Comparative characterization of Santolina insularis chemotypes by essential oil composition, 5S-rRNA-NTS sequencing and EcoRV RFLP-PCR

Comparative characterization of *Santolina insularis* chemotypes by essential oil composition, 5S-rRNA-NTS sequencing and *Eco*RV RFLP-PCR

Giorgio Gnavi[a, 1], Cinzia M. Bertea[a, 1], Marianna Usai[b], Massimo E. Maffei[a,]

[a] Plant Physiology Unit, Department of Plant Biology, Innovation Centre, University of Turin, Via Quarello 11/A, 10135 Turin, Italy
[b] Dipartimento di Scienze del Farmaco, University of Sassari, Via Muroni 23/A, 07100 Sassari, Italy

Abstract

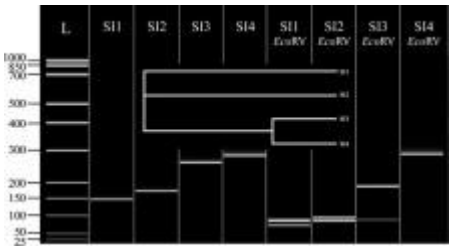
Santolina insularis (Genn ex Fiori) Arrig. is a medicinal plant whose essential oil shows antiviral and antibacterial activities and potent and selective cytotoxic activity against the human colon carcinoma cell line. The occurrence of several chemotypes makes the taxonomic identification of *S. insularis* hard to achieve.

GC–MS essential oil analyses of four chemotypes (SI1, SI2, SI3 and SI4) revealed the presence of different percentages of santolina triene, β-pinene, myrcene, β-phellandrene, artemisia ketone and *cis*-chrysanthemol, allowing a chemical discrimination. Single fragments of the 5S-rRNA-NTS region of approximately 150, 170, 260 and 280 bp were produced by SI1, SI2, SI3 and SI4, respectively, and the sequence alignment of the 5S-rRNA spacer region flanked by the 3′-and 5′-ends of the coding region confirmed a consistent difference between chemotypes. Furthermore, a PCR-RFLP method was applied. From the identified sequences, an *Eco*RV site could be found in chemotypes SI1, SI2 and SI3 in the 5S-rRNA spacer regions at 81 bp position; however, this site was absent in the chemotype SI4.

This study, by showing remarkable chemical variation in the terpenoid profile and consistent genomic difference in the 5S-rRNA spacer regions, identified four chemotypes of *S. insularis* which could be grouped into two ecotypes, based on chemical and genomic analyses. The identification of specific gene sequences of the 5S-rRNA-NTS region and of a *Eco*RV site identified in this work can be used for a rapid and precise identification of the plant chemo-/ecotypes, complementing the essential oil chemical analysis.

Graphical abstract

GC–MS analyses of *Santolina insularis* essential oils and sequencing of the 5S-rRNA-NTS region allowed the chemical and genomic characterization of four distinct chemo-/genotypes, indicating the existence of two ecotypes (mountain and coast).



Keywords

- *Santolina insularis*;
- Compositae;
- Essential oils;
- RFLP-PCR;
- 5S-rRNA non-transcribed spacer region;
- *Eco*RV site restriction mapping

---

1. Introduction

The genus *Santolina* (Compositae, tribe Anthemiadeae) is a taxonomically complex assembly of species whose classification has been subjected to numerous revisions. *Santolina insularis* (Genn. ex Fiori) Arrigoni and *Santolina corsica* Jord. et Fourr. are the two *Santolina* species reported in the Island of Sardinia (Poli et al., 1997). *S. insularis* is reported to be mainly distributed on the Gennargentu (Central Sardinia) and Marganai-Linas massifs (South-West Sardinia) (Pignatti, 1982) and the occurrence of several chemotypes makes its identification hard to achieve. *S. insularis* is known since ancient times in the Sardinian folk medicine as an intestinal vermifuge against horse strongyloidiasis and as a parasite repellent (Ballero et al., 1994). Several products obtained from *S. insularis* have been investigated for their biological activities. The essential oil obtained *in toto* from *S. insularis* has been investigated for its antiviral activity on herpes simplex type 1 (HSV-1) and type 2 (HSV-2) *in vitro*, with the antiviral activity mainly due to direct virucidal effects (De Logu et al., 2000 and Valenti et al., 2001). The biological activity of *S. insularis* was also demonstrated against *Staphylococcus aureus*, *Escherichia coli*, *Candida albicans*, *Candida tropicalis* and *Cryptococcus neoformans* (Cherchi et al., 2001). Phytochemical analysis of an acetone extract from the defatted aerial parts of *S. insularis* revealed the plant to be a prolific producer of terpenoids, characterized by several eudesmane sesquiterpenoids, chrysanthemane monoterpenoids, and a mosquito-repellent *p*-menthane derivative (Fattorusso et al., 2004). Some germacrane sesquiterpenes also showed a potent and selective cytotoxic activity against the human colon carcinoma cell line (Appendino et al., 2005). Furthermore, bioactivity-guided fractionation of the methanol extract from the leaves of *S. insularis* led to the isolation of a xanthone, together with some flavonoids. Among the latter, the most active compound, luteolin, was shown to actively prevent ear oedema (Cottiglia et al., 2005).

S. insularis is thus an important source of bioactive compounds, but at the same time has a high compositional variability. It is known that the chemical composition of essential oils is affected by several factors. According to Barra (2009), endogenous (e.g., anatomical, biochemical and physiological characteristics of the plants) and exogenous (i.e., environmental) factors lead to ecotypes or chemotypes in the same plant species. Chemotaxonomy has been widely used to classify plants with essential oils characterized by intra-specific chemical polymorphism. Since chemotypes are frequently genotypes, the application of DNA barcoding, coupled with powerful

