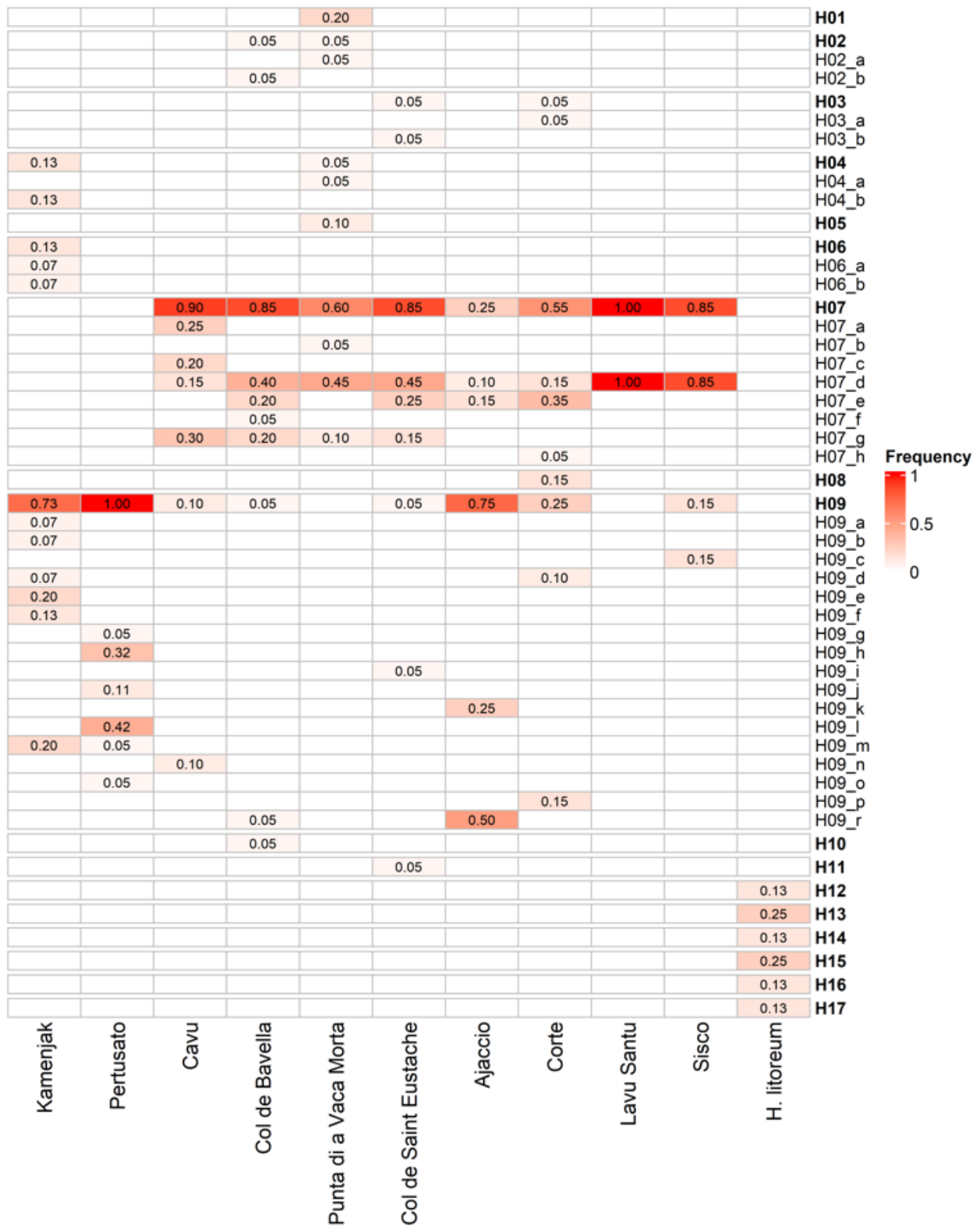
g) P-09 (excluded from the analysis)



h) P-12



Figure A.1 Multiple sequence alignment of sequenced alleles (Figure will be finalized before submission).

