Applications in Plant Sciences

Botanical Society of America
leading scientists and educators since 1893

PRIMER NOTE

Development of novel microsatellite markers for *Alkanna tinctoria* by comparative transcriptomics

Muhammad Ahmad[1,2], Desanka Lazic[1], Karin Hansel-Hohl[1], Christian Lexer [2], and Eva Maria Sehr[1,3]

Manuscript received 22 July 2019; revision accepted 9 September 2019.

[1] Center for Health and Bioresources, AIT Austrian Institute of Technology GmbH, 3430, Tulln, Austria

[2] Department of Botany and Biodiversity Research, Faculty of Life Sciences, University of Vienna, 1030, Vienna, Austria

[3] Author for correspondence: eva-maria.sehr@ait.ac.at

Citation: Ahmad, M., D. Lazic, K. Hansel-Hohl, C. Lexer, and E. M. Sehr. 2019. Development of novel microsatellite markers for *Alkanna tinctoria* by comparative transcriptomics. *Applications in Plant Sciences* 7(10): e11296.

doi:10.1002/aps3.11296

PREMISE: *Alkanna tinctoria* (Boraginaceae) is an important medicinal herb with its main distribution across the Mediterranean region. To reveal its genetic variation and population structure, microsatellite markers were developed and validated in four Greek populations.

METHODS AND RESULTS: RNA-Seq data of the related species *Arnebia euchroma* and *Echium plantagineum* were assembled and mined to identify conserved ortholog sets containing simple sequence repeat motifs. Fifty potential loci were identified and then tested on *A. tinctoria*, of which 17 loci were polymorphic. The number of alleles ranged from one to nine, and the levels of observed and expected heterozygosity ranged from 0.000 to 1.000 and 0.000 to 0.820, respectively. Most of these loci could be successfully amplified in eight other species of Boraginaceae (*Alkanna graeca*, *A. hellenica*, *A. sfikasiana*, *Echium vulgare*, *E. plantagineum*, *Lithospermum officinale*, *Borago officinalis*, and *Anchusa officinalis*).

CONCLUSIONS: This study provides the first set of microsatellite loci for studying the genetic variation and population structure of *A. tinctoria* and shows their potential usefulness in other Boraginaceae species.

KEY WORDS *Alkanna tinctoria*; Boraginaceae; conserved ortholog set; microsatellites; population structure.

Alkanna tinctoria Tausch (Boraginaceae) is a perennial herbaceous plant that is found across southern Europe, northern Africa, and southwestern Asia, with a central distribution across the Mediterranean region (Valdés, 2011). In Greece, its occurrence has been reported from all floristic regions including the islands (Dimopoulos et al., 2013). The use of *A. tinctoria* red root extracts as a coloring agent and in traditional medicine to treat wounds can be traced back to the period of Hippocrates and Theophrastus (Papageorgiou et al., 1999). Recently, several studies have confirmed the wound-healing, antimicrobial, anticancer, and antioxidant properties of root extracts from *A. tinctoria*, which are attributed to alkannin or shikonin produced in roots (Deng et al., 2018; Yan et al., 2019; Zhang et al., 2019). Because of these properties, these active ingredients have received increasing attention from the pharmaceutical, cosmetic, and food industries in recent years (Malik et al., 2016). However, most of the research on *A. tinctoria* until now has focused either on its chemical composition or on deciphering the function of active ingredients. The extent of genetic variation and the population structure of *A. tinctoria* in Greece or elsewhere have never been studied, which may restrict the management and planned utilization of this valuable resource. This is also evident from the National Center for Biotechnology Information (NCBI) databases where, except for barcoding sequences, no other information could be found.

Because limited sequence information is available for this species, we employed a comparative transcriptomics approach using RNA-Seq data sets of the closely related species *Echium plantagineum* L. and *Arnebia euchroma* (Royle) I. M. Johnst., both in the Boraginaceae, to identify a conserved ortholog set (COS) containing simple sequence repeat (SSR) motifs, a method that has previously been used to develop markers for different species in Fabaceae (Chapman, 2015). By employing this strategy, we have successfully developed a first set of SSR markers for *A. tinctoria*, which will facilitate future genetic diversity and gene flow studies. Additionally, we tested the cross-species transferability of these markers in eight other species of Boraginaceae.

