

# Statistical analyses of population genetic data: new tools, old concepts

François Rousset and Michel Raymond

Up to a few years ago, most genetic data analyses were conducted using *Biosys*, a package that was already 'showing its age' at the time of final release, in the words of one of its authors<sup>1</sup>. Today, it is possible to use the so-called 'exact' tests – that is, tests not affected by an imperfect knowledge of some parameter values – where this was practically impossible before, as well as various resampling methods. The recent emphasis on hypervariable DNA markers<sup>2,3</sup>, where the number of alleles is large relative to the sample size, revives some classical statistical problems and justifies the current interest in exact tests in situations

where traditional asymptotic procedures can be misleading<sup>4-6</sup>. More-refined asymptotic procedures, such as some implementations of the bootstrap, may also be constructed.

The user is not only confronted with a choice of various resampling techniques or exact tests versus more traditional methods, but also with a choice of different test statistics (to define a rejection zone) and different algorithms. One of the most commonly encountered problem is testing independence in two-way contingency tables. Although this is a limited approach to data analysis, two-way contingency tables have always been an excellent introduction to important statistical concepts. Standard uses include testing for genetic differentiation between populations or linkage disequilibrium<sup>6-8</sup>, or for random species associations in ecological studies<sup>9</sup>.

To define a test, it is necessary to define a probability distribution under some 'null' hypothesis, and to compute probabilities given this distribution. A common problem in the definition of tests is that probability distributions under the 'null' hypothesis may depend on unknown parameters. Consider two subpopulations, from which two samples of size  $n_1$  and  $n_2$  are analyzed at one locus presenting two alleles  $a$  and  $A$ . For example,  $n_{1a}$  genes of allelic type  $a$  are sampled in population 1, and a total of  $n_{.a} = n_{1a} + n_{2a}$  genes of allelic type  $a$  are sampled among a total of  $n_{..} = n_1 + n_2$  genes. The probability of some sample  $(n_{1a}, n_{1A}, n_{2a}, n_{2A})$ , under the null hypothesis of no subpopulation differentiation, is:

$$\frac{n_1!n_2!}{n_{1a}!n_{2a}!n_{1A}!n_{2A}!} (1-p)^{n_{.a}} p^{n_{.A}} \quad (1)$$

(where  $n_i = n_{i1} + n_{i2}$ ) and depends on the unknown frequency  $p$  of allele  $A$  in the population.

It is well known that whatever the frequency  $p$ , the distribution of the chi-square statistic can be approximated by the chi-square distribution and this approximation improves as sample size increases. This illustrates the first traditional answer to this problem, where the probability computed is

**Several methods or tests and various software are currently being developed for analyzing data in population genetics and ecology, which often rely on computer-intensive algorithms. The potential user is thus confronted with the painful experience of freedom and, in particular, has to make *a priori* choices between different methods. Using examples drawn from population genetics, we explain some of these recently developed tools.**

François Rousset and Michel Raymond are at the Laboratoire Génétique et Environnement, Institut des Sciences de l'Évolution, Université de Montpellier II, 34095 Montpellier, France.

an asymptotic approximation to the true value that depends on  $p$ . The other solution, more rarely applicable, is to find a probability distribution that is independent of the unknown parameter  $p$ . In the present example, although the probability distribution of samples is dependent upon the allelic frequency, the conditional probability among all possible samples with the same observed total ('marginal') allelic counts  $n_{.a}$  and  $n_{.A}$  (as in Box 1) is independent of this frequency. This conditional probability is

$$\frac{n_1!n_2!n_{.a}!n_{.A}!}{n_{1a}!n_{1A}!n_{2a}!n_{2A}!} \quad (2)$$

A test based on such a distribution is 'exact' in the sense that exact probability values can be computed, although  $p$  is unknown. Exactness of tests does not imply that they 'tell the truth' about some hypothesis, since the outcome of the analysis can only be a statement about probabilities of events. Examples of exact tests in population genetics are listed in Table 1.

