

ARTICLE

# Ecogeography of haplotype composition of *Sagittaria trifolia* L. (Alismataceae): environment, space, vicariance and selective sweeps

Youhua Chen

Department of Renewable Resources, University of Alberta, Edmonton, Canada

**ABSTRACT** In the present report, the relative influence of environment and space was evaluated for explaining the variation of haplotype composition of 42 populations of *Sagittaria trifolia* L. (Alismataceae) in China. The results showed that, neither environment nor space could explain current haplotype composition patterns of *S. trifolia*, and most variation in haplotype composition could not be explained. Vicariance was recognized to explain the pattern that most haplotypes (25 out of 27 ones) were rare, being found in only one or two populations of *S. trifolia* in China. Finally, an emerging selective sweep from increasing human activity and habitat destruction explained the dominance patterns of haplotypes 5 and 8 among populations of *S. trifolia*.

Acta Biol Szeged 58(1):7-10 (2014)

**KEY WORDS**

environmental filtering,  
genetic drift,  
population genetic structure,  
vicariance,  
variation partitioning

Environmental filtering and dispersal limitation are known to influence the phylogeographic patterns of many species (Kelly et al. 2006; Viruel et al. 2012). Phylogeographic and population genetic structure of some herb species in the genus *Sagittaria* (family Alismataceae) have been quantified in recent studies in China (Chen et al. 2008; Tan et al. 2008; Liu et al. 2010). However, the role of environment and dispersal in generating the observed genetic structure has not been assessed thus far.

Among these species whose population genetic structure has been recently characterized is *S. trifolia*, a perennial herb that is widely distributed in Eurasian areas (Chen et al. 2008). This species is an ideal candidate to investigate the influence of historical contingency and contemporary environment on its distribution (Chen et al. 2008). In the present study, the effects of environment and space on haplotype composition and distribution of the species range was examined.

*S. trifolia* can disperse by several mechanisms including water and animals, especially bees and other insects (Cook 1990; Chen et al. 2008). The wide distribution of this species in Eurasia could be a result of its long-distance dispersal ability. I thus predicted that the variation in haplotype composition among different populations of *S. trifolia* in China would be closely related to the geographic distance between populations (field locations). The wide distributional range of *S. trifolia* over the territory of China, however, also includes many distinct climatic domains (Chen et al. 2006; Qin et al.

2006). The influence of environmental filtering (or environmental gradients) should thus also be critical in determining the haplotype composition pattern of *S. trifolia*. Given that both dispersal limitation and environmental filtering might function simultaneously in determining the geographic distribution of *S. trifolia*, I sought to quantify the relative importance of both mechanisms. I utilized a variation partitioning technique to quantitatively assess the relative contribution of space and environment on the population genetic structure of *S. trifolia* in China in terms of haplotype composition.

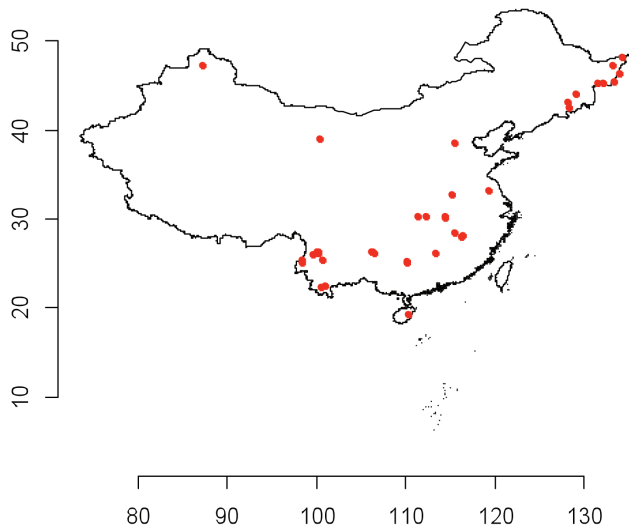
## Materials and Methods

### Distribution and haplotype information

Forty-two *S. trifolia* populations were considered in the present study that covered a wide range of the territory of China, representing almost all the natural distributional ranges for the species (Chen et al. 2008). Details on the field locations have been recorded in a previous study (Chen et al. 2008). In general, there are 11 northeast populations, 2 northwest populations, 1 north population, 5 east populations, 6 central populations, 3 south populations and 14 southwest populations (Chen et al. 2008). Figure 1 presented the geographic distribution of these populations over the territory of China. Haplotype information for the 42 *S. trifolia* populations was determined by using and comparing the atpB-rbcL noncoding intergenic spacer region of Chloroplast DNA of the 108 individuals over different populations (Chen et al. 2008). In total, 27 haplotypes were identified constituting a 42 × 27 population-haplotype matrix for analysis (Table 1).