METHODS AND RESULTS

Raw RNA-Seq reads of *E. plantagineum* (SRR4034890) and *A. euchroma* (SRR4034892) obtained from the NCBI Sequence Read Archive (SRA) were uploaded to Galaxy public server (https://usegalaxy.eu).

After filtering of reads for low-quality (Q < 30), poly-N, and adapter sequences using Trimmomatic Galaxy version 0.36.0 (Bolger et al., 2014), species-specific de novo assemblies were generated using Trinity Galaxy version 2.2.0 (Grabherr et al., 2011). The assembled contigs were then clustered (90% identity threshold) to generate 91,002 and 106,240 unigenes in *E. plantagineum* and *A. euchroma*, respectively, using CD-HIT-EST software Galaxy version 1.3 (Li and Godzik, 2006). Mining for SSR regions (≥6 di-, ≥5 tri-, ≥5 tetra-, and ≥5 pentanucleotide repeats) in each of the transcriptomes by MISA (Beier et al., 2017) resulted in 4999 (*E. plantagineum*) and 4821 (*A. euchroma*) SSR-harboring transcripts. Thirty-two to forty-five percent of SSR motifs were located in the first or last 50 bp of the transcripts and were discarded because they were not suitable for primer design. To identify the COS containing SSRs, transcripts predicted to contain SSR motifs were BLASTed against each other by running the BLAST Reciprocal Best Hits (RBH) Galaxy version 0.1.11 (Cock et al., 2015). A successful BLAST-RBH was considered as a potential COS-SSR locus. By using this strategy, 89 loci were identified. After pairwise alignments, we discarded 49 loci, either because there was no variation between aligned sequences or regions flanking SSR motifs were not suitable for primer design. Primers were designed from the remaining 50 COS-SSRs.

Fresh leaves of *A. tinctoria* (n = 67) were collected from four different locations in Greece (Appendix 1). Genomic DNA was isolated using an in-house optimized protocol based on a cetyltrimethylammonium bromide (CTAB) method described in van der Beek et al. (1992). Briefly, 50 mg of powdered leaf tissues were homogenized in 400 µL of grinding buffer (0.35 M sorbitol, 0.1 M Tris-HCl, 5 mM EDTA, 20 mM $NaHSO_3$, and 4% PVP 40; pH 8.9) followed by addition of 400 µL of lysis buffer (0.2 M Tris-HCL, 50 mM EDTA, 2 M NaCl, 2% CTAB; pH 8.6), 100 µL of sodium dodecyl sulfate (SDS; 10%), and 4 µL of Proteinase K. The mixture was incubated at 60°C for 1 h and centrifuged at 15,000 rpm and 4°C for 10 min. The supernatant was transferred to new 2-mL Eppendorf tubes containing 450 µL of buffer III (3 M CH_3COOK and 5 M CH_3COOH) and incubated on ice for 10 min, followed by centrifugation as described previously. The supernatant was precipitated with 560 µL of isopropanol by centrifugation at 15,000 rpm and 4°C for 25 min. The DNA pellet was washed with chilled 70% ethanol and was dissolved in 100 µL of double-autoclaved water containing 5 µL of RNase (10 mg/mL). For initial amplification, primers were tested on two individuals of *A. tinctoria*. For 30 primer pairs that showed bands of expected size, 16 individuals from four populations were then

TABLE 1. Characteristics of 17 polymorphic conserved ortholog set–simple sequence repeat markers developed for *Alkanna tinctoria*.[a]