statistical methods, appeared to be useful in the further discrimination of different genotypes. Recently, the phylogenetic relationship of some higher plant species has been evaluated using sequences of a 5S-rRNA gene spacer region. The 5S-rRNA is a component of all ribosomes except in the mitochondria of certain species (Brown and Carlson, 1997). In all higher eukaryotes, 5S-rRNA is transcribed from hundreds to thousands of genes. Genes encoding 5S-rRNA are located separately from the 18S-26S rRNA gene clusters and organized into tandem repeats with alternative arrays of sequences coding 5S-rRNA and non-transcribed spacers (NTS) in one or more sites in the genome (Park et al., 2000). The high level of conservation of the 5S-rRNA gene is associated to the precise function of 5S-rRNA as a component of the large ribosomal subunit in all eukaryotic organisms. Some regions of the gene are more conserved than others, which are explained by the regulation of 5S-rRNA transcription (Negi et al., 2002). Sequence conservation of coding region and high divergence in the spacer regions provided a good model for studying the organization and evolution of multigenes in different plant species (Scoles et al., 1988 and Cox et al., 1992). Based on these assumptions, the variation in the NTS region has been used in a number of plant species to evaluate intra-specific variation, mapping 5S-rDNA arrays, genome evolution, and phylogenetic reconstruction (Udovicic et al., 1995, Baker et al., 2000, Negi et al., 2002 and Gnavi et al., in press). Bertea et al. (2005) showed that molecular approaches are a powerful tool to distinguish the *Acorus calamus* diploid β-asarone-free cytotype from the other cytotypes containing it. The same group also used specific *Salvia divinorum* primers designed on the sequence of the 5S-rRNA gene spacer region (Bertea et al., 2006) to develop a Real-Time PCR-based mathematical model to quantify *S. divinorum* in commercial plant samples or hallucinogenic preparation (Luciano et al., 2007). More recently, a combination of biomolecular and chemical techniques allowed to characterize chemotypes of *Artemisia umbelliformis*, complementing the analysis of their essential oil and sesquiterpene lactones with a molecular characterization by PCR and PCR-RFLP of the 5S-rRNA-NTS region of their genome (Rubiolo et al., 2009).

By using the same methodology, we analysed the essential oil composition of some *S. insularis* chemotypes and we further characterized the 5S-rRNA-NTS region of their genome. The results of this work show a marked difference in both oil composition and DNA fingerprinting among the different *S. insularis* chemotypes and the existence of two ecotypes.

2. Results and discussion

This work aims to characterize some chemotypes of *S. insularis* by combining results from chemical and genomic analyses of several samples from plants growing spontaneously in the island of Sardinia. Chemical analyses were performed by GC–MS of distilled essential oil, whereas genomic results were obtained throughout characterization of the 5S-rRNA-NTS region and further RFLP-PCR discrimination by using *Eco*RV.

2.1. Essential oil composition

The essential oil chemical composition of the four *S. insularis* chemotypes is characterized by the presence of monoterpene and sesquiterpene hydrocarbons and oxygenated compounds. Santolina triene **1**, β-pinene **2**, myrcene **3**, β-phellandrene **4**, artemisia ketone **5** and *cis*-chrysanthemol **6** were the most abundant compounds (Fig. 1 and Table 1).

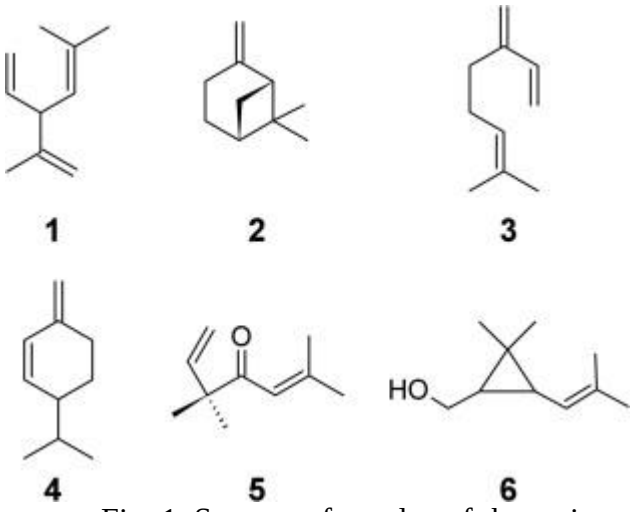


Fig. 1. Structure formulae of the main terpenoids characterizing the essential oil of the four *S. insularis* chemotypes. Santolina triene **1**, β-pinene **2**, myrcene **3**, β-phellandrene **4**, artemisia ketone **5** and *cis*-chrysanthemol **6**.

Table 1. Essential oil chemical composition of four *S. insularis* chemotypes (SI1–SI4) collected from populations growing spontaneously in the Sardinian island. At least 50 plants for each chemotype were used for the essential oil distillation. Values are expressed as μg g^{-1} dry wt (±standard error). KI = Kovats Index.

Compounds	**KI**	**SI1**	**SI2**	**SI3**	**SI4**
Santolina triene	909	304.13 (±17.40)	58.72 (±8.79)	209.59 (±13.01)	2230.13 (±81.89)
Artemisia triene	930	38.67 (±3.88)	51.26 (±16.96)	0.00	89.41 (±59.33)
α-Thujene	930	0.00	0.00	33.33 (±18.06)	49.53 (±23.78)
α-Pinene	939	0.00	58.48 (±1.55)	46.68 (±28.44)	0.00
Sabinene	975	0.00	62.06 (±0.51)	473.78 (±27.82)	345.29 (±11.23)
β-Pinene	979	166.35 (±8.03)	777.22 (±31.81)	528.64 (±15.52)	516.31 (±8.84)
Myrcene	991	24.81 (±0.98)	994.85 (±40.18)	1276.59 (±49.24)	249.55 (±1.74)
Yomogi alcohol	999	102.25 (±0.94)	412.14 (±8.71)	0.00	0.00
α-Phellandrene	1003	0.00	0.00	0.00	374.03 (±305.24)
α-Terpinene	1017	0.00	0.00	25.02 (±21.33)	0.00
p-Cymene	1025	39.45 (±4.65)	0.00	62.84 (±43.42)	54.33 (±28.09)
o-Cymene	1026	0.00	56.14 (±4.78)	57.66 (±38.99)	39.84 (±15.47)
Limonene	1029	32.87 (±1.17)	110.10 (±5.34)	0.00	0.00

