Development of 18 microsatellite markers for *Salvia pratensis*

Karol Krak1,2, Petr Vít1,2, Jan Douda1, and Bohumil Mandák1,2,3

**PREMISE:** Microsatellite markers were developed for the perennial herb *Salvia pratensis* (Lamiaceae), a species representative of European dry grasslands. The development of microsatellite markers is needed for genetic and phylogeographical studies of species from the genus *Salvia*.

**METHODS AND RESULTS:** We used low-coverage Illumina sequencing to identify microsatellite loci. Based on these data, we have developed 18 polymorphic microsatellite markers with the number of alleles per locus ranging from two to 15. The levels of observed and expected heterozygosity ranged from 0.05 to 0.95 and from 0.05 to 0.89, respectively. The majority of the markers successfully cross-amplified in other *Salvia* species.

**CONCLUSIONS:** The markers were shown to be suitable for population genetic and phylogeographical studies in *S. pratensis* as well as in related species (*S. aethiopis*, *S. austriaca*, *S. glutinosa*, *S. nemorosa*, *S. nutans*, and *S. verticillata*) and will be used in the broader context to trace the origins of European dry grasslands.

**KEY WORDS** Lamiaceae; microsatellites; phylogeography; *Salvia pratensis*.

---

*Salvia pratensis* L. (Lamiaceae) is a perennial herb distributed throughout Europe from the Volga River in the east, to northern Spain in the west, to the southern European peninsulas (Balkan and Apennine) in the south, and to the Netherlands, Poland, and Germany in the north (Kaplan et al., 2018). A few locations are also recorded in the British Isles, where the species is protected (Kaplan et al., 2018). The species is diploid with 2n = 2x = 18 chromosomes (e.g., Van Loon, 1980). It is a typical representative of European temperate dry grassland vegetation. Grasslands represent one of the most species-rich but threatened ecosystems in temperate Europe (Habel et al., 2013). Understanding the phylogeographic history of such species is crucial for determining the postglacial history of these endangered ecosystems and is expected to facilitate the development of efficient conservation strategies.

A total of 29 microsatellite markers have already been developed for *S. officinalis* L. (Molecular Ecology Resources Primer Development Consortium et al., 2010; Radosavljević et al., 2011, 2012), a species representing a sister group to *S. pratensis* (Will and Claßen-Bockhoff, 2017). Twenty of these loci were tested for cross-amplification to *S. pratensis*, but successful cross-amplification has been reported for only five (Radosavljević et al., 2011). For a state-of-the-art computational approach, this low number of microsatellite markers is inadequate for a comprehensive phylogeographical or population genetic study. Therefore, we set out to develop a set of markers specifically for *S. pratensis*. This marker set was tested for cross-amplification in other species of *Salvia* L. (*S. aethiopis* L., *S. austriaca* Jacq., *S. glutinosa* L., *S. nemorosa* L., *S. nutans* L., and *S. verticillata* L.).

**METHODS AND RESULTS**

We used low-coverage Illumina sequencing to identify microsatellite loci for *S. pratensis*. Total genomic DNA was extracted from one individual using the DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany) and a sequencing library was prepared using the NEBNext Ultra DNA Library Prep Kit (New England Biolabs, Ipswich, Massachusetts, USA), as described in Belyayev et al. (2019). Paired-end sequencing with 2 × 300-bp reads was carried out using the Illumina MiSeq platform (Macrogen, Seoul, South Korea). The *S. pratensis* library was sequenced together with nine other libraries.

The sequencing resulted in 812,212 read pairs. The raw reads were trimmed using Trimmomatic version 0.36 (Bolger et al., 2014) to remove adapters and low-quality bases using the following settings: ILLUMINACLIP:2:30:10, LEADING:20, TRAILING:20 SLIDINGWINDOW:4:20, and MINLEN:85. The trimmed reads were mapped on the chloroplast genome of *S. miltiorrhiza* Bunge (Qian et al., 2013) to remove reads of cpDNA origin using Bowtie 2 (Langmead and Salzberg, 2012) with default settings. The unmapped reads were used for microsatellite locus filtering using SSR_pipeline 0.951 (Miller et al., 2013).
with the following microsatellite search parameter settings: minimum flanking region length 40 bp; minimum number of repeats seven, six, and five for di-, tri-, and tetranucleotides, respectively. Primers were designed for the reads containing microsatellites by running additional microsatellite filtering using MSATCOMMANDER 0.8.2 (Faircloth, 2008), using the same microsatellite search settings as in SSR_pipeline and the ‘design primers with M13 tag’ option turned on. This approach yielded 2356 candidate microsatellite loci.