Locus	Primer sequences (5′–3′)	Repeat motif	Allele size range (bp)	Putative function (Organism)	*E*-value	T_a (°C)
C2	F: GAAGCCATTTTGGAGAGAAG R: CGAGTGTSACCTCTCTGTAC	$(TCA)_6$	148–167	None	3e-164	57
C3	F: AGTTGGTCTSTAGCATTTCC R: TCTTTGCTGWCCACTCACTC	$(ACC)_8$	151–181	tRNA ligase-like (*Manihot esculenta*)	0	57
C5	F: TCATCATCAGAATATTCATCATCG R: CAAATGAAGGGACTAAACATGC	$(ATC)_6$	197–239	HM-associated protein (*Olea europea*)	3e-5	55
C9	F: CAATTTCCCAATTCTACCCC R: GAGGAACTTYTGGGAATCAGG	$(CAA)_7$	166–227	NP complex protein (*Solanum tuberosum*)	2e-144	58
C12	F: TCCAAATAAAGAAATCTACCTACG R: TGTAGAACAGACTACTATGAATGG	$(CAT)_7$	204–213	Glucosylceramidase (*Coffea arabica*)	0	50
C13	F: TTAGGTGGTGGTGGTAGTGG R: CTGACCGTGCTCCTGATCC	$(TGG)_6$	258–269	Glycine-rich cell wall structural protein 1 (*Latuca sativa*)	0.093	58
C14	F: AACTGAAGAAGAAAACAAAGGTGAC R: TTGTTGGAGCATTTGAGTC	$(TAG)_7$	142–161	None	2e-132	57
C16	F: TCACCACCCAAAAMACYACC R: TCTTCACCATAYGGYCCTCCTG	$(CCA)_6$	137–145	Polyamine transporter (*Nicotiana tomentosiformis*)	0	58
C20	F: TCTTCCCTGTTCTTGTTTCC R: CTTTCCTCACATGCAGCAC	$(TCA)_6$	108–151	Reticulata-related 4 (*Hevea brasiliensis*)	3e-165	57
C24	F: GCTATTCAAATTCAACATCACTGAAG R: CACCAAGAGGCTCTTCTGGTGA	$(ATT)_8$	233–242	B2 protein (*Sesamum indicum*)	2e-163	58
C29	F: TGGCATACCATGAAGCATTG R: GTAATGCTGCCTATGAGAGG	$(ACC)_6$	149–195	DNA-binding protein (*Nicotiana sylvestris*)	3e-95	56
C30b	F: CGATGATCGATTGAAGCGCKTCC R: GTCAACCACCCCCAAYTRATCG	$(GAG)_7$	132–148	Hypothetical protein F511 (*Dorcoceras hygrometricum*)	4e-21	56
C32	F: CAAACCCATCTTCGTATTTCRCC R: CCCACTGAACTCAACWAKGTCC	$(GAG)_5$	157–178	None	1e-139	53
C35	F: CATTGCTGAAGATCTCCAATCC R: CATTACATGCCGTACTACCACC	$(CTG)_6$	124–153	Polyadenylate-binding protein RBP47 (*Nicotiana sylvestris*)	3e-172	56
C41	F: GTAATAGCTCTAACTCAAATCAGCAG R: ACAAACTTTCCAAGCGTCTTGATAAG	$(AGC)_6$	127–142	None	7e-119	53
C42	F: GCCAAGCATTCGTCRGGGAG R: CACACTACTCTCCCTACACCC	$(TGG)_8$	230–260	F-box protein (*Nicotiana tabacum*)	0	56
C45	F: AGWMTGATGAGCAAACACAATA R: GTGGTTTTGGCTTGTTCTTG	$(TCA)_7$	235–238	DNA-binding transcription factor (*Durio zibethinus*)	2e-80	56

Note: T_a = annealing temperature.
[a]Nucleotide sequences of each locus are provided in Appendix S1.

TABLE 2. Genetic diversity indices of 17 polymorphic conserved ortholog set–simple sequence repeat markers in four Greek populations of *Alkanna tinctoria*.[a]