Compounds	KI	SI1	SI2	SI3	SI4
β-Phellandrene	1030	763.08 (±100.44)	235.19 (±22.11)	1626.91 (±139.70)	1743.67 (±22.78)
γ-Terpinene	1060	0.00	0.00	68.29 (±48.18)	48.61 (±2.04)
Artemisia ketone	1062	959.10 (±136.06)	0.00	0.00	1323.02 (±15.25)
Artemisia alcohol	1084	28.67 (±0.17)	0.00	0.00	196.81 (±7.38)
Terpinolene	1089	31.40 (±1.95)	74.44 (±1.74)	134.59 (±7.31)	102.23 (±0.53)
cis-Chrysanthemol	1164	0.00	2737.97 (±153.86)	441.70 (±370.72)	1389.88 (±15.07)
Lavandulol	1181	0.00	112.18 (±24.70)	166.56 (±133.58)	449.86 (±10.72)
Terpinen-4-ol	1185	143.11 (±1.29)	0.00	218.70 (±178.91)	345.52 (±9.61)
α-Terpineol	1189	0.00	0.00	97.12 (±72.12)	60.11 (±1.66)
cis-Verbenyl acetate	1283	44.30 (±5.58)	0.00	0.00	183.35 (±1.62)
trans-Chrysanthemol	1287	0.00	153.57 (±2.52)	0.00	0.00
Lavandulyl acetate	1290	0.00	0.00	0.00	47.65 (±0.86)
iso-Ascaridol	1303	0.00	59.62 (±1.58)	0.00	0.00
Methyl eugenol	1404	0.00	50.78 (±0.35)	121.05 (±94.03)	0.00
trans-Caryophyllene	1419	0.00	0.00	0.00	0.00
Allo-aromadendrene	1460	0.00	0.00	0.00	52.05 (±0.96)
trans-9-epi-Caryophyllene	1466	32.59 (±2.69)	117.49 (±11.31)	64.04 (±44.53)	97.14 (±2.41)
Dehydro-sesquicineole	1471	0.00	215.90 (±0.61)	0.00	0.00
γ-Curcumene gamma	1481	52.17 (±0.48)	149.76 (±1.80)	128.72 (±99.50)	43.90 (±0.35)
ar-Curcumene	1483	99.96 (±4.18)	135.89 (±7.82)	115.58 (±89.29)	57.55 (±1.52)
Bicyclogermacrene	1500	0.00	0.00	63.01 (±42.59)	0.00
10-epi-Italicen ether	1516	33.69 (±2.27)	0.00	27.91 (±12.35)	0.00
cis-Nerolidol	1533	0.00	53.58 (±11.32)	0.00	0.00
γ-Calacorene	1546	0.00	76.86 (±12.38)	51.37 (±33.72)	53.44 (±27.16)
Spathulenol	1578	0.00	0.00	0.00	65.66

Compounds	KI	SI1	SI2	SI3	SI4
					(±37.94)
Caryophyllene oxide	1583	44.19 (±5.37)	204.24 (±3.29)	310.47 (±256.94)	50.86 (±24.76)
Caryophylla-3(15),7(14)-dien-6-ol	1641	0.00	706.88 (±8.86)	559.09 (±13.25)	357.92 (±11.01)

Chemotype SI1, from a population growing wild in Bruncu Spina (collected at 1570 m a.s.l. near Punta La Marmora mountain, central-eastern Sardinia), is characterized by high contents of artemisia ketone **5** and β-phellandrene **4**, whereas chemotype SI2, from populations growing wild in Buggerru (collected at 100 m a.s.l., Carbonia-Iglesias province, South-Western Sardinia), shows high contents of *cis*-chrysanthemol **6**, myrcene **3** and β-pinene **2**. The content of β-phellandrene **4** and myrcene **3** is particularly high in chemotype SI3, from the Marganai massif (collected at 500 m a.s.l. near Iglesias, South-Western Sardinia), whereas chemotype SI4, from populations growing wild in Genna Silana (collected at 1.017 m a.s.l. near Urzulei, Ogliastra province, central-eastern Sardinia), is characterized by high contents of santolina triene **1**, β-phellandrene **4**, artemisia ketone **5** and *cis*-chrysanthemol **6** (Table 1).

A direct comparison of the results shown in Table 1 with those found in the literature, confirms the high variability in the monotepene and sesquiterpene composition and content of *S. insularis* essential oils. Samples of *S. insularis* collected from the Marganai massif and analysed by Poli et al. (1997) contained mainly 3,3,6-trimethyl-1,5-heptadien-4-one (artemisia ketone **5**) (21%) and 10-H-cyclopropyl-1,1,7-trimethyl-4-methylen-decahydro azulene (CMDA) (12.7%). A few years later, the same results were reported by Valenti et al. (2001). Samples collected from the same area (Marganai) were also analysed by Cherchi et al. (2001) who described a different essential oil composition, characterized by the presence of myrcene **3** (14.8–17%), β-phellandrene **4** (8–9%), *trans*-β-terpineol (5–6%) and *ar*-curcumene (6–10%). The latter essential oil composition is similar to chemotype SI3 reported in our work, except for *cis*-chrysanthemol **6**, which is present in chemotype SI3 and not reported by Cherchi et al. (2001). More recently, samples collected near Arzana (at about 10 km from the sampling area of SI4), were characterized by eudesmane sesquiterpenoids and a *trans*-chrysanthemyl monoterpenoid (Fattorusso et al., 2004 and Appendino et al., 2005).

In order to assess a chemical relationship among chemotypes, a principal component analysis (PCA) was calculated. The eigenvalues of Factors 1 and 2 were 2.164 and 1.038, respectively. The percent of total variance explained by Factors 1 and 2 was 54.09% and 25.95%, respectively. Fig. 2 shows a factor loadings plot of the PCA along with factor scores of the main essential oil components. Positive loadings of Factor 1 characterize all chemotypes, whereas positive loadings of Factor 2 separate chemotype SI2 and SI3 from all other chemotypes. With regards essential oil components, positive factors scores of Factor 1 and Factor 2 separate *cis*-chrysanthemol **6**, β-pinene **2** and myrcene **3**, by placing them in the upper right quadrant where SI2 and SI3 are located. Positive factors scores of Factor 1 and negative factors scores of Factor 2 separate β-phellandrene **4**, santolina triene **1** and artemisia ketone **5** in the lower right quadrant, where SI1 and SI4 chemotypes are located. All other minor essential oil compound factor scores are spread uniformly in the upper and lower right quadrants (data not shown for the clarity of the graph).