Persistent doubts are expressed about the use of the probability distribution defined by eqn 2, and methods based on the estimation of unknown parameter values are still proposed particularly in the ecological literature<sup>10,11</sup>. The logic underlying the definition of exact tests shows that (1) the choice of a distribution where marginal counts are fixed

**Table 1. Some exact tests available in population genetics**

Type of test <sup>a</sup>	Probability distribution		Statistics	Refs
	Distribution of	conditional on		
Hardy-Weinberg	genotypic counts	allelic counts	various	5,19
Population structure	allelic counts	allelic counts	various	7,8,28
	genotypic counts	genotypic counts	various	28,42
Linkage disequilibrium	two-loci	genotypic	Pr(S) <sup>b</sup>	43,44
	genotypic counts	counts for each locus		
Ewens-Watterson	allelic counts	allelic counts	Pr(S)	6,45
	allelic counts	number of alleles	$\Sigma p_i^2$ , Pr(S)	46,47

<sup>a</sup>Each test is characterized by a probability distribution generated under the null hypothesis, and a test statistic used to define the rejection zone and to compute the  $P$ -value. For each probability distribution, there are as many exact tests as possible statistics, and only some of them are indicated. Computer programs for most of these tests are available as described in the references given.

<sup>b</sup>Pr(S) is the probability of the sample under the null hypothesis.

<sup>c</sup> $\Sigma p_i^2$  is the sample homozygosity.

**Table 2. Test statistics and ranking**

Genotypes <sup>a</sup>					Pr <sup>c</sup>	Ranks <sup>b</sup>			
AA	AB	BB	AC	CD		Pr	$\hat{\theta}$	$\hat{\theta}_{RH}$	$G_a$
<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>0</b>	$\frac{3}{35}$	2	2	2	3
<b>0</b>	<b>0</b>	<b>0</b>	<b>2</b>	<b>1</b>	$\frac{1}{35}$	1	3	4	4
1	1	1	0	1	$\frac{3}{35}$	2	9	12	10
0	0	0	3	0	$\frac{3}{35}$	2	4	7	7
1	1	0	1	1	$\frac{3}{35}$	2	2	3	3
0	0	1	2	0	$\frac{1}{35}$	1	1	1	2
1	0	0	2	1	$\frac{3}{35}$	2	11	14	11
0	1	1	1	0	$\frac{3}{35}$	2	10	13	10
1	0	0	3	0	$\frac{3}{35}$	2	2	5	5
0	1	1	0	1	$\frac{3}{35}$	2	6	6	6
1	0	1	2	0	$\frac{3}{35}$	2	8	10	9
0	1	0	2	1	$\frac{1}{35}$	1	7	8	8
1	0	1	0	1	$\frac{3}{35}$	2	5	9	8
0	0	1	2	1	$\frac{1}{35}$	1	9	11	9
1	1	0	1	0	$\frac{1}{35}$	1	1	1	1
0	0	1	3	0	$\frac{1}{35}$	1	1	1	1
1	1	0	0	1	$\frac{1}{35}$	1	1	1	1
0	0	0	3	1	$\frac{1}{35}$	1	1	1	1
1	1	1	0	0	$\frac{1}{35}$	1	1	1	1

<sup>a</sup>The table describes genotypic data from two populations. For the five distinct genotypes indicated, all possible contingency tables corresponding to the row marginal genotypic values of 4 and 3 and the column marginal values of 1, 1, 1, 3, and 1 are depicted.

<sup>b</sup>Pr can be used as a statistic, and used to rank all tables. Other statistics can be computed and used to define other rankings. Some examples of possible statistics are given here: two estimators of  $F_{ST}$  ( $\hat{\theta}$  and  $\hat{\theta}_{RH}$ , defined in Ref. 41), and  $G_a$ , the log likelihood ratio computed from the allelic counts<sup>42</sup>. The  $P$ -value is defined as the sum of probabilities (Pr) of tables with an equal or lower rank than the observed table. As an illustration, if the observed sample is the one in bold characters, then the exact  $P$ -value is 1 (probability test), 0.314 ( $\hat{\theta}$  test), 0.143 ( $\hat{\theta}_{RH}$  test) or 0.229 ( $G_a$  test).

<sup>c</sup>Pr is the conditional probability of each table under the null hypothesis.

must be based on explicit probability arguments, (2) it is not necessary for the marginal counts to be fixed in advance by the experimental set-up<sup>12</sup>, and (3) it is not necessary to estimate the unknown frequencies. These points are important to keep in mind because they are also the basic justification of permutation procedures: permutation techniques are relevant when the distribution generated by permutation is independent of the unknown parameters under the 'null' hypothesis.