Accepted June 11, 2014

\*Corresponding author. E-mail: haydi@126.com



**Figure 1.** Forty-two sampling locations for species *S. trifolia* across territory of China as presented in Fig. 1 of the previous study (Chen et al. 2008).

### Spatial and environmental variables

Latitude and longitude were the two most important spatial variables used to quantify the strength of dispersal in structuring haplotype composition of *S. trifolia* populations. However, because the raw spatial variables may not have been accurate in characterizing the structure of spatial autocorrelation within the haplotype data, I performed a transformation using principal coordinates of neighboring matrices (PCNM) (Dray et al. 2006; Legendre et al. 2009).

Environmental data characterizing the field locations were gathered from the IPCC (<http://www.ipcc-data.org/>) and Hydro 1k ([http://gcmd.nasa.gov/records/GCMD\\_HYDRO1k.html](http://gcmd.nasa.gov/records/GCMD_HYDRO1k.html)) databases. Fifteen variables were considered including diurnal temperature range, ground frost frequency, precipitation, solar radiation, temperature (minimum, mean and maximum), vapor pressure, wet day frequency, elevation, wetness index, aspect, slope, flow direction and flow accumulation. To avoid potential collinearity among the environmental variables in the variation partitioning, I performed a principal component analysis and extracted the first five principal axes as the representative of important environmental gradients influencing population genetic structure of *S. trifolia*.

### Variation partitioning technique

Variation partitioning has been widely applied in community ecology (Borcard et al. 1992; Legendre and Legendre 1998; Meot et al. 1998; Legendre 2007; Legendre et al. 2009). This technique can elucidate the relative contribution of different groups of explanatory variables on the variation in species composition (*i.e.*, beta diversity). The larger the proportion

**Table 1.** Geographic locations of 42 *S. trifolia* populations in China and associated haplotype presence/absence patterns from the previous study (Chen et al. 2008). Codes: 0: absence; 1: presence.

| Population ID | Longitude (E) | Latitude (N) | Haplotype composition          |
|---------------|---------------|--------------|--------------------------------|
| JL-1          | 128.13        | 43.19        | 00000001000000000100000000     |
| JL-2          | 128.15        | 43.23        | 00001001000000000000000000     |
| JL-3          | 128.28        | 42.57        | 00001000000000000000000000     |
| HLJ-1         | 131.51        | 45.3         | 0000000000000000000001000100   |
| HLJ-2         | 132.16        | 45.27        | 00000001000000000000000000     |
| HLJ-3         | 132.16        | 45.27        | 000000000000000000000010100000 |
| HLJ-4         | 129.11        | 44.06        | 00101001000000000000000000     |
| HLJ-5         | 134.01        | 46.44        | 00001001000000000000000000     |
| HLJ-6         | 134.24        | 48.2         | 00001000000000000000000000     |
| HLJ-7         | 133.18        | 47.25        | 00001010000000000000000000     |
| HLJ-8         | 133.35        | 45.47        | 000000000000101000000010000    |
| XJ-1          | 87.19         | 47.37        | 0000000000000000000000000001   |
| GS-1          | 100.27        | 39.03        | 00000001000000000000000000     |
| HEB-1         | 115.56        | 38.54        | 00000001000000000000000000     |
| YN-1          | 100.11        | 26.14        | 00101001000000000000000000     |
| YN-2          | 100.1         | 26.36        | 00000001000000000000000000     |
| YN-3          | 98.33         | 25.43        | 0000000100000000000100000000   |
| YN-4          | 100.15        | 26.34        | 00001001000000000000000000     |
| YN-5          | 100.56        | 22.4         | 00011000000000000000000000     |
| YN-6          | 98.33         | 25.07        | 00001001000001000000000000     |
| YN-7          | 101.01        | 22.45        | 110000000000000000000000010    |
| YN-8          | 99.56         | 26.06        | 00001001000000000000000000     |
| YN-9          | 100.6         | 25.42        | 00001000000000000000000000     |
| YN-10         | 100.1         | 26.36        | 00001001000000000000000000     |
| GZH-1         | 106.4         | 26.24        | 10001000000000000000000000     |
| GZH-2         | 106.2         | 26.25        | 00000001000000000000000000     |
| GZH-3         | 106.27        | 26.29        | 00001001000000000000000000     |
| GZH-4         | 106.16        | 26.25        | 00001001000000000000000000     |
| HUB-1         | 114.37        | 30.24        | 00001001000000000000000000     |
| HUB-2         | 114.41        | 30.2         | 00001001000000000000000000     |
| HUB-3         | 111.35        | 30.28        | 00001001000000000000000000     |
| HUB-4         | 112.22        | 30.31        | 00001000000100000000000000     |
| HUN-1         | 113.4         | 26.16        | 00001001000000000000000000     |
| HEN-1         | 115.24        | 32.7         | 00001000010000000000000000     |
| JS-1          | 119.29        | 33.24        | 00000000000000011000000000     |
| JS-2          | 119.29        | 33.19        | 00000001001000000000000000     |
| JX-1          | 116.37        | 28.14        | 0000100000000000000000001000   |
| JX-2          | 115.51        | 28.45        | 00001000000000000000000000     |
| JX-3          | 116.31        | 28.05        | 00000001000000000000000000     |
| HAIN-1        | 110.34        | 19.21        | 00001100100000000000000000     |
| GX-1          | 110.18        | 25.23        | 00001000000000000000000000     |
| GX-2          | 110.16        | 25.05        | 00001001000000000000000000     |

