Adaptive Genetic Diversity in Dalmatian sage (*Salvia officinalis* L.)

addresses

1 Institute for Adriatic Crops and Karst Reclamation, Split, Croatia
*e-mail: Marija.Jug-Dujakovic@krs.hr
2 University of Zagreb, Facutly of Science, Zagreb, Croatia
3 University of Zagreb, Faculty of Agriculture, Zagreb, Croatia

Marija JUG-DUJAKOVIĆ[1*], Zlatko LIBER[2], Martina GRDIŠA[3], Sandro BOGDANOVIĆ[3], Zlatko ŠATOVIĆ[3]

summary

One of the main challenges in molecular ecology and population genetics is the detection of genes, or genomic regions, that have been targeted by natural selection. The aim of this study was to quantify the relative contribution of natural selection in shaping the genetic variation observed among Dalmatian sage (*Salvia officinalis* L.) populations. Genetic diversity of 25 Dalmatian sage populations from Croatia and Bosnia and Herzegovina, each consisting of 20 to 25 plants, was assessed using AFLP markers. Two alternative methods for the identification of F_{ST}-outlier loci have been used: the frequentist methods implemented in Mcheza and the Bayesian method implemented in BayeScan. Moreover, the spatial analysis method as implemented in Samβada was used to compute multiple univariate logistic regressions to test the probability of presence of an allelic variant for a polymorphic marker given the environmental conditions of the sampling locations. The climate data for the sampling locations were obtained from the WorldClim database. The ecological characteristics were described using 19 bioclimatic variables representing the annual trends, seasonal variations and extremes in temperature and precipitation. The comparison of the outlier marker loci detected by the three approaches could help to identify the factors that are responsible for the observed spatial structuring of genetic diversity in Dalmatian sage populations.

key words

landscape genetics
genome scan
F_{ST}-outlier loci
adaptation
Salvia officinalis L.

*Fig.1. Dalmatian sage (*Salvia officinalis *L.)*

introduction

Dalmatian sage (*Salvia officinalis* L.) is an outcrossing, insect-pollinated, perennial subshrubby plant of the family Lamiaceae. The natural distribution area of Dalmatian sage is coastal region of the western Balkan and central and southern Apennine Peninsulas while some relict populations can be found in continental parts of the Balkan Peninsula (Hedge, 1982; Pignatti, 1982; Di Pietro, 2011). Dalmatian sage is well adapted to the dry Mediterranean climate and it thrives on sunny slopes with poor, rocky soil, rising from sea level to more than 1,000 m.a.s.l. (Kintzios 2000).

In Croatia, Dalmatian sage naturally grows in a continuous range along the Adriatic coast and on islands across northwest-southeast environmental gradient, generally characterized by increasing temperatures and decreasing precipitation over a span of 700 km. Molecular analyses of genetic diversity and structure of Dalmatian sage population in Croatia using RAPD (Liber et al., 2014), AFLP (Jug-Dujaković, 2010) and SSR (Greguraš, 2013) markers showed a high within-population diversity, moderate differentiation among populations and a typical pattern of isolation-by-distance.

The aim of this study was to use a landscape genomics approach (Beaumont and Nichols, 1996; Manel et al., 2010; Joost et al., 2013) to focus on divergent selection of a widespread, thermophilous species across the environmental gradient of eastern Adriatic coast. Thus, the objectives were to (1) identify outlier loci under divergent selection and (2) assess the association between loci and bioclimatic variables of sampling locations.

materials and methods

Twenty-five natural populations of Dalmatian sage originating from Croatia (23) and Bosnia and Herzegovina (2) were sampled, each consisting of 20 to 25 plants. The seed samples are preserved *ex situ* as a part of the Collection of Medicinal and Aromatic Plants in Zagreb, Croatia (available at: cpgrd.zsr.hr).

Genomic DNA samples were extracted from dried leaves using the GenElute™ Plant Genomic DNA Miniprep Kit (Sigma-Aldrich). The plants were genotyped using four AFLP primer combinations. The products were run on an ABI 3730XL analyzer (Applied Biosystems) using the commercial GeneScan service (Macrogen). The results were analyzed using GeneMapper 4.0 software (Applied Biosystems).