As an illustration, we can consider Mantel tests, which are used increasingly in population studies<sup>13-15</sup>. Suppose that the 'null' hypothesis is that genetic differentiation follows an island model (Box 2), with each subsample assumed to be sampled from a different subpopulation. For any given parameter values (i.e. values of  $4Nmp$ ,  $4Nm q$  in Box 2), the probability distribution of each subsample<sup>16</sup> is independent

of the observed subsamples in other subpopulations so that the probability of observing subsamples  $s_1$  and  $s_2$  in geographic positions  $a$  and  $b$ , respectively, is the same as that of the reverse observation under the 'null' hypothesis. It is then possible to test for a relationship between genetic data and geographic distance by permutation of the subsamples between the different geographic locations. Such a test will be 'exact'. By contrast, no conclusion can be drawn from random permutations of pairwise statistics (such as pairwise  $F_{ST}$  estimates) between the different pairs of locations (cf. Box 3).

Some extensions of Mantel tests have been proposed such as partial Mantel tests, where the 'null' hypothesis corresponds to a correlation of some pairwise parameter with one parameter and aims to test for a partial correlation with another one<sup>14,15,17</sup>, but the validity of the permutation procedures proposed in such cases remains to be formally investigated.

**The choice of the statistic**

A statistic is computed from the data, and is used to rank all possible samples of the distribution. The  $P$ -value of the test is then computed by summing up the probabilities of all samples of one side of the ranking, starting from (and including) the value of the observed sample (Table 2). Thus, different test statistics define different rankings of possible samples, but the  $P$ -value is in all cases defined as a sum of exact probabilities of samples with more extreme ranks, so all tests are exact tests. The probability of the observed sample is used as a test statistic in many situations (thus defining an exact probability test), but this choice, which traces back to Fisher's exact test for contingency tables<sup>18</sup>, has no general justification.

How do we choose the test statistics? A test statistic is generally chosen to maximize the power of the test when some specified alternative hypothesis is true. It is therefore pivotal to have an idea of what could be a (biologically) plausible alternative hypothesis in order to design a powerful test. When specific alternative hypotheses have been formulated and probability distributions under the alternative hypothesis are known, it is possible to define more efficient test statistics and more efficient estimators.

For example, phenomena such as selfing and other types of inbreeding lead to a specific form of heterozygote deficiency which is often the appropriate alternative hypothesis to design tests of random mating<sup>19</sup>. Sometimes it is less easy. For example, a sampling distribution is known for the island model, and has been used to define an estimator of  $Nm$  that is efficient for this model<sup>20</sup>, but there is no equivalent result for other possible models of population structure.

Likewise, departure from neutrality could be the result of selection for particular genotypes, which leads to various predictions concerning the allele frequency distribution or the genotypic constitution. In most cases, the kind of selection acting on the locus studied is not known *a priori*; thus the formulation of a biologically sound alternative hypothesis of a test of neutrality is difficult. This is illustrated by the low power of the Ewens-Watterson test (which uses the sample homozygosity as a test statistic) with respect to specific types of selection<sup>21</sup>.

**Computing or estimating P-values**

To compute the  $P$ -value of an exact test, all possible samples of the probability distribution must be considered and ranked according to a particular test statistic (see details in Table 2). This complete enumeration is not always possible, either because algorithms generating this exhaustive

sampling are not presently available or because computation time is limiting due to the large number of possible configurations to be considered. As an illustration, the number of possible genotypic tables corresponding to a sample of 40 individuals with 4 equiproportional alleles is 6671, and is higher than  $10^7$  if the sample size is increased fourfold. One solution is to estimate this  $P$ -value by considering only a 'random' subset of possible samples. This subset is generated using two possible sampling techniques: permutation or Markov Chain Monte Carlo (MCMC) algorithms. For both methods, the probability of sampling a particular element is equal to its probability of occurrence under the 'null' hypothesis. The principle of permutation is to generate independent random samples under this hypothesis. The principle of MCMC is explained in Box 1. The main weakness of MCMC is that the speed of convergence to the correct value, though usually much faster than that of permutation algorithms for many applications<sup>5,9</sup>, cannot be predicted in advance<sup>22</sup>. This issue is a current area of investigation<sup>22</sup>.