	AT3 (n = 17)				AT4 (n = 18)				AT9 (n = 16)				AT10 (n = 16)			
Locus	A	H_o	H_e	$H_{e(ad)}$	A	H_o	H_e	$H_{e(ad)}$	A	H_o	H_e	$H_{e(ad)}$	A	H_o	H_e	$H_{e(ad)}$
C2[b]	5	0.471	0.438	0.390	7	0.778	0.671	0.583	3	0.800	0.626	0.559	4	0.938	0.723	0.662
C3	6	0.588	0.716	0.728	9	0.722	0.819	0.810	2	0.267	0.265	0.260	2	0.188	0.216	0.213
C5[b]	5	0.294	0.704	0.684	8	0.444	0.623	0.618	7	0.667	0.713	0.624	8	0.562	0.809	0.806
C9[b]	5	0.471	0.562	0.539	9	0.111	0.269	0.258	6	0.733	0.815	0.803	6	0.562	0.680	0.681
C12	1	0.000	0.000	0.000	4	0.000	0.000	0.000	4	0.062	0.094	0.122	1	0.000	0.000	0.000
C13[b]	4	0.353	0.229	0.205	4	0.444	0.391	0.410	2	0.062	0.032	0.032	2	0.188	0.192	0.213
C14[b]	5	0.912	0.631	0.595	6	1.000	0.662	0.615	3	1.000	0.548	0.533	4	0.938	0.584	0.553
C16[b]	5	0.824	0.716	0.683	5	0.722	0.647	0.605	3	0.562	0.639	0.610	2	0.375	0.208	0.201
C20	5	0.471	0.517	0.492	6	0.722	0.652	0.648	4	0.667	0.542	0.585	6	0.438	0.742	0.704
C24	3	0.688	0.533	0.528	4	0.556	0.550	0.534	3	0.200	0.108	0.102	2	0.188	0.100	0.098
C29[b]	7	1.000	0.817	0.779	9	1.000	0.821	0.795	6	0.875	0.667	0.621	4	0.938	0.622	0.582
C30b	3	0.471	0.399	0.406	6	0.556	0.523	0.462	4	0.688	0.581	0.521	3	0.250	0.529	0.536
C32	2	0.176	0.094	0.092	3	0.111	0.058	0.058	3	0.250	0.138	0.128	2	0.125	0.066	0.065
C35[b]	5	1.000	0.711	0.654	5	1.000	0.731	0.652	2	1.000	0.516	0.516	3	0.938	0.599	0.568
C41[b]	5	1.000	0.719	0.694	6	1.000	0.649	0.606	5	0.875	0.670	0.635	4	0.938	0.632	0.571
C42	2	0.176	0.094	0.092	3	0.111	0.058	0.057	3	0.133	0.305	0.295	2	0.062	0.032	0.032
C45	2	0.059	0.030	0.030	2	0.111	0.058	0.057	1	0.000	0.000	0.000	1	0.000	0.000	0.000

Note: A = number of alleles; H_e = expected heterozygosity; $H_{e(ad)}$ = expected heterozygosity after allele dosage correction; H_o = observed heterozygosity.
[a]Voucher and locality information are provided in Appendix 1.
[b]Loci that showed tetraploid peaks.

selected to assess polymorphism. The PCR reaction was performed incorporating the FAM-labeled M13 primer according to Schuelke (2000) and consisted of 4 μL of HOT FIREPol Blend Master Mix (Solis BioDyne, Tartu, Estonia), 0.24 μM of each reverse primer, 0.24 μM of M13-labeled fluorescent dye (FAM or HEX), 0.08 μM of each forward primer modified with an M13 tail, and 10–20 ng of genomic DNA template, in a final volume of 20 μL. PCR was performed using PTC-220 DYAD Thermal Cycler (MJ Research, Waltham, Massachusetts, USA) with the following parameters: 15 min of enzyme activation at 95°C; followed by 35 cycles of denaturation at 95°C for 20 s, annealing at 53–57°C for 45 s, extension for 1 min at 72°C; with a final extension period of 15 min at 72°C. The amplified products were visualized on 2% agarose gels and, after dilution to an appropriate concentration (3 : 17 to 3 : 57), were separated on a capillary sequencer (ABI PRISM 3130xl; Applied Biosystems, Foster City, California, USA) using the GeneScan 350 internal size standard (Applied Biosystems). Allelic profiles of each individual were determined using GeneMapper version 5 (Applied Biosystems). Of the 30 loci, 20 loci were found to be polymorphic; however, peaks from three loci were difficult to interpret and were not included in further analyses (Table 1). Characteristics of 10 monomorphic SSR loci are given in Appendix 2. It is noteworthy that the allelic profiles of some loci indicated that *A. tinctoria* is a tetraploid species. This observation is consistent with Coppi et al. (2006), who also observed a 4*x* ploidy level in *A. tinctoria*. For further analysis, we treated the data as tetraploid based on the available literature evidence (Coppi et al., 2006) and our own observation of >50% of loci (nine out of 17) with multi-allelic genotypes consistent with tetraploidy (Table 2). The nucleotide sequences of all loci assembled from the RNA-Seq data set of *A. euchroma* are given in Appendices S1 and S2.

