GENETIC DIVERSITY OF WILD vs. CULTIVATED/NATURALIZED POPULATIONS OF DALMATIAN SAGE

Zlatko Šatović
University of Zagreb, Faculty of Agriculture, Zagreb, Croatia
Centre of Excellence for Biodiversity and Molecular Plant Breeding, Zagreb, Croatia
E-mail: zsatovic@agr.hr
(1) Dalmatian sage (*Salvia officinalis* L.)
(2) Genetic diversity and relationships
(3) Genetic structure
(4) Ecological niche modelling
(5) Demographic history
Plan

1. Dalmatian sage (*Salvia officinalis* L.)
2. Genetic diversity and relationships
3. Genetic structure
4. Ecological niche modelling
5. Demographic history
DALMATIAN SAGE

- *Salvia officinalis* L.
- outcrossing, insect-pollinated
- perennial subshrub
- widely used since ancient times for medicinal, culinary and ornamental purposes
- natural distribution:
  coastal region of the western Balkan and central and southern Apennine Peninsulas
- cultivation:
  throughout the Mediterranean region, in Australia, Germany, USA etc.
- naturalized populations:
  plants that have escaped from earlier cultivation
(1) Dalmatian sage (*Salvia officinalis* L.)
(2) Genetic diversity and relationships
(3) Genetic structure
(4) Ecological niche modelling
(5) Demographic history
(1) Collecting
- leaf tissue for DNA extraction
- 30 populations / 709 samples (20 to 25 per populations)
- origin:
  11 Balkan countries
- status:
  23 wild populations
  7 cultivated/naturalized populations

(2) DNA extraction

(3) Molecular marker analysis
- 8 microsatellite markers developed for Dalmatian sage
  (simple sequence repeats; SSRs)
- 165 alleles
COLLECTING

Populations:
- wild
- cultivated (naturalized)

Map showing the distribution of wild and cultivated populations in the region.
Komiža, island of Vis

Cavtat
(A) Isolation and characterization of microsatellites
- construction of microsatellite libraries from genomic DNA for GA, GT, AGA and ACA repeats
- primer pairs flanking microsatellite repeats were designed for 29 clones
- 29 microsatellites were tested using 25 plants

(B) Microsatellite marker analysis
- eight most polymorphic markers: 165 alleles
<table>
<thead>
<tr>
<th>Population</th>
<th>Individual</th>
<th>SoUZ001</th>
<th>SoUZ002</th>
<th>SoUZ002</th>
<th>...</th>
<th>SoUZ019</th>
<th>SoUZ019</th>
</tr>
</thead>
<tbody>
<tr>
<td>P01</td>
<td>S001</td>
<td>173</td>
<td>189</td>
<td>183</td>
<td>185</td>
<td>135</td>
<td>150</td>
</tr>
<tr>
<td>P01</td>
<td>S002</td>
<td>165</td>
<td>199</td>
<td>185</td>
<td>185</td>
<td>150</td>
<td>153</td>
</tr>
<tr>
<td>P01</td>
<td>S003</td>
<td>183</td>
<td>185</td>
<td>195</td>
<td>197</td>
<td>150</td>
<td>159</td>
</tr>
<tr>
<td>P01</td>
<td>S004</td>
<td>163</td>
<td>187</td>
<td>189</td>
<td>195</td>
<td>150</td>
<td>150</td>
</tr>
<tr>
<td>P01</td>
<td>S005</td>
<td>165</td>
<td>167</td>
<td>187</td>
<td>195</td>
<td>135</td>
<td>150</td>
</tr>
<tr>
<td>P01</td>
<td>S006</td>
<td>165</td>
<td>183</td>
<td>185</td>
<td>187</td>
<td>135</td>
<td>156</td>
</tr>
<tr>
<td>P01</td>
<td>S007</td>
<td>173</td>
<td>183</td>
<td>185</td>
<td>195</td>
<td>150</td>
<td>159</td>
</tr>
<tr>
<td>P01</td>
<td>S008</td>
<td>165</td>
<td>167</td>
<td>183</td>
<td>195</td>
<td>135</td>
<td>135</td>
</tr>
<tr>
<td>P01</td>
<td>S009</td>
<td>165</td>
<td>171</td>
<td>181</td>
<td>183</td>
<td>147</td>
<td>150</td>
</tr>
<tr>
<td>P01</td>
<td>S010</td>
<td>165</td>
<td>173</td>
<td>195</td>
<td>201</td>
<td>150</td>
<td>159</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P30</td>
<td>S709</td>
<td>161</td>
<td>165</td>
<td>185</td>
<td>195</td>
<td>135</td>
<td>150</td>
</tr>
</tbody>
</table>
GENETIC DIVERSITY