of variation explained by a group of explanatory variables is, the higher influence it has on the species composition.

A simplified version for the variation partitioning method was presented here (Borcard et al. 1992; Legendre and Legendre 1998; Chen 2013; Goncalves-Souza et al. 2014): Suppose that we have the covariate matrices  $E$  and  $S$  for environmental and spatial variables respectively (each column represents one factor).  $G$  is the combined matrix by merging both  $E$  and  $S$ .  $Y$  is the response matrix (here is the presence-absence matrix for the haplotypes of *S. trifolia*). Then, I

perform redundancy analysis (or multivariate regression analysis) on the covariate matrices to the response matrix as:  $Y_e = E(E^T E)^{-1} E Y$ ;  $Y_s = S(S^T S)^{-1} S Y$ ;  $Y_g = G(G^T G)^{-1} G Y$ . Defining the variation operator as  $COV(X) = \frac{1}{n} \text{Trace}((X - \bar{X})^T (X - \bar{X}))$ . Then, the variation explained solely by the environmental covariates  $E$  is given by  $COV(Y_e) - COV(Y_s)$ ; the variation explained solely by the spatial covariates  $S$  is given by  $COV(Y_s) - COV(Y_e)$ ; and the variation jointly explained by both environmental and spatial covariates is given by  $COV(Y_e) + COV(Y_s) - COV(Y_g)$ . At last, the unexplained variation is  $COV(Y) - COV(Y_g)$ .

Variance partitioning was applied at the population level for *S. trifolia*. I used composition (or distribution) of the 27 haplotypes over the 42 *S. trifolia* populations and evaluated the relative influence of environmental and spatial descriptors on influencing the haplotype composition among populations. The program “vegan” (Oksanen et al. 2012) in the R environment (R Development Core Team 2013) was used to perform variation partitioning with the function “varpart”. The multivariate method, canonical correspondence analysis (CCA) (ter Braak 1986), was implemented when performing variation partitioning.

## Results

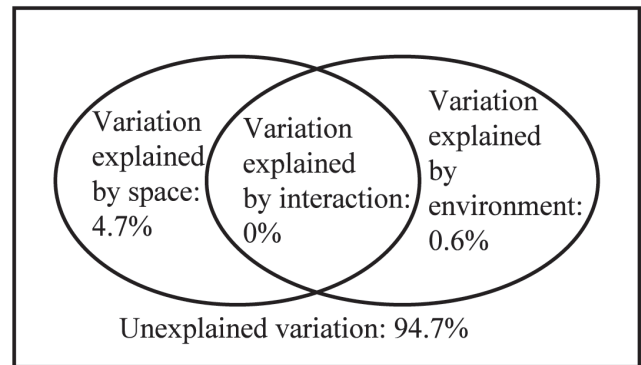
Interestingly, neither space nor environment could strongly explain the haplotype composition of *S. trifolia* populations (Fig. 2). Space explained only 4.7% of total variation, while environment explained essentially none (0.6%) of the total variation of the haplotype composition of *S. trifolia* populations. The interaction between space and environment also did not explain any (0%) of the variation in haplotype composition. Most variation (94.7%) in haplotype composition could not be explained using the selected variables.