Subsequently, 67 individuals from four populations (Appendix 1) were characterized by the 17 polymorphic loci. The DNA isolation method is similar to that described above, except a lithium chloride–containing buffer (50 mM Tris-HCl, 0.7 M NaCl, 20 mM EDTA, 0.4 M LiCl, and 2% PVP 40) was used instead of the grinding and lysis buffer. In addition, 20 μL of dithiothreitol (1 M) was added before incubation at 60°C (Lefort and Douglas, 1999). PCR reaction mixture and cycling conditions were the same as described above. GenoDive version 2.023 (Meriman and van Tienderen, 2004) was used to estimate the number of alleles and the levels of observed (H_o) and expected heterozygosity (H_e), with and without allele dosage correction. Allele dosage was inferred using the maximum likelihood method implemented in GenoDive (Table 2). The number of alleles ranged from one to nine, with an average of 5.94 alleles per locus. Levels of H_e and H_o of individual loci ranged from 0.000 to 0.821 and 0.000 to 1.000 in the studied populations, respectively. We note that in-depth cytogenetic work would be needed to clarify the precise cytotype status of each accession and population. Nevertheless, H_e represents a useful estimate of gene diversity for any ploidy level, and we report values with and without allele dosage correction; therefore, we regard our estimates of diversity and information content of the markers as robust and conservative (Table 2).

The cross-species transferability of 17 markers was tested on three *Alkanna* Tausch species: *A. graeca* Boiss. & Spruner, *A. hellenica* Rech. f., and *A. sfikasiana* Tan, Vold & Strid (Appendix 1). In addition, we also tested cross-species amplification in *Echium vulgare* L., *E. plantagineum*, *Lithospermum officinale* L., *Borago officinalis* L., and *Anchusa officinalis* Gouan (Appendix 1) to ascertain whether these markers are applicable to genera other than *Alkanna*.

All primer pairs gave expected size products in *E. plantagineum*, and 14 amplified in *L. officinale*, *E. vulgare*, and all tested species of *Alkanna* (Table 3). In contrast, only ~50% of the tested primer pairs had expected size bands in *B. officinalis* and *A. officinalis* (Table 3).

CONCLUSIONS

Through mining COS markers, a first set of microsatellite markers was developed for *A. tinctoria*. The observed genetic diversity indices and level of polymorphism in *A. tinctoria* showed that these markers could be useful for genotyping and genetic structure

TABLE 3. Cross-species amplification of 17 polymorphic microsatellite markers developed for *Alkanna tinctoria* in eight different species in the Boraginaceae, with approximate sizes of amplified products given in base pairs.[a]

Locus	*Alkanna graeca* (*n* = 6)	*Alkanna hellenica* (*n* = 6)	*Alkanna sfikasiana* (*n* = 6)	*Echium vulgare* (*n* = 6)	*Echium plantagineum* (*n* = 2)	*Lithospermum officinale* (*n* = 2)	*Anchusa officinalis* (*n* = 2)	*Borago officinalis* (*n* = 2)
C2	160	160	160	160	160	160	–	–
C3	–	150	150	150	150	150	300	400
C5	230	230	230	230	230	230	–	–
C9	200	200	200	200	200	–	–	200
C12	220	220	220	220	220	220	220	220
C13	≡	≡	≡	≡	260	≡	≡	≡
C14	120	120	120	120	120	120	–	120
C16	120	120	120	120	120	120	–	–
C20	≡	≡	≡	≡	ND	ND	ND	ND
C24	230	230	230	250	250	250	230	–
C29	200	200	200	200	200	200	200	200
C30b	130	130	130	130	130	130	130	130
C32	150	150	150	–	150	150	–	–
C35	150	150	150	150	150	150	150	150
C41	130	130	130	130	130	130	130	130
C42	250	250	250	200	200	200	–	200
C45	250	250	250	250	250	250	250	250

Note: – = no amplification; ≡ = multiple bands; ND = not determined.
[a]Voucher and locality information for *A. graeca*, *A. hellenica*, *A. sfikasiana*, and *E. vulgare* are provided in Appendix 1. *Lithospermum officinale* (Clone 10 and Clone 21) and *E. plantagineum* (Clone 2 and Clone 3) individuals were provided by INOQ GmbH, Germany. *Anchusa officinalis* (accession no. BVAL-901753) and *B. officinalis* (accession no. BVAL-901040) seeds were purchased from the Austrian Agency for Health and Food Safety.

analysis in this species. Successful amplification of these loci suggests their potential utility to other related taxa in the Boraginaceae.