**The bootstrap: an alternative?**

The principle underlying the bootstrap is that if we do not know the true distribution of the difference between an unknown parameter  $\theta$  and some estimator of this parameter, we can approximate it by the distribution of the difference between an estimate  $\hat{\theta}$  obtained from a particular data set and estimates computed from different imaginary data sets generated using information provided by the sample. Instructive examples are given in the context of mark-recapture experiments in Ref. 23 and in the context of phylogenetic analysis in Ref. 24.

There are two ways to generate such imaginary data sets. In the nonparametric bootstrap, they are obtained by sampling with replacement within the real data set. For example, the distribution of the difference between an estimator of  $F_{ST}$  and the true value of  $F_{ST}$  could be approximated by the distribution of the difference between an estimate  $\hat{F}$  obtained from several loci and further estimates computed from different resamples of loci from this data set. The parametric bootstrap is similar except that one generates imaginary data sets according to a probability distribution whose parameters are estimated from the real data set. For example, the data may be used to estimate the parameters  $p$  and  $Nm$  of the distribution of subpopulation gene frequency  $x$  in an island model (Box 2). Then imaginary subpopulations can be generated by sampling from the distribution with estimated parameter values  $\hat{p}$  and  $\hat{Nm}$ , and imaginary data sets can be generated by sampling from such imaginary populations. In both cases it is possible to define confidence intervals for the parameter values (see Ref. 25 for a recent review and discussion of the different methods).

Though superficially similar to permutation methods, bootstrap methods are based on large sample approximations and are not an alternative answer to problems that can be solved by exact methods. Rather, they can address other problems. Like other asymptotic techniques, they may be the only ones available for most questions of interest. However, they can be misleading and may need laborious studies to be validated. There are as yet few investigations of possible uses of the bootstrap for population genetic analyses<sup>6,26,27</sup> and it has not always been found accurate<sup>6</sup>. However, these studies did not consider improved uses of the bootstrap such as those discussed in Ref. 25.

*Applying the bootstrap to population structure*

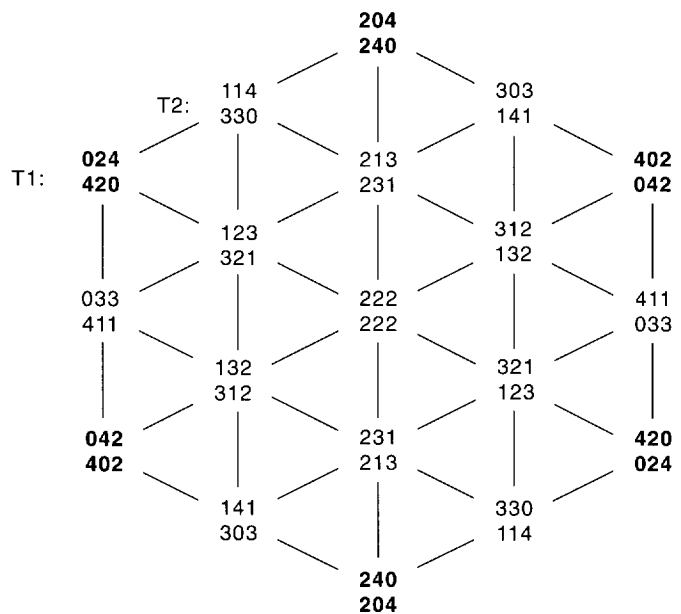
How can we define bootstrap confidence intervals for the value of  $F_{ST}$ ? Consider the example of an island model with

**Box 1. Principle of Markov Chain Monte Carlo algorithms**

Consider all possible contingency tables, shown below, with margins

		6	
		6	
4	4	4	

The principle of MCMC is to define a 'random walk' between these different states, with the probabilities of transition from one state to another chosen so that the probability of being in some state is that expected under the hypothesis considered. This is done using the Metropolis-Hastings algorithm<sup>34</sup> where transition probabilities between any two states are a function of the ratio of probabilities of being in each of them. For example, the probability of tables T1 and T2, according to eqn (2), are  $\Pr(T1) = 6!6!4!4!4! / 12!2!4!2!4! = 0.006464$  and  $\Pr(T2) = 6!6!4!4!4! / 12!4!3!3! = 0.01732$ , but the transition probability from T1 to T2 is a function of  $\Pr(T2)/\Pr(T1)$  which is simply 8/3.