Given the fact that there are too much variation unexplained based on the results, the possible explanations on the relative importance of spatial and environmental filtering on structuring haplotype distribution of *S. trifolia* distributed over China are challenging. However, because the composition of haplotypes among the natural populations of *S. trifolia* presented interesting patterns (*i.e.*, some haplotypes dominate across different populations) (Chen et al. 2008), some discussions and conclusions still can be draw as in the discussion section.

## Discussion

Spatial and environmental effects did not play a substantial role in influencing the population genetic structure of *S. trifolia* in China in terms of haplotype composition. Thus, my hypothesis that space and environment structure haplotype composition of *S. trifolia* populations was not supported.

My results suggest that haplotype divergence of *S. trifolia* may not be related to environmental filtering and may be only slightly related to spatial limitation. However, it should be acknowledged that the present study suffers certain limitations



**Figure 2.** Result of variation partitioning explaining haplotype composition of *S. trifolia* populations in China. The term “interaction” inside the intersection of the two ellipses indicates the interaction of space and environment.

that affect this interpretation. Firstly, the 47 populations of *S. trifolia* included in the study were all within the territory of China. We might have found evidence of environmental filtering or space if populations in Russia or Europe were also included (Chen et al. 2008). Secondly, haplotype information examined in this study derived from only one DNA marker (a noncoding region of chloroplast +DNA), which may be not sufficient to reflect the true haplotype condition of the species over different populations. The combination of haplotype information from different microsatellite loci (Hosid et al. 2010; Viruel et al. 2012) is necessary to systematically and holistically reveal the haplotype composition of the species and uncover the emerging roles of environment and space on influencing haplotype diversity and composition of the species.

A different mechanism that may be more important than space and environment in determining the haplotype composition of *S. trifolia* populations is vicariance (*i.e.*, geographic isolation) (Estep et al. 2005; Viruel et al. 2012). There is indication in the data that vicariance could be an important driver of *S. trifolia* composition in that most haplotypes were rare, being found in only one or two of the 42 populations examined. Indeed, out of a total 27 haplotypes found in the samples, 23 haplotypes were found in just a single population and 2 haplotypes were present in just 2 populations (Table 1). This high divergence of haplotypes between populations could be a result of dispersal barriers, habitat discreteness or isolated areas (Kelly et al. 2006). A high vicariance level among populations leads to higher divergence when comparing different populations (Estep et al. 2005; Kelly et al. 2006; Viruel et al. 2012). It is noted, however, that limited sampling may also explain why the occurrence of haplotypes was so rare between populations.

Interestingly, it was observed that haplotypes 5 and 8 were highly predominant among the different populations. It is possible that recent dispersal, genetic drift or a selective sweep

(Depaulis et al. 1999; Depaulis et al. 2000; Messer and Neher 2012) is emerging among isolated populations, leading to the increasing dominance of haplotypes 5 and 8 in China. Mutations in specific nucleotide sites for haplotypes 5 and 8 may provide a survival advantage and increased fitness. A selective sweep resulting in the dominance of haplotypes 5 and 8 might be also driven by human activity and habitat destruction, as high-speed economic development and environmental change in China (Chen et al. 2006; Qin et al. 2006) have increased anthropogenic impacts on the native habitats of *Sagittaria* (Chen et al. 2008; Tan et al. 2008; Liu et al. 2010).

In conclusion, to better elucidate haplotype composition of *S. trifolia*, more field sampling and DNA sequencing are advised. As discussed above, the haplotype information should be as comprehensive as possible if ones want to better quantify the relative influence of neutrality and niche mechanisms in structuring the spatial distribution patterns of *S. trifolia* haplotypes.

## Acknowledgements

I thank Gennifer Meldrum for her great help on improving the language aspect of the work. I appreciate an anonymous reviewer for providing constructive comments on helping improve an earlier version of the work. The study is supported by China Scholarship Council.

