Molecular Ecology Notes (2006) **6**, 965–967 doi: 10.1111/j.1471-8286.2006.01414.x

PROGRAM NOTE

PERM: a computer program to detect structuring factors in social units

PIERRE DUCHESNE, CÉDRIC ÉTIENNE and LOUIS BERNATCHEZ
Département de Biologie, Université Laval, Ste-Foy, Québec, Canada G1K 7P4

Abstract

PERM is a permutation program designed to detect statistical connections between grouping structures and grouping factors or correlates. Groups may be of various kinds such as herds, flocks, schools and mating couples provided they make up meaningful social units. Relatedness, population membership and genotypic contents are among several aggregating variables which may be processed. Typically, PERM takes in a collection of grouped data and outputs a *P* value. The latter is computed on the basis of random membership among groups (H_O). All files, including input, output and program, are of Excel type (.xls). PERM can be downloaded free of charge at: www.bio.ulaval.ca/louisbernatchez/downloads.htm.

Keywords: computer software, grouping variables, *P* value, permutation, randomization, social units

Received 20 December 2005; revision accepted 31 March 2006

Living organisms are frequently grouped together in more or less cohesive units such as schools of fish, bird flocks, herds of mammals, mating couples, etc. The organizing principles of these groups may vary according to species and context, and there may be several factors at work within a single organization, for instance kinship groups within an otherwise randomly mating population. One way to detect statistical connections between a candidate grouping variable such as relatedness and the grouping structure found in a collection of groups is to evaluate the *P* value of a meaningful statistic S (e.g. the intragroup sum of Rxy values) against its probability distribution under the hypothesis that grouping structure is independent of the candidate variable.

Classical tests such as *t*-tests or *F*-tests are not always applicable for this purpose since they usually assume statistical properties such as symmetry, homocedasticity and groups of equal or nearly equal sizes. By contrast, groups found in natural settings are often of very unequal sizes and the probability distributions or densities of S are usually unknown. Randomization tests are a way around this problem. The latter are defined as procedures that involve comparing an observed test statistic with a distribution that is generated by randomly reordering the data values in some sense (Manly 2001). Randomization tests are increasingly being used in various fields of biology, including single species ecology (Suzuki *et al*. 2005), mate choice (Landry *et al*. 2001), population genetics (Belkhir *et al*. 2004), including kinships aggregation (Belkhir *et al*. 2002) but most often, community ecology (Manly 2001; Goheen *et al*. 2005).

The goal of PERM is to statistically detect grouping factors or correlates without imposing any prior conditions. PERM tackles this problem by randomly permuting specimens across groups within a collection of groups while preserving all original data as well as group sizes. Hence, the empirical grouping structure is destroyed and replaced by random group membership. After each permutation, S is computed and compared to S_O, the value of S computed from the observed collection. The proportion of the permutation *S* values that are equal or higher than S_O is one *P* value estimate. Each *P* value computed from many permutations is an estimate. The user may ask for several iterations to be run, each iteration producing a distinct S distribution and thus a distinct *P* value estimate. The number of permutations and the number of iterations are user-defined parameters. Based on the above permutation procedure, PERM 1.0 adresses four distinct problems each with a specific procedure. Below, we describe these four procedures in more details.

Correspondence: P. Duchesne, Fax: +1418-656-2043; E-mail: Pierre.Duchesne@bio.ulaval.ca

Here are some examples of the types of questions which PERM was designed to answer. The names of the associated procedures are in brackets.

1 Do specimens from the same group tend to be more related/inbred than specimens from distinct groups? (Groups: pairwise relationship)
2 Are there differences between groups in terms of size, weight, isotope contents, etc.? (Groups: comparisons of means)
3 Do specimens from the same school/herd/flock tend to belong to the same population? (Groups: category membership)
4 Is there a statistical connection between mating couple formation and MHC alleles? (Pairs/couples: pairwise relationship)

The first procedure, the Groups: pairwise relationship test, aims at detecting a possible statistical link between a collection of groups and some pairwise relationship statistic Sxy such as Rxy, Mxy, half or fullsibship. Relations other than those pertaining to relatedness may also be tested. This procedure takes in an Excel document containing a matrix of the Sxy statistic between each pair of specimens found in the collection, irrespective of group membership. This matrix may be square, inferior triangular or superior triangular. Diagonal values are not used. A distinct input Excel document describes the observed grouping structure. The *P* value computations are based on the SOM statistic: the sum of all Sxy values for all pairs of specimens that belong to the same group, i.e. the intragroup sum of Sxy (see Fig. 1). PERM outputs an Excel sheet comprising several sheets of information among which a global *P* value and a matrix of *P* values with rows corresponding to groups and columns to iterations. Therefore groups contributing to lower the global *P* value may be identified. Also, the stability of the *P* value estimates over iterations may be assessed. *P* value instability can be decreased by increasing the number of permutations per iteration.