Allelic frequencies at AFLP marker loci were calculated from the observed frequencies of fragments in each population using Bayesian method with non-uniform prior distribution of allele frequencies proposed by Zhivotovsky (1999), as implemented in AFLP-Surv v. 1.0 (Vekemans et al., 2002). Hereby, we assumed that the populations were in Hardy-Weinberg equilibrium ($F_{IS} = 0$) justified by the outcrossing nature of Dalmatian sage.

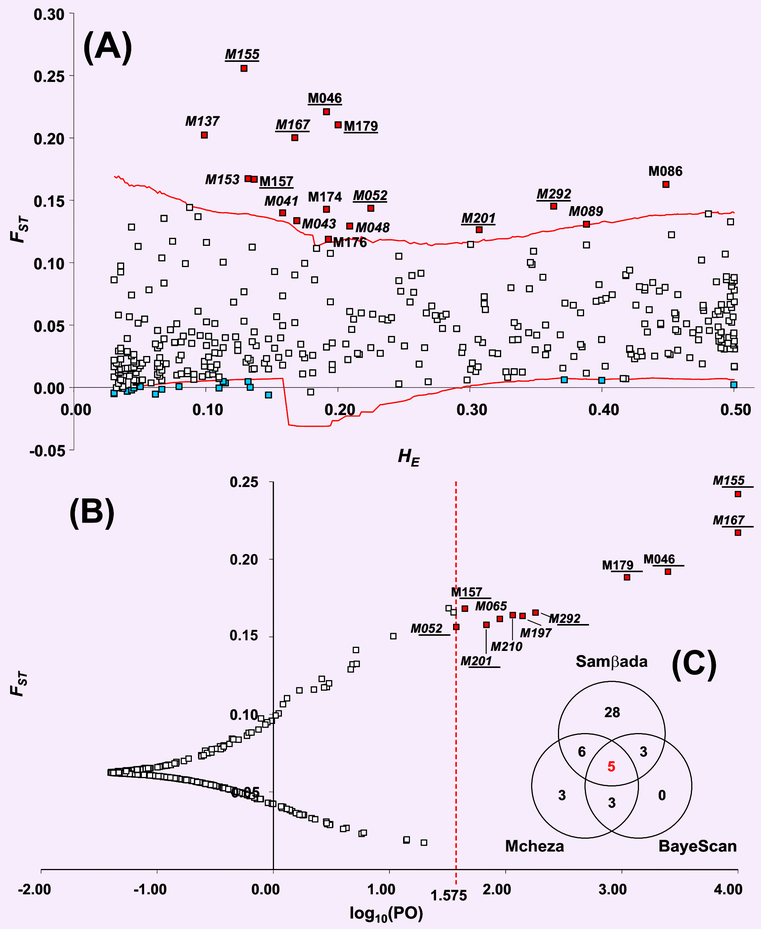
The identification of F_{ST}-outlier loci was carried out using the frequentist method of Beaumont and Nichols (1996) implemented in Mcheza (Antao and Beaumont, 2011) and the Bayesian method implemented in BayeScan ver. 2.01 (Foll and Gaggiotti, 2008). Outliers detected by both methods are more likely to represent truly adaptive regions within the genome because the two separate approaches have completely different assumptions and employ separate algorithms.

The spatial analysis method described by Joost et al. (2007) and implemented in Samβada (Stucki and Joost, 2015) was used to compute multiple univariate logistic regressions to test the probability of presence of an allelic variant for a polymorphic marker given the environmental conditions of the sampling locations. The climate data for 25 sampling sites were obtained from the WorldClim database at a spatial resolution of approximately 1 km^2 (Hijmans et al., 2005; www.worldclim.org). The ecological characteristics of the sampling sites were described using 19 bioclimatic variables (11 temperature- and 8 precipitation-related) representing the annual trends, seasonal variations and extremes in temperature and precipitation (Table 1).

