PSAwinD Version 1.1.1: A Program for Calculating Spatial Indices

M. TAKAHASHI

From the Laboratory for Environmental Amelioration Breeding Division, Breeding Department, Forest Tree Breeding Center, 3809-1 Ishi, Juo, Ibaraki, 319-1301 Japan.

Address correspondence to Makoto Takahashi at the address above, or e-mail: makotot@affrc.go.jp.

The investigation of within-population spatial genetic structure has been an important aspect of population genetics, allowing determination of the major genetic processes that have established the genetic structure within a population by examining relevant population genetic parameters. I have developed a user-friendly Windows software program, the Program for Spatial Autocorrelation for Windows, Delphi
version 1.1.1 (PSAwinD), for examining within-population spatial genetic structure. The features of the software are described herein. The software can calculate Moran’s I, standard normal deviate (SND), and the number of alleles in common (NAC) with various optional settings. Those indices deal with different hierarchical levels of genetic data. Different levels of investigation provide different kinds of information, and comprehensive analysis gives us a greater chance of obtaining a deeper insight into the genetic processes operating on the studied population. In addition, a demographic perspective is important in spatial genetic studies because different life-history groups exhibit different genetic structure and variation. Furthermore, the software also provides functions for calculating spatial indices for combinations of data subsets.

The investigation of within-population spatial genetic structure has been an important aspect of population genetics since Sokal and Oden (1978a,b) introduced the method of spatial autocorrelation analysis into population genetic studies. This technique allows determination of the major genetic processes that have established the genetic structure within a population by examining relevant population genetic parameters (Berg and Hamrick 1995; Sokal et al. 1997). Early spatial genetic studies were primarily based on allozyme data (Alvares-Buylla et al. 1996; Berg and Hamrick 1995; Epperson and Allard 1989; Knowles et al. 1992; Takahashi et al. 2000). Subsequently, investigations using data obtained from DNA markers such as random amplified polymorphic DNA (RAPD) and ISSR (Tani et al. 1998), and microsatellites (Streiff et al. 1998; Ueno et al. 2000) also became common.

At first, the indices introduced by Sokal and Oden (1978a), such as standard normal deviates (SND) and Moran’s I, were frequently used (Epperson and Allard 1989; Knowles et al. 1992). Other indices, such as the number of alleles in common (NAC) and the coefficient of coancestry, were also devised and became popular (Berg and Hamrick 1995; Hamrick et al. 1993). Spatial genetic studies deal with the genetic data of individuals within a population in conjunction with their location data. Populations are comprised of individuals from different life-history groups, and the genetic variation and structure of these groups differ (Alvarez-Buylla et al. 1996; Epperson and Alvarez-Buylla 1997; Hamrick et al. 1993). Thus the demographic perspective is important when examining within-population spatial genetic structure.

Genetic data have three hierarchical levels: allelic, genotypic, and multigenotypic, which makes thorough examination of datasets laborious. Examinations usually include calculating spatial indices for some sets of distance classes, and software designed to perform such calculations is valuable as it facilitates spatial genetic analysis. Computer software such as SAAP (Wartenberg 1989), FijAnal (Nason 1997), ECO-GENE (Degen et al. 1996), and its later version, SGS (Degen et al. 2001), provide functions for calculating spatial indices. The software described here, the Program for Spatial Autocorrelation for Windows, Delphi version 1.1.1 (PSAwinD) can calculate three spatial indices: Moran’s I (Sokal and Oden 1978a), SND (Sokal and Oden 1978a), and NAC (Berg and Hamrick 1995; Surle et al. 1990).

**Program Description**

**Data Files**

PSAwinD requires at least two data files to calculate the spatial indices: a DAT file and a LOC file. The DAT file should include genetic data of the genotyped individuals and the LOC file should include the location data for each individual, in the required format. Users can examine spatial autocorrelation within and/or between the data subsets when additional data files are defined (see below).