Altogether, 78 candidate loci containing perfect repeat motifs were tested for amplification using seven individuals from four distinct populations (for the origin of populations, see Appendix 1). The primer tailing method, as described in Schuelke (2000), was used to amplify and fluorescently label the resulting PCR products. The PCRs were carried out in 5-μL reactions containing 1× concentration of QIAGEN Multiplex PCR Mix (QIAGEN), 0.05 μM of the forward primer, 0.2 μM of both the reverse and M13 primers, and 10 ng of template DNA. The cycling conditions were as follows: initial denaturation at 95°C for 15 min followed by 25 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and elongation at 72°C for 2 min, complemented by an additional 10 cycles with the annealing temperature decreased to 50°C, and a final elongation step at 72°C for 10 min. The PCR products were first checked by agarose gel electrophoresis. Markers yielding clear non-ladder-like banding patterns within the range of 100–500 bp in at least six out of the seven individuals analyzed were further examined on the ABI 3500 automatic sequencer (Applied Biosystems, Foster City, California, USA). For this, 1 μL of non-diluted PCR products was mixed with 12 μL of Hi-Di Formamide (Applied Biosystems) and 0.1 μL of GeneTrace 500 LIZ (Carolina Biosystems, Prague, Czech Republic) internal size standard prior to the fragment analysis. Twenty-six markers showing distinct and easily readable fragments with no more than two alleles per individual were considered for further experiments.

Variability was quantified by amplifying the suitable candidate loci using additional plant material from the above-mentioned four populations (each population represented by 19–20 individuals). During these tests, eight markers were shown to amplify more than two alleles in at least some of the individuals, indicating potential

| Locus | Primer sequences (5′–3′) | Repeat motif | Allele size range (bp) | GenBank accession no. |
|-------|--------------------------|--------------|------------------------|-----------------------|
| SP_Tri_36 | F: ACTCCCTCTCCGCATTTTC | (GTT)₉ | 160–202 | MN219563 |
| R: CGTCTGTGGTGCCCGAC |
| SP_Di_24 | F: CCAGCATTTGAGCAGGAATGG | (AG)₉ | 209–221 | MN219552 |
| R: GGACCTGACAAACCAAGAACG |
| SP_Di_25b | F: TCCATGCGAGTATCTCACC | (GT)₄ | 302–337 | MN219553 |
| R: ACAAGAGAGCAGCCTTACC |
| SP_Di_31 | F: TGATGAGTTGAGTATCCTACG | (AG)₉ | 282–332 | MN219549 |
| R: GAATGCAATGGCCGGCAG |
| SP_Di_27b | F: TCCTCCTCAGCAACTGCC | (CT)₇ | 247–352 | MN219554 |
| R: AGGCGCAATTCCAGCACACC |
| SP_Tri_34 | F: TCTCCCTCTCCAATCTCAACACC | (AGC)₉ | 223–259 | MN219562 |
| R: AGGAAGACTCCAGATCCGAC |
| SP_Di_30 | F: CCTTAGGCACCTCTGATCC | (GT)₁₁ | 156–204 | MN219556 |
| R: CAGGACCCCTAATGAATACC |
| SP_Di_02 | F: TCCAGAAATCTCTAAATCCGAC | (AG)₉ | 174–214 | MN219548 |
| R: TCCAGAAATCTCTAAATCCGAC |
| SP_Di_06c | F: CGTTGGGCGCGCCTTTCAAC | (AT)₇ | 219–236 | MN219551 |
| R: GTATCAAGCAAGACCTACCCC |
| SP_Di_04b | F: ATTCCCTCTTGAGCGGACGG | (AT)₇ | 231–260 | MN219550 |
| R: AGGATCTAATAGATAGACTCCACAAATG |
| SP_Di_08 | F: TGGCGGAGGTATTTGTTGCC | (CT)₇ | 162–208 | MN219555 |
| R: TCTGTGTCAGTATTTGCC |
| SP_Tri_17 | F: GTGCGGCTGCTGCTCAACAG | (AAT)₆ | 202–262 | MN219561 |
| R: CGAACAGCTGCCTCCTACAAAG |
| SP_Tri_23 | F: CAACAGGTTGGACACGAGG | (CTT)₉ | 230–245 | MN219564 |
| R: CTGGCACTGCGCAAAGAAATTAAG |
| SP_Tri_04b | F: CTTGTTACCTCTCACTTCCC | (ACG)₆ | 196–207 | MN219557 |
| R: TCATCTAATCCTTCATTGACCC |
| SP_Te_07b | F: ATATAAACGGTGCTCACAG | (ATG)₉ | 165–215 | MN219547 |
| R: TTCCAGTCAACCAGCAAGGG |
| SP_Tri_06 | F: CCTGAGTTGCTCAGACGAGG | (AAT)₉ | 165–195 | MN219558 |
| R: TATGGTCAGCTCAGTCTCC |
| SP_Tri_15 | F: TGAGTCAATATTCTCGCAGG | (AAG)₇ | 297–330 | MN219560 |
| R: CGCCTTCGGGTGTTGTTAG |
| SP_Tri_10 | F: TGAAACAAACAGTGGGCCGG | (ATC)₉ | 166–199 | MN219559 |
| R: AAGAACCCGCGCTGTTAAG |