The second procedure, the Groups: comparisons of means test, compares average values among groups for some measure. The groups of values are rebuilt through random permutations. After each permutation, the averages of the g groups, M1 … Mg, are recalculated and compared. When the collection comprises only two groups, the user may test for M1 > M2, M2 > M1, M1 <> M2 based on statistics D = M1–M2, D = M2–M1 or D = |M1 – M2|, respectively. With more than two groups, all paired bilateral tests are performed. Number of permutations and iterations are under user control.

Third, the Groups: category membership procedure searches for connections between group structure and category, e.g. population membership (e.g. Fraser *et al.* 2005). It is based on a homogeneity statistics H. Given hG, the number of specimens belonging to the most frequently identified category within group G, H is the sum of hG computed over all groups. For example, given the following six allocation groups:
A B B B A, B A B A A A, A A, B B B B A B B B B A, A A A B B A A B B B, A B B
H = 3 + 4 + 2 + 8 + 5 + 2 = 24

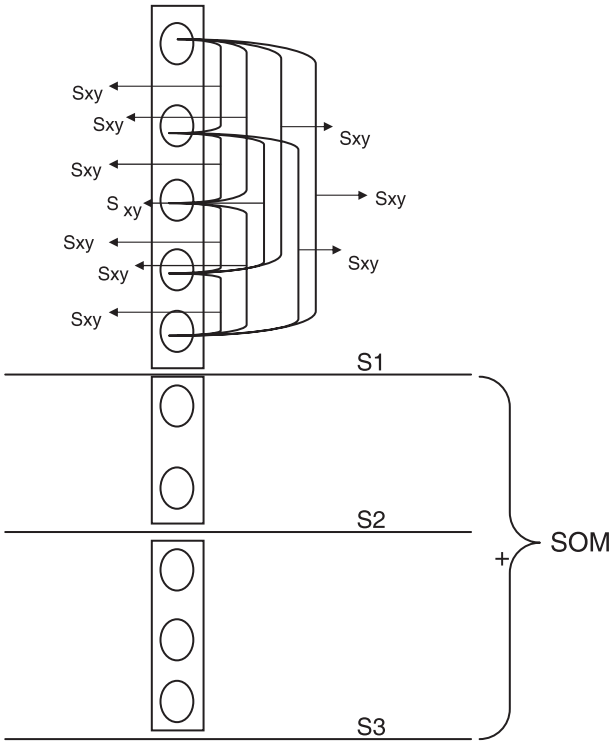

Fig. 1 The SOM statistics for the Groups: pairwise relationship test. All Sxy values are added within each group of the collection. SOM is obtained by summing over all group sums.

The number of permutations and iterations are under user control. PERM outputs a matrix (iteration X group) of *P* values in addition to a global *P* value.

In order to assess the power of a given category membership test and also to interpret the *P* value from empirical groups, PERM allows to compute *P* values from artificial groups built on the basis of perfect homogeneity within each group. The user must provide estimates of category sizes, absolute or relative, such as census sizes of populations. Also an allocation matrix must be given with each column containing the probabilities that, say, A be correctly identified as A or incorrectly as B or C, etc. Very unequal census sizes, e.g. when nearly all specimens belong to category A, as well as high probabilities of incorrect allocations reduce the power of the category test.