## References

- Borcard D, Legendre P, Drapeau P (1992) Partialling out the Spatial Component of Ecological Variation. *Ecology* 73:1045.
- Ter Braak CJF (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67:1167-1179.
- Chen J, Liu F, Wang Q, Motley T (2008) Phylogeography of a marsh herb *Sagittaria trifolia* (Alismataceae) in China inferred from cpDNA atpB-rbcL intergenic spacers. *Mol Phylogenet Evol* 48:168-175.
- Chen Y (2013) A comparison of synonymous codon usage bias patterns in DNA and RNA virus genomes: quantifying the relative importance of mutational pressure and natural selection. *Biomed Res Int* 2013:406342.
- Chen Y, Ding Y, She Z, et al. (2006) Assessment of climate and environment changes in China (II): Impacts, adaptation and mitigation of climate and environment changes. *Adv Clim Chang Res* 2:6-12.
- Cook C (1990) Aquatic plant book. SPB Academic Publishing, The Hague, The Netherlands.
- Depaulis F, Brazier L, Mousset S, Turbe A, Veuille M (2000) Selective sweep near the In(2L)t inversion breakpoint in an African population of *Drosophila melanogaster*. *Genet Res* 76:149-158.
- Depaulis F, Brazier L, Veuille M (1999) Selective sweeps at the *Drosophila melanogaster* suppressor of hairless locus and its association with the In(2L)t inversion polymorphism. *Genetics* 152:1017-1024.
- Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecol Modell* 196:483-493.
- Estep MC, Connell MU, Henson RN, Murell ZE, Small RL (2005) Testing a vicariance model to explain haplotype distribution in the psammophilic scorpion *Paruroctonus utahensis*. *Southwest Nat* 50:150-157.
- Goncalves-Souza T, Diniz-Filho J, Romero G (2014) Disentangling the phylogenetic and ecological components of spider phenotypic variation. *PLoS One* 9:e89314.
- Hosid E, Yusim E, Grishkan I, Frenkel ZM, Wasswer SP, Nevo E, Korol A (2010) Microsatellite diversity in natural populations of ascomycetous fungus, *Emerizella nidulans*, from different climatic-edaphic conditions in Israel. *Isr J Ecol Evol* 56:119-134.
- Kelly D, Maclsaac H, Heath D (2006) Vicariance and dispersal effects on phylogeographic structure and speciation in a widespread estuarine invertebrate. *Evolution (NY)* 60:257-267.
- Legendre P (2007) Studying beta diversity: ecological variation partitioning by multiple regression and canonical analysis. *J Plant Ecol* 1:3-8.
- Legendre P, Legendre L (1998) Numerical ecology. Elsevier Science BV, Amsterdam
- Legendre P, Mi X, Ren H, Ma K, Yu M, Sun I-F, He F (2009) Partitioning beta diversity in a subtropical broad-leaved forest of China. *Ecology* 90:663-674.
- Liu F, Zhao S-Y, Li W, Chen J-M, Wang Q-F (2010) Population genetic structure and phylogeographic patterns in Chinese endemic species *Sagittaria lichuanensis*, inferred from cpDNA atpB-rbcL intergenic spacers. *Botany* 88:886-892.
- Meot A, Legendre P, Borcard D (1998) Partialling out the spatial component of ecological variation: questions and propositions in the linear modelling framework. *Environ Ecol Stat* 5:1-27.
- Messer P, Neher R (2012) Estimating the strength of selective sweeps from deep population diversity data. *Genetics* 191:593-605.
- Oksanen J, Blanchet G, Kindt R, et al. (2012) vegan: Community Ecology Package. R package version 2.0-4.
- Qin D, Ding Y, Su J, et al. (2006) Assessment of climate and environment changes in China (I): Climate and environment changes in China and their projections. *Adv Clim Chang Res* 2:1-5.
- R Development Core Team (2013) R: A Language and Environment for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Tan B, Liu K, Yue X-L, Liu F, Chen J-M, Wang Q-F (2008) Chloroplast DNA variation and phylogeographic patterns in the Chinese endemic marsh herb *Sagittaria potamogetifolia*. *Aquat Bot* 89:372-378.
- Viruel J, Catalan P, Segarra-Moragues J (2012) Disrupted phylogeographical microsatellite and chloroplast DNA patterns indicate a vicariance rather than long-distance dispersal origin for the disjunct distribution of the Chilean endemic *Dioscorea biloba* (Dioscoreaceae) around the Atacama Desert. *J Biogeogr* 39:1073-1085.