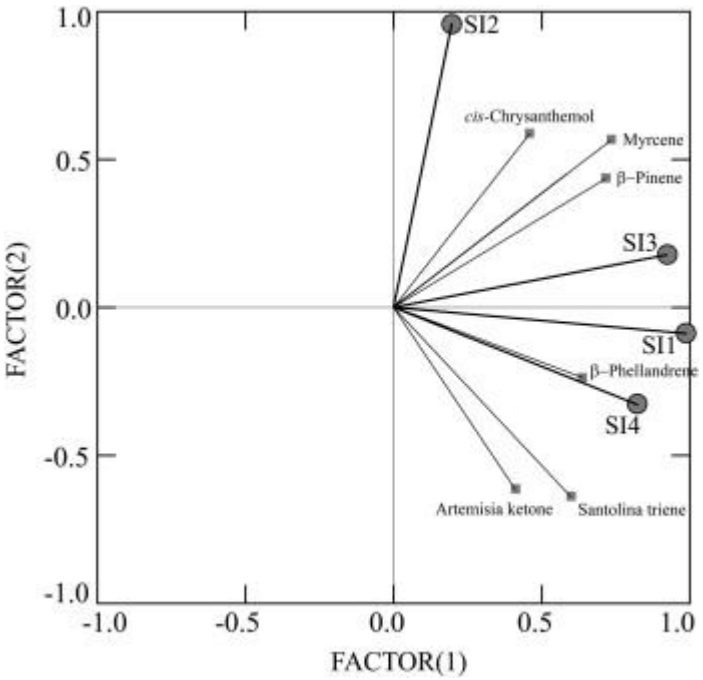


Fig. 2. Factor loadings plot of the PCA along with factor scores of the main essential oil components of *S. insularis* chemotypes. Positive loadings of Factor 1 characterize all chemotypes, whereas positive loadings of Factor 2 separate chemotype SI2 and SI3 from all other chemotypes. *cis*-Chrysanthemol **6**, β-pinene **2** and myrcene **3**, are located in the upper right quadrant along with SI2 and SI3, whereas β-phellandrene **4**, santolina triene **1** and artemisia ketone **5** are placed by PCA in the lower right quadrant, where SI1 and SI4 chemotypes are located.

The results from the PCA show that all chemotypes growing from 500 to 1570 m a.s.l. are gathered together, whereas SI2, collected close to the sea, appears to be isolated from the others. These results indicate that the four chemotypes can be divided into two ecotypically similar groups: one represented by populations growing spontaneously at middle to high elevations and another represented by populations growing almost at the sea level. The phenotypic expression of the *S. insularis* chemo-/ecotypes under study appears to be sensitive to exogenous factors which affect the essential oil composition, as it is typical in many other essential oil producing plants (Barra, 2009).

By having assessed the presence of two ecotypes responding with a distinct essential oil composition, we deepen our comparative analysis by looking at the DNA level, in the attempt to seek for possible genotypic/ecotypic similarities.

2.2. DNA fingerprinting

In order to look for a taxonomic indicator less affected by phenotypic plasticity, we extracted DNA from each chemotype and analysed the non-transcribed spacer region of the *S. insularis* 5S-rRNA gene. In higher eukaryotes, the 5S-rRNA gene is separated by simple spacers. The diversity of the spacer region can be used as an identification basis (Cai et al., 1999 and Gnavi et al., in press). Here, two primers flanking the spacer region of 5S-rRNA were used in PCR analysis of genomic DNA isolated from chemotypes SI1, SI2, SI3 and SI4, by using the same strategy that has been successfully employed for differing *A. calamus* chemotypes (Sugimoto et al., 1999), *A. calamus* cytotypes (Bertea et al., 2005), *S. divinorum* pure plants (Bertea et al., 2006) and plant mixtures (Luciano et al., 2007) as well as *A. umbelliformis* chemotypes (Rubiolo et al., 2009).

Single fragments of approximately 150, 170, 260 and 280 bp were produced by SI1, SI2, SI3 and SI4, respectively (Fig. 3). No sequence variation was observed among the individual chemotypes. Fragments derived from all chemotypes were ligated into pGEM®-T Easy vector and the nucleotide sequence was determined. The sequenced region spans 144 bp for SI1 (NCBI GenBank Accession No. GU339173), 172 bp for SI2 (NCBI GenBank Accession No. GU339174), 258 bp for SI3 (NCBI GenBank Accession No. GU339175) and 284 bp for SI4 (NCBI GenBank Accession No. GU339176) (Fig. 3). Sequence alignment of the 5S-rRNA spacer region flanked by the 3′-and 5′-ends of the coding region surprisingly showed a consistent difference between chemotypes. SI1 presented a difference of a few nucleotides with respect to SI2, but differences increased with respect to SI3 and SI4 (Fig. 3 and Fig. 4). These differences are quite consistent but are not uncommon between chemotypes or cytotypes, as it has been previously demonstrated for other plant species (Sugimoto et al., 1999, Bertea et al., 2005, Rubiolo et al., 2009 and Gnavi et al., in press). A cladogram of sequence data shows a clade that gathers SI3 and SI4, whereas SI1 and SI2 are present in distinct clades (Fig. 4, bottom graph).