ACKNOWLEDGMENTS

The authors thank A. Sessitsch and our partners in MICROMETABOLITE project, especially N. Fokialakis, A. N. Assimopoulou, N. Krigas, E. Maloupa, G. Brader, and P. Roedel for the provision of plant material used in this study. This project has received financial support from the European Union's Horizon 2020 research and innovation program (grant agreement no. 721635).

AUTHOR CONTRIBUTIONS

M.A. designed the study and performed in silico analysis and lab work. D.L. and K.H.H. provided support in lab work, and C.L. and E.M.S. supervised the work. All authors read and approved the final version of the manuscript.

DATA AVAILABILITY

Nucleotide sequences for the developed markers are provided in Appendix S1 and Appendix S2.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

APPENDIX S1. Nucleotide sequences of polymorphic conserved ortholog set–simple sequence repeat (COS-SSR) markers developed in *Alkanna tinctoria*. These nucleotide sequences are assembled from the RNA-Seq data set of *Arnebia euchroma*.

APPENDIX S2. Nucleotide sequences of monomorphic conserved ortholog set–simple sequence repeat (COS-SSR) markers developed in *Alkanna tinctoria*. These nucleotide sequences are assembled from the RNA-Seq data set of *Arnebia euchroma*.

LITERATURE CITED

Beier, S., T. Thiel, T. Münch, U. Scholz, and M. Mascher. 2017. MISA-web: A web server for microsatellite prediction. *Bioinformatics* 33: 2583–2585.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.

Chapman, M. A. 2015. Transcriptome sequencing and marker development for four underutilized legumes. *Applications in Plant Sciences* 3(2): 1400111.

Cock, P. J. A., J. M. Chilton, B. Grüning, J. E. Johnson, and N. Soranzo. 2015. NCBI BLAST+ integrated into Galaxy. *GigaScience* 4: 39.

Coppi, A., F. Selvi, and M. Bigazzi. 2006. Chromosome studies in Mediterranean species of Boraginaceae. *Flora Mediterranea* 16: 253–274.

Deng, X., Q. Chen, L. Qiang, M. Chi, N. Xie, Y. Wu, M. Yao, et al. 2018. Development of a porcine full-thickness burn hypertrophic scar model and investigation of the effects of shikonin on hypertrophic scar remediation. *Frontiers in Pharmacology* 9: 590.

Dimopoulos, P., T. Raus, E. Bergmeier, T. Constantinidis, G. Iatrou, S. Kokkini, A. Strid, and D. Tzanoudakis. 2013. Vascular plants of Greece: An annotated checklist. Englera, vol. 31. Botanischer Garten und Botanisches Museum Berlin-Dahlem, Berlin, Germany.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644–652.

Lefort, F., and G. C. Douglas. 1999. An efficient micro-method of DNA isolation from mature leaves of four hardwood tree species *Acer*, *Fraxinus*, *Prunus* and *Quercus*. *Annals of Forest Science* 56: 259–263.

Li, W., and A. Godzik. 2006. CD-HIT: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22: 1658–1659.

Malik, S., S. Bhushan, M. Sharma, and P. S. Ahuja. 2016. Biotechnological approaches to the production of shikonins: A critical review with recent updates. *Critical Reviews in Biotechnology* 36: 327–340.

Meriman, P. G., and P. H. van Tienderen. 2004. Genotype and Genodive: Two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* 4: 792–794.

Papageorgiou, V. P., A. N. Assimopoulou, E. A. Couladouros, D. Hepworth, and K. C. Nicolaou. 1999. The chemistry and biology of alkannin, shikonin, and related naphthazarin natural products. *Angewandte Chemie International Edition* 38: 270–301.

Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18: 233–234.

Valdés, B. 2011. Boraginaceae. *In* Euro+Med Plantbase: The information resource for Euro-Mediterranean plant diversity. Website http://ww2.bgbm.org/EuroPlusMed/ [accessed 30 January 2019].

van der Beek, J. G., R. Verkerk, P. Zabel, and P. Lindhout. 1992. Mapping strategy for resistance genes in tomato based on RFLPs between cultivars: Cf9 (resistance to *Cladosporium fulvum*) on chromosome 1. *Theoretical and Applied Genetics* 84: 106–112.