The fourth and last procedure, the Pairs/couples: pairwise relationship test, aims at detecting a statistical connection between mating couple formation and some

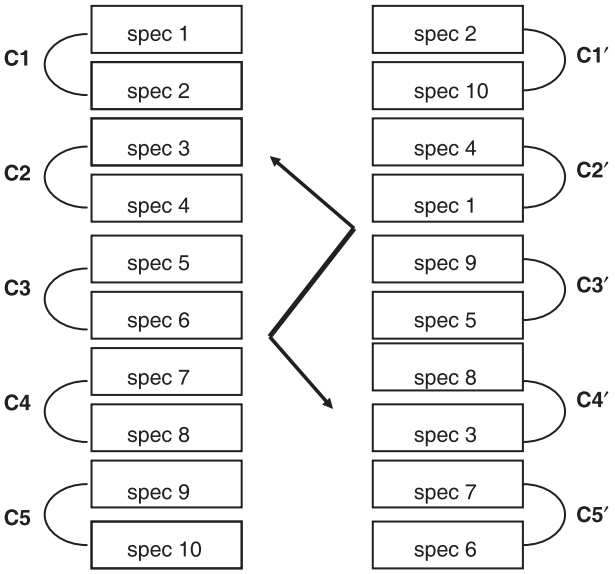

Fig. 2 The nonsexed permutation procedure for the Pairs/couples: pairwise relationship test. The C1 … C5 symbols refer to the original, empirical, couples. After all specimens have been randomly permuted, new couples C1′ … C5′ are built by pairing the first specimen with the second, the third with the fourth, etc.

pairwise relationship between partners such as allelic distance on an MHC locus (e.g. Landry *et al*. 2001). It is analogous to the Groups: pairwise relationship procedure except that it offers a choice of two permutation types according to whether the specimens are sexed or not. In sexed permutations, males are permuted among themselves and females are also permuted among themselves thus keeping constant the sets of male and female specimens. However, if the sex of the specimens is unknown, all genotypes are put together in the same set before each permutation (see Fig. 2). Even when all specimens have been sexed, if the user is convinced that sex and Sxy are statistically independent variables, then the nonsexed permutation type may be selected to increase the number of possible pairings, and hence the power of the test.

Within all four procedures, PERM outputs a histogram of S (the procedure statistic) for the last distribution/iteration, i.e. including the value of S computed at each permutation. The part of the histogram corresponding to the last *P* value is highlighted. Besides illustrating the meaning of the *P* value, the purpose of drawing a graph of the last distribution is to visualize its shape and smoothness level. A jigsaw graph is an indication that the number of permutations should probably be increased especially when associated with *P* value instability over iterations.

Each of the procedures found in PERM has also been coded in the mathematical programming language Maple 9.50. Outputs from PERM were compared with those from the corresponding Maple procedures on several sets of data. Any discrepancy between outputs was thoroughly investigated until the underlying bug was found and corrected. Given the considerable distance between the programming logics of Maple and VBA and the fact that the two codes were written by two different people, there is little doubt that PERM is now free of consequential bugs.

In addition to the Excel sheet containing the VBA code of the program, the user will find on the PERM website a folder of 'Demo Files' and a 'Readme' file providing the basic operating information and detailed description of each procedure.

PERM has been designed with a view to tackle other problems involving the detection of statistical connections between grouping structure and candidate aggregating variables. Future additions are possible and will depend largely on subsequent requests. Researchers are invited to express specific calculation needs around the general topic of group formation.

References

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, logiciel sous Windows™ pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier, France.

Belkhir K, Castric V, Bonhomme F (2002) IDENTIX, a computer program to test for relatedness in a population using permutation methods. *Molecular Ecology Notes*, **2**, 611–614.

Fraser D, Duchesne P, Bernatchez L (2005) Migratory brook charr schools exhibit population and kin associations beyond juvenile stages. *Molecular Ecology*, **14**, 3133–3146.

Goheen JR, White EP, Ernest SKM *et al*. (2005) Intra-guild compensation regulates species richness in desert rodents. *Ecology*, **86**, 567–573.

Landry C, Garant D, Duchesne P, Bernatchez L (2001) 'Good genes as heterozygosity': MHC and mate choice in Atlantic salmon (*Salmo salar*). *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **1473**, 1279–1285.

Manly FJ (2001). *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 2nd edn. Texts in Statistical Science.

Maple 9.50 Copyright © 1981–2004 by Maplesoft, a division of Waterloo Maple Inc.

Suzuki RO, Suzuki JI, Kachi N (2005) Change in spatial distribution patterns of a biennial plant between growth stages and generations in a patchy habitat. *Annals of Botany*, **96**, 1009–1017.