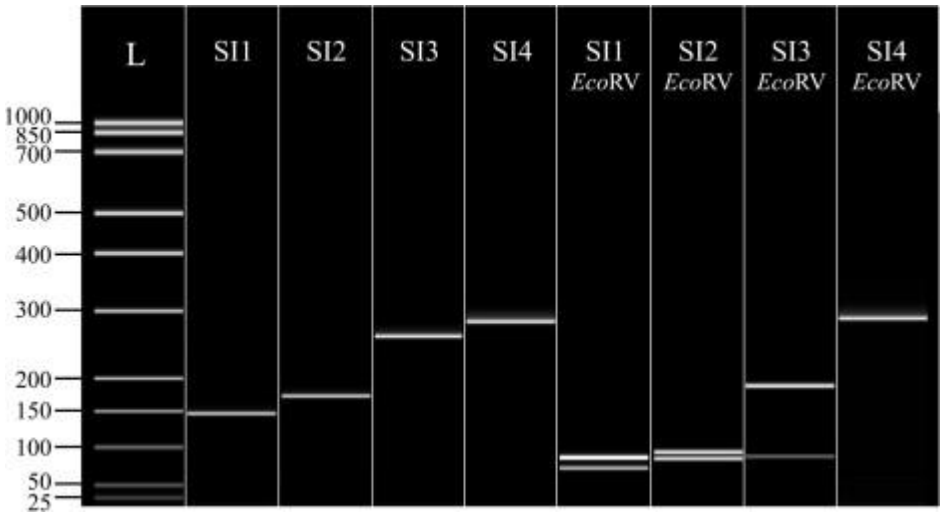


Fig. 3. PCR products generated by primers flanking the spacer region of 5S-rRNA gene using DNAs from *S. insularis* chemotypes. Single fragments of approximately 150, 170, 260 and 280 bp were produced by SI1, SI2, SI3 and SI4, respectively. PCR-RFLP analysis using *Eco*RV *S. insularis* digested PCR products gives a common band of about 80 bp in SI1, SI2 and SI3 and an additional band of about 60 bp in SI1, 90 bp in SI2, and 170 bp in SI3. Digestion of the PCR products from SI4 gives a single band of about 280 bp. L = bp markers. The PCR products were separated by using the Agilent 2100 Bioanalyzer and the DNA 1000 LabChip Kit (Agilent Technologies).

```
SI1  GTGCTTGGGCGAGAGTAGTACTGGCTTGGGTGACCCGCTGGGAAGTCCTCGTGTTGCACC  60
SI2  GTGCTTGGGCGAGAGTAGTACTGGCTTGGGTGACCCGCTGGGAAGTCCTCGTGTTGCACC  60
SI3  GTGCTTGGGCGAGAGTAGTACTGGCTTGGGTGACCCGCTGGGAAGTCCTCGTGTTGCACC  60
SI4  GTGCTTGGGCGAGAGTAGTACTAGGATGGGTGACCTCCTGGGAAGTCCTTGTGTTGCACC  60
     ********************* *  *********  ************ **********

SI1  CCTCTT-----------------TTTTTGCCAACCGATATCACAAAACTTTAACAT---  99
SI2  CCTCTT-----------------TTTTTGCCAACCGATATCACAAAACTTTGTGATTCC  102
SI3  CCTCTT-----------------TTTTTGCCAACCGATATCACAAAACTTTGTGATTCC  102
SI4  TCTTTTTGCCAAAACTGCGTGTTCTTTTTGTTAATTAGCGTAATTTGACATCTTTTTTTT  120
       * **                  ******  **      * *     * *     *

SI1  ------------------------------------------------------------  99
SI2  GTGGTTGTTTGTTTTTTTCGTCGTG---------------------------------  127
SI3  GTGGTTGTTTGTTTTTTTCGTCGTGAATACTACCCGGACTTGTACATCTTTTAAACTAGT  162
SI4  GAAAA-ATAAATTTAGATTGCTCTCATTTTTTTCGTTAATTATTGAAACTTACAACTTTA  179

SI1  ----------------------------------------------------AATA-----  103
SI2  ----------------------------------------------------AATA-----  131
SI3  TCTGAATTATGACCGGCGTCATATGCGTTTATTATATAATGCCTTTAACATAATA-----  217
SI4  TTCGGTTAATTAGCGTAATTTGACGTCTTTTTTGAAAAATAAACTTACAAATGCATATAT  239
                                                          *

SI1  ACCTTTGCAAACTATGAC----GGGTGCGATCATACCAGCACTAA  144
SI2  ACCTTTGCAAACTATGAC----GGGTGCGATCATACCAGCACTAA  172
SI3  ACCTTTGCAAACTATGAC----GGGTGCGATCATACCAGCACTAA  258
SI4  ACAATTACAACGCTTTACTTTCAGGTGCGATCATACCAGCACTAA  284
     **  ** ***    * **     **********************
```

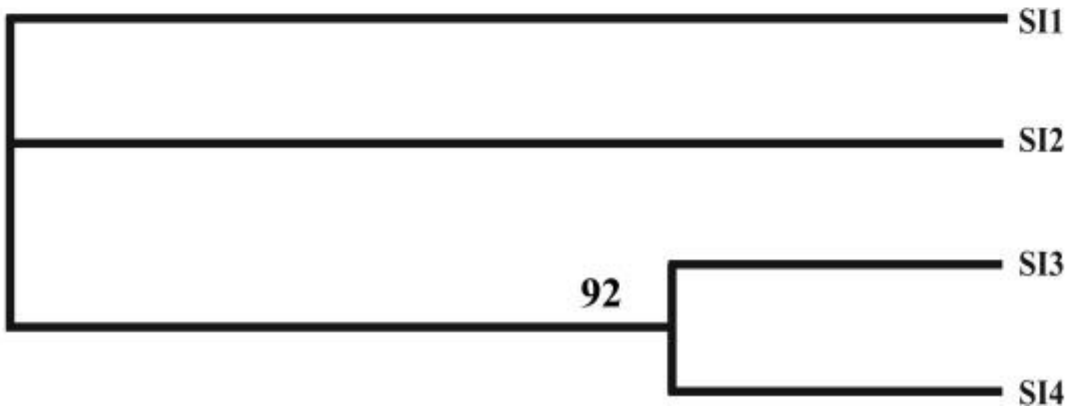


Fig. 4. Sequence alignment of the spacer regions of 5S-rRNA genes from *S. insularis* chemotypes. Primer sequences used for amplification are in bold. Identical sequences are indicated by (*). Gaps (–) are introduced for the best alignment. *Eco*RV site is evidenced by the dotted squared box. The bottom graph shows the cladogram performed on sequence data, evidencing a clade gathering SI3 and SI4, and two other clades represented by SI1 and SI2.

In order to better characterize those genotypes showing DNA fragments of similar size (i.e., SI1 and SI2), a PCR-RFLP method was applied. From the identified sequences, an *Eco*RV site could be found in chemotypes SI1, SI2 and SI3 in 5S-rRNA spacer regions at 81 bp position, but this site was absent in SI4 5S-rRNA spacer region (Fig. 4, dotted box). As expected, PCR products from the different chemotypes were digested by *Eco*RV, giving a common band of about 80 bp in SI1, SI2 and SI3 and an additional band of about 60 bp in SI1, 90 bp in SI2, and 170 bp in SI3 (Fig. 3). Digestion of the PCR products from SI4 gave a single band of about 280 bp, as expected (Fig. 3). RFLP-PCR using *Eco*RV allowed a much better discrimination between SI3 and SI4.

3. Conclusions

DNA analysis and phytochemical analyses (e.g., GC–MS, HPLC, LC–MS, etc.) are increasingly used hand in hand, rather than in isolation. Although it cannot be excluded the possibility that whenever the chemotypes for which a DNA fingerprinting has been assessed would mix, their DNA marker might lose their value due to meiotic rearrangements, the stability of DNA fingerprinting is anyway a solid method allowing the unequivocal identification of critical plant samples. Presently, some controversy exists over the value of DNA barcoding, largely because of the perception that this new identification method would diminish rather than enhance traditional morphology-based taxonomy. However, more and more gene sequences are now suitable for DNA barcoding of flowering plants (Kress et al., 2005).

The result of this comparative characterization confirms that *S. insularis*, a plant with a high phytochemical potential, shows a remarkable intra-specific variability, both at the genomic and gene products (essential oil components) levels. This study, by showing remarkable chemical variation in the terpenoid profile, and consistent genomic difference in the 5S-rRNA spacer regions, has identified four chemotypes of *S. insularis* which can be grouped into two ecotypes, based on chemical and genomic analyses. The habitat, in this case elevation, poses a selective pressure on *S. insularis* that can be observed not only at the phenotypic level (similar essential oil composition) but also at the genomic level (similar 5S-rRNA-NTS sequences).