aAnnealing temperature was 55°C for all loci.
bOne microvariant allele was detected in each of these markers; these microvariants differ by 1 bp from the nearest (shorter) allele, and all other alleles derived from these microvariants differ by the expected length (according to the repeat motif).
cAlleles deviating by 1 bp from the expected allele size were observed in part of the individuals from the Hungarian population. These alleles are at least 9 bp longer than the last “regular” allele.
duplication of these microsatellite loci, and were therefore discarded. The remaining 18 markers were all polymorphic (Table 1). The allele sizes were determined using the software GeneMarker 2.6.4 (SoftGenetics, State College, Pennsylvania, USA). The number of alleles, levels of observed and expected heterozygosity, inbreeding coefficient, and deviation from Hardy–Weinberg equilibrium, using Fisher's exact test, were estimated using the R package diveRsity (Keenan et al., 2013) for each locus and population (Table 2). The allele sizes were determined using the software GeneMarker 2.6.4 (SoftGenetics, State College, Pennsylvania, USA). The number of alleles, levels of observed and expected heterozygosity, inbreeding coefficient, and deviation from Hardy–Weinberg equilibrium, using Fisher's exact test, were estimated using the R package diveRsity (Keenan et al., 2013) for each locus and population (Table 2). The number of alleles per locus and population varied from two to 15. The levels of observed and expected heterozygosity per locus and population ranged from 0.05 to 0.95 and from 0.05 to 0.89, respectively. For all loci, except SP_Di_24, SP_Tri_04, and SP_Tri_06, significant deviation (P < 0.05) from Hardy–Weinberg equilibrium was observed in at least one of the populations tested. However, only one locus (SP_Tri_17) deviated in all four populations.

We further cross-amplified the newly developed markers in six other Salvia species (S. aethiopis, S. austriae, S. glutinosa, S. nemorosa, S. nutans, and S. verticillata), each represented by five individuals (Appendix 1). Markers yielding microsatellite-like peaks in at least four individuals per species were considered to successfully cross-amplify. The highest cross-amplification rate was obtained for S. nemorosa, for which 15 markers amplified (Table 3), whereas cross-amplification to the distantly related S. glutinosa (see Will and Claßen-Bockhoff, 2017) failed almost completely.
(Table 3). For the remainder of the Salvia species examined, 9–13 primer pairs yielded successful amplifications (Table 3).

CONCLUSIONS

We have developed 18 polymorphic microsatellite markers for S. pratensis, a perennial plant species representing one of the key taxa of European dry grasslands. In the near future, we will use these markers to reconstruct the phylogeography of this species in Europe and contribute to the reconstruction of the post-glacial history of European dry grasslands, which are threatened habitats of high conservation value. Furthermore, at least nine of these markers were successfully cross-amplified in five other Salvia species, providing resources for further genetic studies in the genus.

ACKNOWLEDGMENTS

The authors thank the Grant Agency of the Czech Republic (grant no. 18-03028S) for financial support. The authors also thank J. Josefirová, K. Machynková, M. Borčinová, and A. Hrdličková for help in the laboratory; K. Boublík, J. Divíšek, P. Dřevojan, T. Urfus, and M. Večeřa for providing plant material; and S. Lang for assistance with English language editing.

AUTHOR CONTRIBUTIONS

K.K., P.V., and B.M. conceived and designed the study. K.K. supervised the laboratory work. K.K., P.V., and J.D. analyzed the data. K.K. and B.M. drafted the manuscript. K.K. and B.M. drafted the manuscript. All authors reviewed the manuscript and approved its final version.

DATA AVAILABILITY

Sequences of the developed microsatellite loci were submitted to the National Center for Biotechnology Information (NCBI; accession numbers MN219547–MN219564). Illumina data have been deposited to the NCBI Sequence Read Archive (BioProject PRJNA564705, BioSample accession number SAMN12721148 [Illumina sequencing], and SRR10092015 [raw Illumina reads]).