Table 1. Bioclimatic variables used to detect selection signature: Total number of markers associated with each bioclimatic variable as detected using Samβada and the number of associated markers identified also by both Mcheza and BayeScan as being under directional selection

No.	Variable	No.of markers	BayeScan / MCHEZA
BIO1	Annual Mean Temperature	2	-
BIO2	Mean Diurnal Range	9	3
BIO3	Isothermality	13	2
BIO4	Temperature Seasonality	3	-
BIO5	Max Temperature of Warmest Month	3	-
BIO6	Min Temperature of Coldest Month	2	-
BIO7	Temperature Annual Range	7	3
BIO8	Mean Temperature of Wettest Quarter	8	1
BIO9	Mean Temperature of Driest Quarter	4	-
BIO10	Mean Temperature of Warmest Quarter	2	-
BIO11	Mean Temperature of Coldest Quarter	3	-
BIO12	Annual Precipitation	8	2
BIO13	Precipitation of Wettest Month	4	-
BIO14	Precipitation of Driest Month	22	3
BIO15	Precipitation Seasonality	16	3
BIO16	Precipitation of Wettest Quarter	5	-
BIO17	Precipitation of Driest Month	22	3
BIO18	Precipitation Seasonality	24	3
BIO19	Precipitation of Wettest Quarter	4	1

Fig. 1. Detection of loci under selection: (A) Mcheza: F_{ST} values of each locus plotted against its heterozygostity. The dashed lines represent the 99% confidence intervals. Loci under positive selection are indicated as red dots, those under balancing selection as light blue dots and neutral as white dots. Loci under positive selection detected also by BayeScan are underlined while those identified by Samβada are shown in italics; (B) BayeScan: F_{ST} values of each locus plotted against the $\log_{10}$ of the posterior odds (PO). The vertical line shows the critical PO used for identifying outlier markers [FDR < 0.01; PO = 37.6; $\log_{10}$(PO) = 1.575]; (C) Wenn diagram illustrating the overlap in outlier detection across three methods.



results and discussion

Four AFLP primer combinations applied on 593 individuals yielded a total of 559 polymorphic markers. The average gene diversity was relatively high ($H_w = 0.178$) ranging from 0.162 to 0.199 while the total gene diversity was $H_T = 0.186$. Wright's index of genetic differentiation was significant but low ($F_{ST} = 0.042$; $P < 0.001$). Pairwise estimates of F_{ST} ranged between 0.006 and 0.081. The results indicated a high within-population diversity and significant, but weak differentiation among populations. These results are consistent with previous results recorded in outcrossing perennial plants but not in endemic species, where much lower diversity within populations was noticed (Hamrick and Godt, 1990; Nybon and Bartish, 2000). Dalmatian sage is an endemic species of the Apennines and the Balkan Peninsula but unlike the majority of endemic species it is widespread and often abundant.

A total of 381 AFLP markers was used to identify the outlier loci (after the removal of markers with a frequency below 3% or above 97%). With a confidence level set to 99%, Mcheza detected a total of 35 outlier loci (9.19%) possibly under selection, among which 17 (4.46%) under directional and 18 (4.72%) under balancing selection (Fig. 2A). BayeScan identified 11 (2.89%) loci exceeding the threshold for very strong evidence of selection [False Discovery Rate (FDR) < 0.01; posteriror odds (PO) > 37], none of them under stabilizing selection (Fig. 2B). A set of eight (2.10%) candidate markers under directional selection was detected by both Mcheza and BayeScan.

As expected, the 19 bioclimatic variables used in the study were highly inter-correlated. Out of 171 pairwise examinations, a strong positive correlation ($r > 0.70$) was found in 29 cases while in four cases a strong negative correlation ($r < -0.70$) was identified. After calculating logistic regressions between all possible marker/bioclimatic variable pairs (a total of 7,239 models), Samβada detected 161 (2.22%) significant models involving 44 markers (11.02%) associated with one up to 11 bioclimatic variables. Bioclimatic variables associated with more than 20 markers were BIO18 Precipitation of Warmest Quarter (BIO18), Precipitation of Driest Month (BIO14), and Precipitation of Driest Quarter (BIO17) (Table 1). Out of 42 markers detected by Samβada, five were identified by both Mcheza and BayeScan as being under directional selection.