Owing to the increased interest and relevance of *S. insularis* as a source of potent and selective cytotoxic activity against the human colon carcinoma (Appendino et al., 2005), the identification of specific gene sequences of the 5S-rRNA-NTS region and of a *Eco*RV site identified in this work offers a valuable tool for a rapid and precise identification of the plant chemo-/ecotypes, complementing the essential oil chemical analysis.

4. Experimental

4.1. Plant material

At least 50 flowering plants of *S. insularis* (Genn. Ex Fiori) Arrigoni were collected at the same time of plant development from populations growing spontaneously in four sites of the Sardinian island, characterized by different pedoclimatic conditions. A voucher specimen of all chemotypes is deposited in HERBARIUM S.A.S.S.A. (Department of Scienze del Farmaco) collective number 732. Plants were air dried overnight at 40 °C before distillation and extraction.

Chemotype SI1 was collected at about 1570 m a.s.l. close to Bruncu Spina, at an elevation where snow covers the land during the winter and where temperatures remain below 0 °C from November to March. During the growing season, temperatures are high during the day and cool at night. Chemotype SI2 was collected at about 100 m a.s.l. close to the sea in the municipality of Buggerru,

Carbonia-Iglesias province, growing wild on a calcareous soil. Located in the south-western Sardinia, the area is characterized by hot and dry summers and mild winters with scattered rain. Chemotype SI3 was collected in the regional park of Monte Linas-Oridda-Marganai in the Marganai area at about 500 m a.s.l. from populations growing spontaneously on calcareous soils. Climatic conditions are similar to those described for SI1, even though with higher season temperatures. Chemotype SI4 was collected at about 1017 m a.s.l. from plants growing wild at the passo Genna Silana. The location has almost the same climatic conditions as described for SI1 (see also Supplementary Figs. S1 for the map of the Sardinia island and S2 for images of *S. insularis* chemotypes).

4.2. Essential oil extraction and GC–MS analysis

Twenty grams of air-dried leaves and flower heads randomly collected from sampled plants of each chemotype were ground by using always the same ratio of leaves, stems and flowering heads and distilled in a modified Clevenger apparatus for 1 h. The essential oil was dried over a column of anhydrous $MgSO_4$ and then analysed by gas-chromatography–mass spectrometry (GC–MS). One microliter of 1:30 v:v essential oil:hexane was injected in splitless mode into a 6890 Agilent Technologies GC equipped with a Zebron 7HG-ZB-5MS column (30.0 m × 0.25 mm, film thickness 0.25 μm). Mass spectra were obtained by coupling the GC to a 5973 N Agilent Technologies Mass Detector. GC–MS conditions were as previously described (Bertea et al., 2005). At least three GC–MS analyses were performed for each chemotype.

Peak identification was performed with the Nist mass spectral search program v2.0 using the libraries NIST 98 and Adams (2001), as well as with direct comparison with pure standards. Standard curves for the quantification analysis were performed by using β-pinene, carvone, terpinene-4-ol and caryophyllene oxide.

4.3. Genomic DNA extraction, PCR amplification, subcloning and sequencing, and EcoRV PCR-RFLP

Samples of the same plant material used for GC–MS analyses (50 mg of dried or powdered material) were frozen in liquid nitrogen and ground to a fine powder with the Tissue Lyser (Qiagen). Genomic DNA was extracted from the ground powder by using the Nucleospin Plant Kit (Macherey Nagel) following manufacturer's instruction. The quantity and quality of the DNA were assessed by spectrophotometric analysis by using the Nanodrop ND-1000 (Thermo Scientific) from several samples of the four chemotypes.

Approximately 20 ng of genomic DNA isolated from powdered leaf material were used as a template for PCR amplification with forward primer (5′-GTGCTTGGGCGAGAGTAGTA-3′) and reverse primer (5′-TTAGTGCTGGTATGATCGCA-3′) flanking the NTS of 5S-rRNA gene (Rubiolo et al., 2009). The amplification was carried out in a 50 μl reaction mixture containing 5 μl 10× PCR reaction buffer (Fermentas), 0.2 mM dNTPs, 20 pmol forward and reverse primers and 0.5 U of *Taq* DNA polymerase (Fermentas). The PCR reactions were carried out in a Whatman Biometra T-Gradient Thermalcycler. Cycling conditions consisted of an initial 4 min at 94 °C, followed by 30 s denaturing at 94 °C, 1 min annealing at 52 °C, and 1 min elongation at 72 °C repeated for 30 cycles and with 5 min final extension at 72 °C.

One microliter of the amplification reaction was analysed by capillary gel electrophoresis (CGE) using the Agilent 2100 Bioanalyzer (Agilent Technologies) and the DNA 1000 LabChip Kit (Agilent Technologies), following manufacturer's instructions. The DNA 1000 LabChip Kit provides sizing and quantitation of dsDNA fragments ranging from 25 to 1000 bp. PCR products

were also analysed by a 2% agarose gel electrophoresis and visualized by ethidium bromide staining under UV. From this gel a band of about 150 bp for SI1, 170 bp for SI2, 260 bp for SI3 and 280 bp for SI4 was purified by using the Nucleospin Extract II Kit (Macherey Nagel) and then subcloned into pGEM-T® Easy vector (Promega). The ligated products were transformed into the *E. coli* Subcloning DH5α Efficiency Competent Cells (Invitrogen). Colonies containing DNA inserts of the correct size were picked and grown overnight in 3 ml of Luria–Bertani (LB) liquid medium. The mini-preparation of plasmid DNAs were performed using QIAprep Spin Miniprep Kit (Qiagen), following manufacturer's instructions. At least 10 plasmid DNAs were employed as a template for sequencing. Both strands of DNA were sequenced at least three times and the sequences were aligned by using ClustalX software.

The purified PCR products of the 5S-rRNA gene spacer region of all *S. insularis* chemotypes were digested with 10 U of *Eco*RV (Amersham Biosciences) at 37 °C for 1 h. One microliter of digestion reactions was fractionated by CGE using the Agilent 2100 Bioanalyzer (Agilent Technologies) and DNA 1000 LabChip Kit (Agilent Technologies) following manufacturer's instructions.

4.4. Statistical analyses

At least three replicates were run for each analysis. Essential oil component amounts were analysed by principal component analysis using Systat 10.0, whereas the cladogram of gene sequences was performed with ClustalX software by using the Neighbour Joining (NJ) method. Bootstrap values were calculated from 100 resamplings of the alignment data.