(1) Average no. of alleles per population ($N_{av}$)

(2) Allelic richness ($N_{ar}$)
- average no. of alleles per population independent of sample size

(3) No. of private alleles ($N_{pr}$)
- number of alleles detected in a single population
  (or in a group of populations: wild vs. cultivated)

<table>
<thead>
<tr>
<th></th>
<th>Wild</th>
<th>Cultivated</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{av}$</td>
<td>8.696</td>
<td>3.821</td>
</tr>
<tr>
<td>$N_{ar}$</td>
<td>7.920</td>
<td>3.672</td>
</tr>
<tr>
<td>Range</td>
<td>5.13-10.30</td>
<td>2.71-4.19</td>
</tr>
<tr>
<td>$P(N_{ar})$</td>
<td>$P &lt; 0.001$</td>
<td></td>
</tr>
<tr>
<td>$N_{pr}$ (total per population)</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>$N_{pr}$ (wild vs. cultivated)</td>
<td>115</td>
<td>0</td>
</tr>
</tbody>
</table>
Allelic richness:

- < 5
- 5 - 7
- 7 - 9
- > 9
- distance matrix: Cavalli-Sforza and Edwards' chord distance
- tree algorithm: Neighbour-joining
- unrooted tree

- seven cultivated/naturalized populations grouped separately from the rest and formed a well-supported clade (bootstrap support 99%) suggesting the common origin of the cultivated plant material
(1) Dalmatian sage (*Salvia officinalis* L.)

(2) Genetic diversity and relationships

(3) Genetic structure

(4) Ecological niche modelling

(5) Demographic history
Bayesian model-based clustering method for inferring population structure

Assumption:
- there are $K$ populations of origin each of which is characterized by a set of allele frequencies at each locus

Goal:
- assign individuals to populations of origin in such a way that within each population the departures from:
  (1) Hardy-Weinberg equilibrium (HWE) and
  (2) linkage equilibrium (LE) are minimized
PROCEDURE

(1) Estimation of the number of Ks populations of origin
   (1.1) Choose different values of K
   (1.2) Compute posterior probabilities for each K
   (1.3) Choose the most likely number of populations (K)

(2) Assignment
   (2.1) Assign individuals to populations (K)
   (2.2) For each individual, estimate the proportion of genome derived from the different population of origin
GENETIC STRUCTURE AT K = 2

- proportions of membership (Q) of each individual plant in each of the two clusters (population of origin):
  - each individual plant is represented by a single vertical line divided into colors representing different clusters
  - the length of the colored segment shows the individual's estimated proportion of membership in that cluster

Wild populations

Cultivated populations

- Wild populations
  - P26: Sićevo Gorge, Serbia

- Cultivated populations
- the cluster A (Wild populations) is at K = 3 split into two clusters according to geographical locations

A1 North-West: Adriatic populations

A2 South-East: Montenegro-Albania-Macedonia-Greece

B Cultivated populations
GENETIC STRUCTURE AT K = 4

- wild populations split further into three clusters
GENETIC STRUCTURE AT K = 4

$r = 0.47$

$P_{Mantel} = 0.0001$

$R^2 = 0.23$
PLAN

(1) Dalmatian sage (*Salvia officinalis* L.)
(2) Genetic diversity and relationships
(3) Genetic structure
(4) Ecological niche modelling
(5) Demographic history
ECOLOGICAL NICHE MODELLING