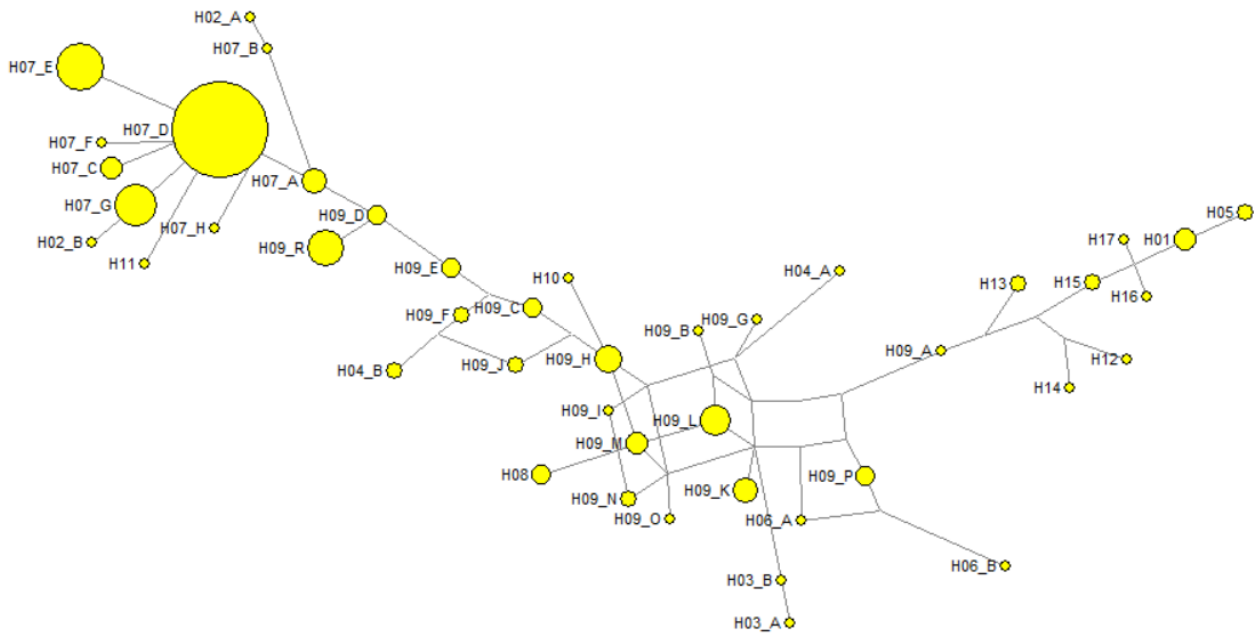


**Fig. A.2.** Heat-table with haplotype frequencies per population. Haplotypes with bold font (from H01 to H22) were obtained with 6 di- and trinucleotide SSR markers, whereas haplotypes with underscore and a letter represent sub-haplotypes obtained with additional four mononucleotide markers. Sub-haplotypes were not identified for haplotype H01, H05, H08, H10, and H11. Six *H. litoreum* haplotypes (H12-H17) were included to highlight differences between species.

**Table A.2**

List of private sub-haplotypes with their population names and corresponding frequencies.

<b>(Sub)-haplotype</b>	<b>Population</b>	<b>Frequencies</b>
H09_k	Ajaccio	0.25
H07_a	Cavu	0.25
H07_c	Cavu	0.20
H09_n	Cavu	0.10
H02_b	Col de Bavella	0.05
H07_f	Col de Bavella	0.05
H10	Col de Bavella	0.05
H03_b	Col de Saint-Eustache	0.05
H09_i	Col de Saint-Eustache	0.05
H11	Col de Saint-Eustache	0.05
H03_a	Corte	0.05
H07_h	Corte	0.05
H08	Corte	0.15
H09_p	Corte	0.15
H04_b	Kamenjak	0.13
H06_a	Kamenjak	0.07
H06_b	Kamenjak	0.07
H09_a	Kamenjak	0.07
H09_b	Kamenjak	0.07
H09_e	Kamenjak	0.20
H09_f	Kamenjak	0.13
H09_g	Pertusato	0.05
H09_h	Pertusato	0.32
H09_j	Pertusato	0.11
H09_l	Pertusato	0.42
H09_o	Pertusato	0.05
H01	Punta di a Vaca Morta	0.20
H02_a	Punta di a Vaca Morta	0.05
H04_a	Punta di a Vaca Morta	0.05
H05	Punta di a Vaca Morta	0.10
H07_b	Punta di a Vaca Morta	0.05
H09_c	Sisco	0.15



**Figure A.3.** Molecular insights into the population structure and haplotype network of *H. italicum* and *H. litoreum* based on the 10 cpSSR markers.