Thus, out of a total of 381 markers, five (1.31%) were identified across three methods as presented in Wenn diagram illustrating the overlap in outlier detection (Fig. 2C). Six bioclimatic variables were associated with (different combinations of) three out of five F_{ST}-outlier markers; three were precipitation-related (BIO14, BIO17, BIO18) while additional three were those describing temperature/precipitation range (BIO2 Mean Diurnal Range; BIO7 Temperature Annual Range; BIO15 Precipitation Seasonality) (Table 1).

references

Antao, T., Beaumont, M.A. 2011. Mcheza: A workbench to detect selection using dominant markers. Bioinformatics 27: 1717–1718.

Beaumont, M.A., Nichols, R.A. 1996. Evaluating loci for use in the genetic analysis of population structure. Proc. R. Soc. Lond. B Biol. Sci. 263: 1619–1626.

Di Pietro, R. 2011. New dry grassland associations from Ausoni-Aurunci mountains (central Italy) - syntaxonomical updating and discussion on the higher rank syntaxa. Hacquetia 10: 183–231.

Foll, M., Gaggiotti, O.E. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180: 977–993.

Greguraš, D., 2013. Genetic diversity and population structure of Dalmatian sage (*Salvia officinalis* L.). PhD Thesis, University of Zagreb, Faculty of Science, Zagreb, Croatia.

Hamrick, J.L., Godt, M.J.W. 1990. Allozyme diversity in plant species. In: Brown, A.H.D., Clegg, M.T., Kahler, A.L., Wier, B.S. (eds.) Plant Population Genetics, Breeding and Genetic Resources. Sinauer, Sunderland, Massachusetts, USA. pp. 43–63.

Hedge, I.C. 1972. *Salvia* L. In: Tutin, T.G., Heywood, V.H., Burges, N.A., Valentine, D.H., Walters, S.M., Webb, D.A. (eds.) Flora Europaea 3, Cambridge University Press, Cambridge, UK. pp. 188–192.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25: 1965–1978.

Joost, S., Bonin, A., Bruford, M.W., Després, L., Conord, C., Erhardt, G., Taberlet, P. 2007. A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomic approach to adaptation. Mol. Ecol. 16: 3955–3969.

Joost, S., Vuilleumier, S., Jensen, J.D., Schoville, S., Leempoel, K., Stucki, S., Widmer, I., Melodelima, Ch., Rolland, J., Manel, S. 2013. Uncovering the genetic basis of adaptive change: on the intersection of landscape genomics and theoretical population genetics. Mol. Ecol. 22: 3659–3665.

Jug-Dujaković, M. 2010. Genetic and biochemical diversity of Dalmatian sage (*Salvia officinalis* L.). PhD Thesis, University of Zagreb, Faculty of Agriculture, Zagreb, Croatia.

Kintzios, S.E. 2000. Sage: The Genus *Salvia*. Harwood Academic Publishers, Amsterdam, The Netherlands.

Liber, Z., Židovec, V., Bogdanović, S., Radosavljević, I., Pruša, M., Filipović, M., Han Dovedan, I., Jug-Dujaković, M., Satovic, Z. 2014. Genetic diversity of Dalmatian sage (*Salvia officinalis* L.) as assessed by RAPD markers. Agric. Conspec. Sci. 79(2): 77–84.

Manel, S., Joost, S., Epperson, B.K., Holderegger, R., Storfer, A., Rosenber, M.S., Scribner, K.T., Bonin, A., Fortin, M.-J. 2010. Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. Mol. Ecol. 19: 3760–3772.

Nybon, H., Bartish, I.V. 2000.. Effects of life history traits and sampling strategies on genetic diversity estimates obtained with RAPD markers in plants. Perspect Plant Ecol. Evol. Syst. 3: 93–114.

Pignatti, S. (ed.) 1982. Flora d' Italia, vol. 2. Edagricole, Bologna, Italy.

Stucki, S., Joost, S. 2015. Samβada: User manual. Version v0.5.1 (available at: lasig.epfl.ch/sambada)

Vekemans, X. Beauwens, B.S., Lemaire, M., Roldan-Ruiz, I. 2002. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homplasy and fragment size. Mol. Ecol. 11: 139–151.

Zhivotovsky, L.A. 1999. Estimating population structure in diploids with multilocus DNA markers. Mol. Ecol. 8: 907–913.

The research was carried out in the framework of the project 'Epigenetic vs. genetic diversity in natural plant populations: a case study of Croatian endemic *Salvia* species' financed by the Croatian Science Foundation.