In addition, it is possible to prevent transitions in both ways between pairs of states, as long as there is still some path leading from any state to any other. Hence for simplicity one allows only transitions between states  $i$  and  $j$  such that the ratio of probabilities  $\Pr(i)/\Pr(j)$  remains simple to compute, such as the transitions between connected states below. This and other simplifications are strong advantages of MCMC algorithms and explain their application in various contexts where permutation techniques are less easily implemented or even irrelevant<sup>5,35,36</sup>, notably for maximum likelihood estimation.

For the probability test, the  $P$ -value is the sum of probabilities of all matrices (in bold) that have equal or lower probability than the observed one (say T1). It is estimated by the fraction of time the Markov chain is in one of them. The precision of this estimate increases with the number of iterations of the Markov chain.

migration rate  $m$  between subpopulations, and two possible alleles  $A$  and  $a$ . The probability distribution of the allele frequency would follow Wright's distribution (see Box 2). In this model,  $F_{ST} = 1/(1 + 4Nm)$ . Now suppose that  $n$  individuals are sampled in each of two subpopulations, and we wish to estimate  $Nm$ . What will happen if we resample individuals?  $F_{ST}$  is estimated from sample allele frequencies, and the larger the sample size  $n$ , the closer we will be to some value, which will

**Box 2. The island model and the definition of  $F_{ST}$**

In the island model, migrants into a subpopulation are supposed to come with identical probabilities  $1/n$  from any other of a number  $n$  of subpopulations. If  $n$  is large and the fraction of migrants is  $m \ll 1$ , and considering only two alleles for simplicity, the probability distribution of the gene frequency  $x$  in a particular subpopulation is approximately  $f(x) \approx Cx^{4Nm(p-1)}(1-x)^{4Nm(q-1)}$ , where  $C$  is a constant,  $p$  and  $q = 1 - p$  are the gene frequencies in the total collection of subpopulations, and  $N$  is the subpopulation size. Thus the probability of identity  $Q_{within}$  of a pair of genes sampled within a subpopulation is the expectation of  $x^2 + (1-x)^2$ :

$$Q_{within} \approx \int_0^1 [x^2 + (1-x)^2] f(x) dx = \frac{1 + 4Nm(p^2 + q^2)}{1 + 4Nm}$$

and the probability of identity of a pair of genes sampled between subpopulations is the sum of squares of expected allele frequencies,  $Q_{among} = p^2 + q^2$ . Hence  $F_{ST}$  defined as  $(Q_{within} - Q_{among}) / (1 - Q_{among})$  has value  $1 / (1 + 4Nm)$ . This is no more than Wright's result<sup>37</sup>, formulated in terms of probabilities of identity in state for pairs of genes, rather than identity by descent or 'standardized variance of gene frequencies'. With this definition, it is easier to grasp basic properties of both the parameter  $F_{ST}$  and its estimators in more complex models<sup>38-40</sup>.

In the present context  $F_{ST}$  is not defined as a function of the allele frequencies in some subpopulations, say  $x_1$  and  $x_2$  in two subpopulations under study. These frequencies as well as probabilities of identities for these particular subpopulations – the so called 'actual values',  $q_{within i} = x_i^2 + (1-x_i)^2$  and  $q_{among} = x_1 x_2 + (1-x_1)(1-x_2)$  – are random variables that will differ for each pair of subpopulations. By contrast,  $F_{ST}$  is a parameter, that is, it is a property of a probability distribution rather than of a particular realization of the process considered.

be determined by the frequencies of the *A* allele in the two subpopulations. But this value will not be  $1 / (1 + 4Nm)$ . This is readily seen by the fact that if we had taken another pair of subpopulations, the frequency of *A* in each of them would be different from those in the previous pair, so when the sample size increases, the  $F_{ST}$  estimates would become closer to another value: we do not get closer to the parameter value, but to a random value different for each pair of subpopulations. Consequently, a test based on a confidence interval constructed in this way should almost always reject the correct parameter value for large sample sizes.