LITERATURE CITED

Belyaev, A., J. Josefirová, M. Kalendar, K. Krak, and B. Mandák. 2019. Natural history of a satellite DNA family: From the ancestral genome component to species-specific sequences, concerted and non-concerted evolution. International Journal of Molecular Sciences 20: 1201. Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimer for Illumina sequence data. Bioinformatics 30: 2114–2120. Faircloth, B. C. 2008. MSATCOMMANDER: Detection of microsatellite repeat arrays and automated, locus-specific primer design. Molecular Ecology Resources 8: 92–94. Habel, J. C., J. Dengler, M. Janíšová, P. Török, C. Wellstein, and M. Wiezik. 2013. European grassland ecosystems: Threatened hotspots of biodiversity. Biodiversity and Conservation 22: 2131–2138. Kaplan, Z., P. Koutecký, J. Danhelka, K. Sumberová, M. Ducháček, J. Stěpánková, L. Eckr, et al. 2018. Distributions of vascular plants in the Czech Republic. Part 6. Preslia 90: 235–346. Keenan, K., P. McCinnity, T. F. Cross, W. W. Crozier, and P. A. Prodöhl. 2013. diveKIT: An R package for the estimation and exploration of population genetics parameters and their associated errors. Methods in Ecology and Evolution 4: 782–788. Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9: 357–359. Miller, M. P., B. J. Knaus, T. D. Mullins, and S. M. Haig. 2013. SSR_pipeline: A bioinformatic infrastructure for identifying microsatellites from paired-end Illumina high-throughput DNA sequencing data. Journal of Heredity 104: 881–885. Molecular Ecology Resources Primer Development Consortium, J. An, A. Bechet, Å. Berggren, S. K. Brown, M. W. Brussels, Q. Cai, A. Cassel-Lundhagen, et al. 2010. Permanent genetic resources added to molecular ecology resources database 1 October 2009 – 30 November 2009. Molecular Ecology Resources 10: 404–408. Qian, J., J. Song, H. Gao, Y. Zhu, J. Xu, X. Pang, H. Yao, et al. 2013. The complete chloroplast genome sequence of the medicinal plant Salvia miltiorrhiza. PLoS ONE 8: e57607. Radosavljević, I., Z. Satovic, Z. Liber. 2012. Development of new microsatellite markers for Salvia officinalis (Lamiaceae) and cross-amplification in closely related species. American Journal of Botany e316–e318. Radosavljević, I., Z. Satovic, J. Jakse, B. Javornik, D. Greguraš, M. Jug-Dujaković, and Z. Liber. 2011. New microsatellite markers for Salvia officinalis L. and its potential use in conservation-genetic studies of narrow endemic Salvia brachyodon Vandans. International Journal of Molecular Sciences 13: 12082–12093. Schuelke, M. 2000. An economic method for the fluorescent labelling of PCR fragments. Nature Biotechnology 18: 233–234. Van Loon, J. C. 1980. Reports. In A. Löve [ed.]. IOPB Chromosome number reports LXIX. Taxon 29: 718–720. Will, M., and R. Claffe-Backhoff. 2017. Time to split Salvia s.l. (Lamiaceae): New insights from Old World Salvia phylogeny. Molecular Phylogenetics and Evolution 109: 33–58.

APPENDIX 1. Voucher and geographic information for Salvia samples used in this study.

| Species           | Voucher no. | Location       | Latitude          | Longitude         | N  |
|-------------------|-------------|----------------|-------------------|-------------------|----|
| S. pratensis L.   | BRNU 665929 | Czech Republic | 50.41022          | 13.807042         | 20 |
| S. pratensis L.   | BRNU 667140 | Hungary, Dorog | 47.75062          | 18.74198          | 19 |
| S. pratensis L.   | BRNU 667298 | Poland, RawaClawice | 50.33827      | 20.23332          | 19 |
| S. pratensis L.   | BRNU 667319 | Germany, KÖNNeRN | 51.67254         | 11.74234          | 20 |
| S. aethiopis L.   | BRNU 150305 | Bulgaria, Prazdkhizh | 42.11396         | 24.38720          | 5  |
| S. aethiopis L.   | BRNU 150305 | Bulgaria, Prazdkhizh | 42.11396         | 24.38720          | 5  |
| S. glutinosa L.   | BRNU 1502854| Iran, Fuman    | 37.17198          | 46.97948          | 5  |
| S. nemorosa L.    | BRNU 1502806| Poland, Skalbmierz | 50.31159         | 20.40273          | 5  |
| S. nutans L.      | BRNU 1503050| Bulgaria, Prazdkhizh | 42.32028         | 24.39677          | 5  |
| S. verticillata L. | BRNU 1503077| Bulgaria, Vratsa | 43.19168          | 23.49425          | 5  |

Note: N = number of individuals.

*All herbarium specimen vouchers are stored in the herbarium of the Department of Botany and Zoology, Masaryk University, Brno, Czech Republic (BRNU).*