Acknowledgements

The authors wish to thank S. Bossi for technical assistance during GC–MS analyses.

Appendix A. Supplementary data

Fig. S1. Map of the Sardinian island and the location were chemotypes were sampled. Chemotype SI1 was collected at about 1570 m. a.s.l. close to Bruncu Spina, at an elevation were snow covers the land during the winter and where temperatures remain below 0 °C from November to March. During the growing season, temperatures are high during the day and cool at night. Chemotype SI2 was collected at about 100 m. a.s.l. close to the sea in the municipality of Buggerru, Carbonia-Iglesias province, growing wild on a calcareous soil. Located in the south-western Sardinia, the area is characterized by hot and dry summers and mild winters with scattered rain. Chemotype SI3 was collected in the regional park of Monte

Linas-Oridda-Marganai in the Marganai area at about 500 m. a.s.l. from populations growing spontaneously on calcareous soils. Climatic conditions are similar to those described for SI1, even though with higher season temperatures. Chemotype SI4 was collected at about 1017 m. a.s.l. from plants growing wild at the passo Genna Silana. The location has almost the same climatic conditions as described for SI1.

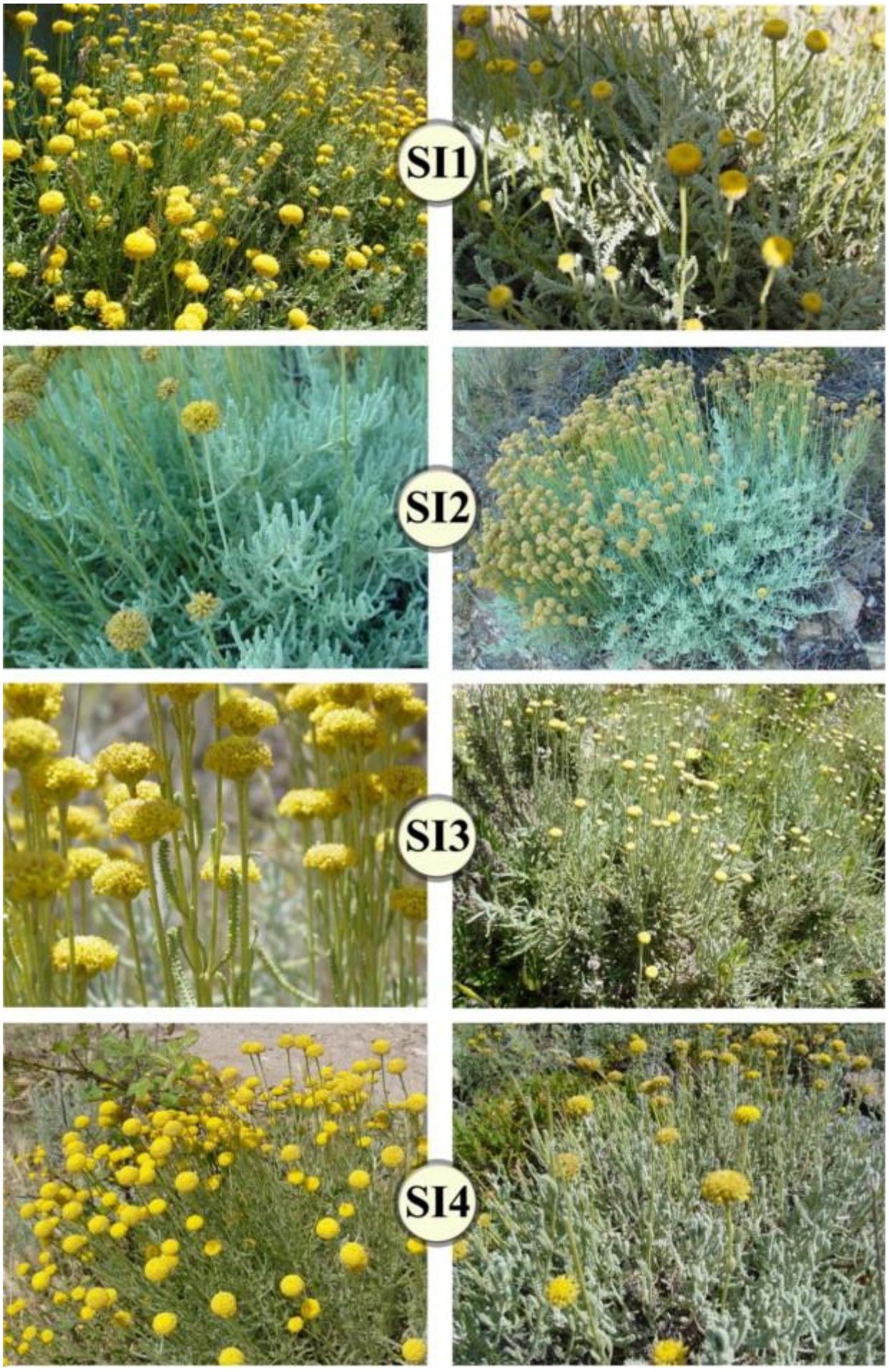

SI1
SI2
SI3
SI4

Fig. S2. Photographs of *Santolina insularis* growing spontaneously in the sites as described in Fig. S1. It is clear from these pictures that the sole morphological characters cannot be used for species distinction, justifying the use of the current approach used in this work (essential oil composition and DNA fingerprinting).

References

Adams, 2001 R.P. Adams Identification of Essential Oil Component by Gas Chromatography/Quadrupole Mass Spectroscopy Allured Publishing Corporation, Carol Stream, IL (2001)

Appendino et al., 2005 G. Appendino, G. Aviello, M. Ballero, F. Borrelli, E. Fattorusso, F. Petrucci, F.U. Santelia, O. Taglialatela-Scafati Cytotoxic germacrane sesquiterpenes from the aerial parts of *Santolina insularis* J. Nat. Prod., 68 (2005), pp. 853–857

Baker et al., 2000 W.J. Baker, T.A. Hedderson, J. Dransfield Molecular phylogenetics of Calamus (Palmae) and related rattan genera based on 5S nrDNA spacer sequence data Mol. Phylogen. Evol., 14 (2000), pp. 218–231

Ballero et al., 1994 M. Ballero, A. Bruni, G. Sacchetti, F. Poli Indagine etnofarmacolobotanica del territorio di Arzana (Sardegna orientale) Annali di Botanica, 52 (1994), pp. 489–500