The appropriate question is whether the value of the variable considered in some population (not its estimate on a sample) is a random variable the distribution of which is determined by the parameter value, or a fixed value identical for different populations exposed to the same process. This is the question to consider in order to choose between fixed and random models in statistics. In models where genetic drift plays a role, quantities such as the subpopulation frequency  $x$  (Box 2) are random variables (sometimes called 'actual' values). Since  $F_{ST}$  is mainly used to measure the effects of genetic drift relative to other evolutionary forces, resampling individuals will generally not be appropriate for inferences about  $F_{ST}$  values<sup>28</sup>. Resampling loci may not be a perfect solution<sup>8,29</sup>. Because typical population genetics studies rarely consider more than twenty loci, studies are needed to check that bootstrap confidence intervals generated by resampling loci are satisfying. But confidence intervals obtained in this way can be free of the problem of convergence to an incorrect parameter value.

**Persistent problems**

As new statistical techniques are proposed, the need to understand traditional concepts of probability and statistics is stronger. We have illustrated this need in some of the simplest problems encountered in population genetics. The recent development of efficient algorithms for exact testing and the large access to computer power provide a new input for the analysis of population data. More efficient tests and estimators can save much effort, but the more general problem of constructing confidence intervals for estimates of

**Box 3. How not to do permutations**

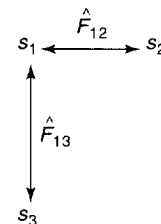
Consider a (semi-)matrix of  $F_{ST}$  estimates  $\hat{F}_{ij}$  computed from all possible pairs of samples  $i$  and  $j$  among four samples:

$\hat{F}_{12}$		
$\hat{F}_{13}$	$\hat{F}_{23}$	
$\hat{F}_{14}$	$\hat{F}_{24}$	$\hat{F}_{34}$

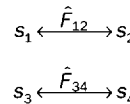
and another matrix obtained by permutation of two cells:

$\hat{F}_{12}$		
$\hat{F}_{34}$	$\hat{F}_{23}$	
$\hat{F}_{14}$	$\hat{F}_{24}$	$\hat{F}_{13}$

A test based on a permutation procedure that generates such permuted data sets could be used if the observed and permuted configurations had the same probability under the null hypothesis. But in the first table the first two elements of the first column are  $\hat{F}_{12}$  and  $\hat{F}_{13}$ , and in the second they are  $\hat{F}_{12}$  and  $\hat{F}_{34}$ . The joint distribution of  $\hat{F}_{12}$  and  $\hat{F}_{13}$  is entirely determined by the joint distribution of subsamples  $s_1, s_2$  and  $s_3$ ,



while in the second table the joint distribution of  $\hat{F}_{12}$  and  $\hat{F}_{34}$  is entirely determined by the joint distribution of subsamples  $s_1, s_2, s_3$  and  $s_4$ :



It will be identical to the joint distribution of  $s_1, s_2$  and  $s_3$  above only when  $s_1 = s_4$ . The probability distribution of a pair of estimates  $(\hat{F}_{12}, \hat{F}_{34})$ , each defined from different subsamples, is not the same as that of  $(\hat{F}_{12}, \hat{F}_{13})$  that involves the same subsample  $s_1$ . Generalizing the argument to the whole matrix, we see that it fails to satisfy the requirements of permutation tests. No conclusion can be drawn from considering all permutations of elements of a matrix of  $F_{ST}$  estimates or any other statistic on pairs of populations. In contrast, the Mantel test is based on consideration of only a restricted subset of permutations, obtained by permuting subsamples between geographic positions rather than pairwise statistics between cells.

standard population genetic parameters has not yet any answer.

Traditional parameters such as  $F_{ST}$  may be used to detect differences in levels of differentiation of populations exposed to different evolutionary processes, for example if a group of populations is expected to display a higher  $F_{ST}$  than another group, or they may be used as measures that allow

the estimation of demographic parameters such as neighborhood sizes. In the first case, one would prefer a statistic whose distributions under the two situations we wish to distinguish do not overlap, such as an estimator with low variance. In the second case an estimator with low bias may be preferred.