**Distance Classes**

The software can determine three types of distance classes: sequential, cumulative, and a Gabriel network. Sequential distance classes most frequently appear in spatial genetic studies. However, the other two types of distance classes are also often important to the dissection of the data. The statistical power of the spatial indices is influenced by the W value, which is twice the sum of the join numbers included in the individual distance classes. When distance classes have a constant interval, the W values are different among the distance classes, resulting in weaker statistical power, with larger variances for both smaller and larger distance classes. To obtain distance classes with equalized statistical power of the indices, the W values of every distance class must be repeatedly calculated until suitable ranges for each class are determined, that is, until every class has an almost equal value. The software provides a function to obtain optimal distance classes automatically.

**Output**

PSAwinD can calculate Moran’s I, SND, and NAC with various optional settings. Those indices deal with different hierarchical levels of genetic data. SND values are calculated using genotypic data, while Moran’s I values are calculated using allelic data. NAC can be calculated by using both single-locus and multilocus genotypic data. Plant species are sessile, with diploid status (a set of genotypes), so examining the spatial pattern of genotypes is useful to reveal the spatial genetic structure of the studied population. Conversely, with regeneration, genetically varied individuals are dispersed by pollen and seed with haploid status (a set of alleles, but not a set of genotypes). Therefore, consideration at the allelic level is also important. Furthermore, multigenotypic examination often provides important information about clonality and the cumulative effect of genetic clustering of an individual locus. Thus different levels of investigation provide different kinds of information and comprehensive analysis gives us a greater chance of obtaining a deeper insight into the genetic processes operating on the studied population.

After calculating the indices, PSAwinD gives the results
as comma-delimited text files with the extensions RT1 and RT2 for Moran’s I, RPS and SND for the SND, and NAC for the NAC data. The RT1 and RPS files include the \( W \) values of all the distance classes at the investigated loci. The RT2 and SND files include the observed and expected values of the indices, their variances, and their significance, with allele (or genotype) frequencies and heterozygosities. The NAC file includes single-locus and multilocus NAC values and their grand mean, differentials from the grand mean, and variances and their significance.

The software can create summary files, SR2 and SSD, from files RT2 and SND, respectively. These files present the number and the proportion of significant alleles (or genotypes) at three levels of significance (the 5%, 1%, and 0.1% significance levels). The summary files facilitate identification of the general trends of the results.

### Defining Layers

In spatial genetic studies, demographic perspective is important because different life-history groups exhibit different genetic structures and variations (Alvarez-Buylla et al. 1996; Epperson and Alvarez-Buylla 1997). Consequently spatial genetic structure within each life-history group and/or genetic associations among different life-history groups, and their comparisons, are important in understanding the spatial distribution of genetic variation within a population comprised of a complex mixture of various life-history groups (Epperson and Alvarez-Buylla 1997; Hamrick et al. 1993). The software can calculate the spatial indices for various combinations of data subsets (layers), including or excluding certain combinations of layers from the calculation, when the subsets are defined in LZR files. These are (1) within identical layers, (2) between different layers, and (3) between specific combinations of layers. For example, when three layers, Large (L), Medium (M), and Small (S), are defined, the first procedure includes only joins between identical layers in the calculations, while joins between different layers are excluded. The second procedure includes only joins between different layers in the calculations. The last procedure is valuable as it includes joins only between specific combinations of layers. In this case, it is possible to calculate the SND and Moran’s I for various combinations of layers. In the example above, selecting “L-M” would include joins between the L layer and the M layer. Inputting “L–not L” would include joins between the L layer and the M layer and joins between the L layer and the S layer in the calculation. Thus the software gives researchers an opportunity to examine the spatial genetic association within and/or between various combinations of layers, using Moran’s I and SND. This version of PSAwinD does not support the calculation of NAC with defined layers.

### Availability

PSAwinD version 1.1.1 is written in Delphi version 5.0 (Borland Inc.) and has been compiled as a 32-bit version for the Microsoft Windows (Windows 95/98/00/Mc and Windows NT) operating system. The program, sample data files, and user’s manual are available on my Internet homepage: http://homepage3.nifty.com/makotot_frbc/.

### References


Received December 10, 2002
Accepted January 27, 2003
Corresponding Editor: Sudhir Kumar
DOI: 10.1093/jhered/esg058