Barra, 2009 A. Barra Factors affecting chemical variability of essential oils: a review of recent developments Nat. Prod. Commun., 4 (2009), pp. 1147–1154

Bertea et al., 2005 C.M. Bertea, C.M.M. Azzolin, S. Bossi, G. Doglia, M.E. Maffei Identification of an *Eco*RI restriction site for a rapid and precise determination of β-asarone-free *Acorus calamus* cytotypes Phytochemistry, 66 (2005), pp. 507–514

Bertea et al., 2006 C.M. Bertea, P. Luciano, S. Bossi, F. Leoni, C. Baiocchi, C. Medana, C.M. Azzolin, G. Temporale, M.A. Lombardozzi, M.E. Maffei PCR and PCR-RFLP of the 5S-rRNA-NTS region and salvinorin A analyses for the rapid and unequivocal determination of *Salvia divinorum* Phytochemistry, 67 (2006), pp. 371–378

Brown and Carlson, 1997 G.R. Brown, J.E. Carlson Molecular cytogenetics of the genes encoding 18s-5.8s-26s rRNA and 5s rRNA in two species of spruce (Picea) Theor. Appl. Genet., 95 (1997), pp. 1–9

Cai et al., 1999 Z.H. Cai, P. Li, T.T.X. Dong, K.W.K. Tsim Molecular diversity of 5S-rRNA spacer domain in *Fritillaria* species revealed by PCR analysis Planta Med., 65 (1999), pp. 360–364

Cherchi et al., 2001 G. Cherchi, D. Deidda, B. De Gioannis, B. Marongiu, R. Pompei, S. Porcedda Extraction of *Santolina insularis* essential oil by supercritical carbon dioxide: influence of some process parameters and biological activity Flav. Fragr. J., 16 (2001), pp. 35–43

Cottiglia et al., 2005 F. Cottiglia, L. Casu, L. Bonsignore, M. Casu, C. Floris, S. Sosa, G. Altinier, R. Della Loggia Topical anti-inflammatory activity of flavonoids and a new xanthone from *Santolina insularis* Z. Naturforsch. C, 60 (2005), pp. 63–66

Cox et al., 1992 A.V. Cox, M.D. Bennett, T.A. Dyer Use of the polymerase chain-reaction to detect spacer size heterogeneity in plant 5S-ribosomal-RNA gene clusters and to locate such clusters in wheat (*Triticum aestivum* L) Theor. Appl. Genet., 83 (1992), pp. 684–690

De Logu et al., 2000 A. De Logu, G. Loy, M.L. Pellerano, L. Bonsignore, M.L. Schivo Inactivation of HSV-1 and HSV-2 and prevention of cell-to-cell virus spread by *Santolina insularis* essential oil *in vitro* Antiviral Res., 48 (2000), pp. 177–185

Fattorusso et al., 2004 E. Fattorusso, F.U. Santelia, G. Appendino, M. Ballero, O. Taglialatela-Scafati Polyoxygenated eudesmanes and trans-chrysanthemanes from the aerial parts of *Santolina insularis* J. Nat. Prod., 67 (2004), pp. 37–41

Gnavi et al., in press Gnavi, G., Bertea, C.M., Maffei, M.E., in press. PCR, sequencing and PCR-RFLP of the 5S-rRNA-NTS region as a tool for the DNA fingerprinting of medicinal and aromatic plants. Flav. Fragr. J. 25. doi: http://dx.doi.org/10.1002/ffj.1970

Kress et al., 2005 W.J. Kress, K.J. Wurdack, E.A. Zimmer, L.A. Weigt, D.H. Janzen Use of DNA barcodes to identify flowering plants Proc. Natl. Acad. Sci. USA, 102 (2005), pp. 8369–8374

Luciano et al., 2007 P. Luciano, C.M. Bertea, G. Temporale, M.E. Maffei DNA internal standard for the quantitative determination of hallucinogenic plants in plant mixtures Forensic Sci. Int. Genet., 1 (2007), pp. 262–266

Negi et al., 2002 M.S. Negi, J. Rajagopal, N. Chauhan, R. Cronn, M. Lakshmikumaran Length and sequence heterogeneity in 5S rDNA of *Populus deltoides* Genome, 45 (2002), pp. 1181–1188

Park et al., 2000 Y.K. Park, K.C. Park, C.H. Park, N.S. Kim Chromosomal localization and sequence variation of 5S rRNA gene in five *Capsicum species* Mol. Cells, 10 (2000), pp. 18–24

Pignatti, 1982 S. Pignatti Flora d'Italia Edagricole, Bologna (1982)

Poli et al., 1997 F. Poli, L. Bonsignore, G. Loy, G. Sacchetti, M. Ballero Comparison between the essential oils of *Santolina insularis* (Genn ex Fiori) Arrigoni and *Santolina corsica* Jord et Fourr from the island of Sardinia (Italy) J. Ethnopharmacol., 56 (1997), pp. 201–208

Rubiolo et al., 2009 P. Rubiolo, M. Matteodo, C. Bicchi, G. Appendino, G. Gnavi, C. Bertea, M.E. Maffei Chemical and biomolecular characterization of *Artemisia umbelliformis* Lam., an important ingredient of the alpine liqueur "Genepi" J. Agric. Food Chem., 57 (2009), pp. 3436–3443

Scoles et al., 1988 G.I. Scoles, B.S. Gill, Z.-Y. Xin, B.C. Clarke, C.L. McIntyre, C. Chapman, R. Appels Frequent duplication and deletion events in the 5S RNA genes and the associated spacer regions of the Triticeae Plant Syst. Evol., 160 (1988), pp. 105–122

Sugimoto et al., 1999 N. Sugimoto, F. Kiuchi, M. Mikage, M. Mori, H. Mizukami, Y. Tsuda DNA profiling of *Acorus calamus* chemotypes differing in essential oil composition Biol. Pharm. Bull., 22 (1999), pp. 481–485

Udovicic et al., 1995 F. Udovicic, G.I. Mcfadden, P.Y. Ladiges Phylogeny of *Eucalyptus* and *Angophora* based on 5S rDNA spacer sequence data Mol. Phylogen. Evol., 4 (1995), pp. 247–256

Valenti et al., 2001 D. Valenti, A. De Logu, G. Loy, C. Sinico, L. Bonsignore, F. Cottiglia, D. Garau, A.M. Fadda Liposome-incorporated *Santolina insularis* essential oil: preparation, characterization and *in vitro* antiviral activity J. Liposome Res., 11 (2001), pp. 73–90