However, in both cases we need realistic models to assess the power of different statistics or the efficiency of different estimators. Many studies have argued that there are discrepancies between direct and indirect estimates of gene flow parameters<sup>30–33</sup>, implying that the models generally considered are inappropriate, or inappropriately interpreted, or that present statistical analyses are inefficient. Whatever the reason(s), a better understanding of real populations will be necessary to show that many of the currently developed computer-intensive methods improve the efficiency of statistical analyses.

### Acknowledgements

We thank F. Cézilly, C. Chevillon, M. Lascoux, N. Pasteur and J. Ronfort for helpful advice and criticisms of the manuscript. This work was partly supported by the Centre de Systématique de Montpellier. This is contribution ISEM 97.058 of the Institut des Sciences de l'Évolution (CNRS, UMR 5554).

### References

- Swofford, D.L. and Selander, R.B. (1989) *BIOSYS-1. A Computer Program for the Analysis of Allelic Variation in Population Genetics and Biochemical Systematics. Release 1.7.*, Illinois Natural History Survey
- Avise, J.C. (1994) *Molecular Markers, Natural History, and Evolution*, Chapman & Hall
- Jarne, P. and Lagoda, J.L. (1996) **Microsatellites, from molecules to populations and back**, *Trends Ecol. Evol.* 11, 424–429
- Roff, D.A. and Bentzen, P. (1989) **The statistical analysis of mitochondrial DNA polymorphisms:  $\chi^2$  and the problem of small samples**, *Mol. Biol. Evol.* 6, 539–545
- Guo, S.W. and Thompson, E.A. (1992) **Performing the exact test of Hardy-Weinberg proportion for multiple alleles**, *Biometrics* 48, 361–372
- Slatkin, M. and Excoffier, L. (1996) **Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm**, *Heredity* 76, 377–383
- Hudson, R.R., Boos, D.D. and Kaplan, N.L. (1992) **A statistical test for detecting geographic subdivision**, *Mol. Biol. Evol.* 9, 138–151
- Raymond, M. and Rousset, F. (1995) **An exact test for population differentiation**, *Evolution* 49, 1283–1286
- Manly, B.F.J. (1995) **A note on the analysis of species co-occurrences**, *Ecology* 76, 1109–1115
- Jackson, D.A., Somers, K.M. and Harvey, H.H. (1992) **Null models and fish communities: evidence of nonrandom patterns**, *Am. Nat.* 139, 930–951
- Lafferty, K.D., Sammond, D.T. and Kuris, A.M. (1994) **Analysis of larval trematode communities**, *Ecology* 75, 2275–2285
- Tocher, K.D. (1950) **Extension of the Neyman-Pearson theory of tests to discontinuous variates**, *Biometrika* 37, 130–144
- Smouse, P.E., Long, J.C. and Sokal, R.R. (1986) **Multiple regression and correlation extensions of the Mantel test of matrix correspondence**, *Syst. Zool.* 35, 627–632
- Leduc, A. *et al.* (1992) **Study of spatial components of forest cover using partial Mantel tests and path analysis**, *J. Veg. Sci.* 3, 69–78
- Thorpe, R.S. *et al.* (1994) **Testing ecological and phylogenetic hypotheses in microevolutionary studies**, in *Phylogenetics and Ecology* (Eggleton, P. and Vane-Wright, R., eds), pp. 189–206, Academic Press
- Rannala, B. and Hartigan, J.A. (1995) **Identity by descent in island-mainland populations**, *Genetics* 139, 429–437
- Legendre, P., Lapointe, F.-J. and Casgrain, P. (1994) **Modeling brain evolution from behavior: a permutational regression approach**, *Evolution* 48, 1487–1499
- Fisher, R.A. (1935) **The logic of inductive inference**, *J. R. Stat. Soc.* 98, 39–54
- Rousset, F. and Raymond, M. (1995) **Testing heterozygote excess and deficiency**, *Genetics* 140, 1413–1419
- Rannala, B. and Hartigan, J.A. (1996) **Estimating gene flow in island populations**, *Genet. Res.* 67, 147–158