- species distribution modelling (SDMs)
  - estimate the relationship between species records at sites and the environmental characteristics of those sites
  - predict the suitability of sites for occupation or persistence of the species
  - produce a modelled distribution of the species (= identify species' suitable environmental space)

- input data:
  (1) geographic distribution
    68 data on occurrence (evenly distributed)
  (2) environmental characteristics of the sites
    19 bioclimatic variables (WorldClim database)
    - 11 temperature-related
    - 8 precipitation-related
    - representing the annual trends, seasonal variations and extremes in temperature and precipitation
ENM: PRESENT DAY CONDITIONS

Suitability
- 0.26 - 0.50
- 0.51 - 0.75
- 0.76 - 0.94

Apennine Peninsula
Adriatic Sea
Balkan Peninsula
PAST DISTRIBUTION

- to model potential species distribution during the Last Glacial Maximum (LGM; ~21,000 years BP)
- to identify putative glacial refugia of the species (= region which made possible the survival of the species and allowed a post-glacial re-colonization)
- input:
  (1) the present model
  (2) past environmental characteristics of the sites
      19 bioclimatic variables
      bioclimatic data for the LGM developed by Paleoclimate Modelling Intercomparison Project Phase II
- two models:
  CCSM (Community Climate System Model; USA)
  MIROC (Model for Interdisciplinary Research on Climate; Japan)
ENM: LAST GLACIAL MAXIMUM CCSM CONDITIONS

- CCSM: severe temperature decline

Suitability

- 0.26 - 0.50
- 0.51 - 0.75
- 0.76 - 0.94
Suitability

- MIROC: moderate temperature decline
(1) Dalmatian sage (*Salvia officinalis* L.)
(2) Genetic diversity and relationships
(3) Genetic structure
(4) Ecological niche modelling
(5) Demographic history
DEMOGRAPHIC HISTORY

- inference on demographic history of a species based on approximate Bayesian computation (ABC)

- **input:**
  - molecular data
  - scenarios describing demographic history

- **procedure:**
  - simulate (a large number of) datasets for each scenario
  - compare simulated and observed datasets
  - the most similar simulated dataset is the most likely

- **goal:**
  - compare competing scenarios
    - posterior probability of each scenario
  - estimate parameters for one or more scenarios
    - effective population size; time of divergence
POPULATIONS

Allelic richness:
- A1: 7.97
- A2: 8.49
- A3: 6.25

ABC analysis:
- A1: 92 inds from 7 pops
- A2: 91 inds from 10 pops
- A3: 82 inds from 4 pops
FIVE SIMPLE HISTORIC SCENARIOS

Scenario 1

Population A1 is derived from population A2, itself derived from population A3

Scenario 2

Population A3 is derived from population A2, itself derived from population A1
Scenario 3

Both populations A1 and A3 derived independently from population A2

Scenario 4

Population A2 was generated by admixture of populations A1 and A3
Scenario 5

All three populations diverged at the same time

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Posterior probability (PP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.166</td>
</tr>
<tr>
<td>2</td>
<td>0.193</td>
</tr>
<tr>
<td>3</td>
<td>0.075</td>
</tr>
<tr>
<td>4</td>
<td>0.126</td>
</tr>
<tr>
<td>5</td>
<td>0.440</td>
</tr>
</tbody>
</table>
HISTORICAL RECONSTRUCTION: ANCIENT PAST

Vegetation map of Europe during the last glacial period

Source: http://www.metatech.org/07/ice_age_global_warming.html

Distribution of Dalmatian sage before the last glaciation?

ABC Effective population sizes:

- Ancestral: 8,550 inds
- A1: 4,330 inds
- A2: 7,190 inds
- A3: 2,440 inds

Glacial refugia:

1. Iberian Peninsula
2. Apennine Peninsula
3. Balkan Peninsula
Distribution of Dalmatian sage during the Last Glacial Maximum

HISTORICAL RECONSTRUCTION: DIVERGENCE

Map of the 52 putative refugia within the Mediterranean region

ABC Time of divergence: 572 generations ago (95% CI: 157–1,540)

Generation time (life span): up to 300 years