Yan, Y., F. Tan, H. Miao, H. Wang, and Y. Cao. 2019. Effect of shikonin against *Candida albicans* biofilms. *Frontiers in Microbiology* 10: 1085.

Zhang, S., Q. Gao, W. Li, L. Zhu, Q. Shang, S. Feng, J. Jia, et al. 2019. Shikonin inhibits cancer cell cycling by targeting Cdc25s. *BMC Cancer* 19: 20.

APPENDIX 1. Geographic coordinates, locality, and voucher information for plant species used in this study.[a]

Species	Population	Geographic coordinates	Location	*N*	IPEN accession no.
Alkanna tinctoria Tausch	AT3	40.630447°N, 22.971729°E	Thessaloniki, Greece	17	GR-1-BBGK-18,6081
Alkanna tinctoria	AT4	40.64277°N, 22.99777°E	Thessaloniki, Greece	18	GR-1-BBGK-18,6091
Alkanna tinctoria	AT9	37.69021°N, 24.05341°E	Sounion, Greece	16	GR-1-BBGK-18,6135
Alkanna tinctoria	AT10	37.87581°N, 23.77331°E	Imittos, Greece	16	GR-1-BBGK-18,6136
Alkanna graeca Boiss. & Spruner	AG1	37.433121°N, 22.685267°E	Parnonas, Greece	6	GR-1-BBGK-18,6138
Alkanna hellenica Rech. f.	AH1	37.900980°N, 2.877485°E	Nea Corintheas, Greece	6	GR-1-BBGK-18,6137
Alkanna sfikasiana Tan, Vold & Strid	AS1	37.339370°N, 22.602229°E	Parnonas, Greece	6	GR-1-BBGK-18,6139
Echium vulgare L.	Ev	48.32097°N, 16.06838°E	Tulln, Austria	6	

Note: IPEN = International Plant Exchange Network; *N* = number of individuals sampled.
[a]All of the living material from *Alkanna* species used in this study is deposited and maintained at the Institute of Plant Breeding and Genetic Resources, Hellenic Agricultural Organization–Demeter (HAO-Demeter), Thessaloniki, Greece.

APPENDIX 2. Characteristics of 10 monomorphic conserved ortholog set–simple sequence repeat markers markers identified in *Alkanna tinctoria*.[a]

Locus	Primer sequences (5′–3′)	Repeat motif	Allele size (bp)	T_a (°C)
C1	F: GAAACTACCCTTCAGSAAGG	$(ATG)_7$	155	50
	R: TCCTTTTCTGACAATTTGCG			
C7	F: ACTTCAGCTCCAGCACCAC	$(GAA)_6$	129	62
	R: CCAATTGGGCAAAAACTGAG			
C8	F: TGATGARAATGATTGGCATG	$(GAT)_6$	113	57
	R: TGGAAYTTGATGATAGRAAGTCCC			
C11	F: TTATGTAGAGCTCTCAAATTCC	$(GAT)_{14}$	155	53
	R: CTGTTTCTTCATAGTATTACCTGG			
C18	F: CCCTCCTCCAAATCTTGATC	$(TC)_7$	205	56
	R: GGTGATGATGTTAGCTTACAC			
C30a	F: GCGGTACCCTCAATAAAATAAGC	$(GAG)_7$	241	56
	R: GCGCTTCAATCGATCATCGC			
C31a	F: GCCTGGGAACAAGTATAAT	$(CTCAGG)_6$	224	53
	R: TTCCAAATATTGTTCCACATATG			
C36a	F: GGATCTTCAGTTGGTACTCTGG	$(ATC)_5$	158	53
	R: AACATTGAACCAACTGAACC			
C39	F: CTTGTGGGGCTTGTAATTTATGC	$(AT)_8$	123	56
	R: GCAGAATGTTGGGGGCTATTGG			
C47	F: AGGCTAATTGGTCTGATGAAGAAG	$(TTC)_6$	181	56
	R: GCATGAGGGAAATCATTATCTG			

Note: T_a = annealing temperature.
[a]Nucleotide sequences of each locus are provided in Appendix S2.