- Spencer, H.G. and Marks, R.W. (1992) **The maintenance of single-locus polymorphism. IV. Models with mutation from existing alleles**, *Genetics* 130, 211–221
- Cowles, M.K. and Carlin, B.P. (1996) **Markov chain Monte Carlo convergence diagnostics: a comparative review**, *J. Am. Stat. Soc.* 91, 883–904
- Potvin, C. and Roff, D.A. (1993) **Distribution-free and robust statistical methods: viable alternatives to parametric statistics?** *Ecology* 74, 1617–1628
- Efron, B., Halloran, E. and Holmes, S. (1996) **Bootstrap confidence levels for phylogenetic trees**, *Proc. Natl. Acad. Sci. U. S. A.* 93, 7085–7090
- DiCiccio, T.J. and Efron, B. (1996) **Bootstrap confidence intervals (with discussion)**, *Stat. Sci.* 11, 189–228
- Van Dongen, S. and Backeljau, T. (1995) **One- and two-sample tests for single locus inbreeding coefficients using the bootstrap**, *Heredity* 74, 129–135
- Long, J.C., Williams, R.C. and Urbanek, M. (1995) **An E-M algorithm and testing strategy for multiple-locus haplotypes**, *Am. J. Hum. Genet.* 56, 799–810
- Weir, B.S. (1996) *Genetic Data Analysis II*, Sinauer
- Van Dongen, S. (1995) **How should we bootstrap allozyme data?** *Heredity* 74, 445–447
- Boileau, M.G., Hebert, P.D.N. and Schwartz, S.S. (1992) **Non-equilibrium gene frequency divergence: persistent founder effects in natural populations**, *J. Evol. Biol.* 5, 25–39
- Campbell, D.R. and Dooley, J.L. (1992) **The spatial scale of genetic differentiation in a hummingbird-pollinated plant: comparison with models of isolation by distance**, *Am. Nat.* 139, 735–748
- Schilthuizen, M. and Lombaerts, M. (1994) **Population structure and levels of gene flow in the mediterranean land snail *Albinaria corrugata* (Pulmonata: Clausiliidae)**, *Evolution* 48, 577–586
- Johnson, M.S. and Black, R. (1995) **Neighborhood size and the importance of barriers to gene flow in an intertidal snail**, *Heredity* 75, 142–154
- Hastings, W.K. (1970) **Monte Carlo sampling methods using Markov chains and their applications**, *Biometrika* 57, 97–109
- Griffiths, R.C. and Tavaré, S. (1994) **Simulating probability distributions in the coalescent**, *Theor. Popul. Biol.* 46, 131–159
- Kuhner, M.K., Yamato, J. and Felsenstein, J. (1995) **Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling**, *Genetics* 140, 1421–1430
- Wright, S. (1969) *Evolution and the Genetics of Populations. II. The Theory of Gene Frequencies*, University of Chicago Press
- Cockerham, C.C. and Weir, B.S. (1987) **Correlations, descent measures: drift with migration and mutation**, *Proc. Natl. Acad. Sci. U. S. A.* 84, 8512–8514
- Cockerham, C.C. and Weir, B.S. (1993) **Estimation of gene flow from *F*-statistics**, *Evolution* 47, 855–863
- Rousset, F. (1996) **Equilibrium values of measures of population subdivision for stepwise mutation processes**, *Genetics* 142, 1357–1362
- Weir, B.S. and Cockerham, C.C. (1984) **Estimating *F*-statistics for the analysis of population structure**, *Evolution* 38, 1358–1370
- Goudet, J. *et al.* (1996) **Testing differentiation in diploid populations**, *Genetics* 144, 1931–1938
- Zaikin, D., Zhivotovsky, L. and Weir, B.S. (1995) **Exact tests for association between alleles at arbitrary numbers of loci**, *Genetica* 96, 169–178
- Raymond, M. and Rousset, F. (1995) **GENEPOP Version 1.2: population genetics software for exact tests and ecumenicism**, *J. Hered.* 86, 248–249
- Slatkin, M. (1994) **Linkage disequilibrium in growing and stable populations**, *Genetics* 137, 331–336
- Watterson, G.A. (1977) **Heterosis or neutrality?** *Genetics* 85, 789–814
- Slatkin, M. (1996) **A correction to the exact test based on the Ewens sampling distribution**, *Genet. Res.* 68, 259–260