

# Appendix 1 – Methods and Statistics in GenAEx 6.4

By Rod Peakall and Peter Smouse

## Overview

This appendix summarizes the methods and statistics available in GenAEx 6.4, along with supporting references. With regard to the references, we point (where possible) to one or more currently available texts in population genetics, although many procedures provided in GenAEx are not yet covered in such texts. The texts we have found particularly useful for teaching include: *Principles of Population Genetics* (3<sup>rd</sup> Ed) by Hartl and Clark [1], *Genetics of Populations* by Hedrick [2-3] and *Introduction to Conservation Genetics* by Frankham et al. [4] and *Conservation and the Genetics of Populations* by Allendorf and Gordon [5]. A series of very useful primers include: *A Primer of Population Genetics* (3<sup>rd</sup> Ed) by Hartl [6], *A Primer of Ecological Genetics* by Conner and Hartl [7] and, *A Primer of Conservation Genetics* by Frankham et al. [8]. Invaluable statistical population genetic resources for advanced users include *Genetic Data Analysis* by Weir [9] and *The Evaluation of Forensic DNA Evidence* by the National Research Council [10] and *Forensic DNA Evidence Interpretation* [11].

## GenAEx Tutorials – An Overview of Topics

Coinciding with the release of GenAEx 6.3, a series of self-paced tutorials on population genetic analysis that employ hand calculations and exercises within GenAEx are now available from the GenAEx web site. These tutorials are based on material we have prepared to support our national graduate workshop series *Genetic Analysis for Populations Studies* offered by Rod Peakall (The Australian National University) and Peter Smouse (Rutgers University, USA) at the Australian National University every 2 to 3 years since 1996. It is strongly recommended that this appendix is read in conjunction with these tutorials.

Part 1 - *An Introduction to Frequency-Based Population Genetic Analysis*: scoring genetic markers, Allele Frequency, Heterozygosity, F-statistics, Nei Genetic Distance, Shannon Diversity Indices and Chi-square tests for Hardy-Weinberg Equilibrium.

Part 2 - *Genetic Distance and Analysis of Molecular Variance (AMOVA)*: Haploid, Codominant and Binary Genetic Distance, AMOVA and F-statistics.

Part 3 - *Spatial Genetic Analysis*: Principal Coordinate Analysis (PCA), Mantel Tests for Matrix Correspondence and Spatial Autocorrelation Analysis.

Part 4 - *Advanced Frequency-Based Analysis*: DNA Profile Probability, Probability of Identity, Probability of Exclusion, Population Assignment and Pairwise Relatedness.

Part 5 - *Advanced Features Including Data Import and Export*: Working with DNA sequences, importing and processing raw genotypic data, and exporting data from GenAEx to other software packages. The Stats menu and how to customise the GenAEx menu are also covered briefly.

Part 6 - *TwoGener*: Male gametic inference, male gametic distances, gametic AMOVA.

**Table 1: An overview of methods in GenAlEx 6.4**

Methods	Overview	GenAlEx Option
Assignment	See Population Assignment and Sex Biased Dispersal and Tutorial Part 4.	
AMOVA	The Analysis of Molecular Variance (AMOVA) procedure follows the methods of Excoffier et al. [12], Huff et al. [13], Peakall et al. [14], and Michalakis and Excoffier [15]. AMOVA is an important, relatively recent statistical procedure that allows the hierarchical partitioning of genetic variation among populations and regions and the estimation of the widely used $F$ -statistics and/or their analogues. The AMOVA framework is important, because it allows such analysis for many types of genetic marker, and it offers statistical testing by random permutation. The data type and choice of distance calculation input into AMOVA can lead to related, but different, analyses. To estimate $F_{st}$ values for codominant data, choose the distance option <i>Codom-Allelic</i> . To estimate $R_{st}$ values, choose the <i>Codom-Microsat</i> option. To suppress within population variance and simply calculate population differentiation based on the genotypic variance, choose the <i>Codom-Genotypic</i> distance option. This later option produces an estimate of $\Phi_{pt}$ , an analogue of $F_{st}$ . $\Phi_{pt}$ is also the estimate of population genetic differentiation provided by GenAlEx when binary or haploid data are analysed. When comparing patterns of molecular variance between codominant (such as SSRs) and non-codominant markers (such as binary AFLPs), $\Phi_{pt}$ should be used for all markers (see [14] and [16] for comparative studies). The software package Arlequin [17] also offers AMOVA and provides other statistical analyses not available in GenAlEx. GenAlEx offers data export options to Arlequin. See also Genetic Distance below and in Table 2 and Tutorial Part 2.	AMOVA
AMOVA and Statistical Tests	Note that there are differences between GenAlEx and Arlequin in how permutational tests are performed and reported for AMOVA. Typically in biology, the null hypothesis concerns the condition of ‘ <i>No Difference</i> ’. In the case of AMOVA: $H_0 =$ No genetic difference among the populations ( $\Phi_{pt} = 0$ , or $F_{ST} = 0$ or $R_{ST} = 0$ ), $H_1 =$ Genetic difference among the populations ( $\Phi_{pt} > 0$ , or $F_{ST} > 0$ or $R_{ST} > 0$ ). Thus, for AMOVA in GenAlEx, our null hypothesis ( $H_0$ ) is that subpopulations can be considered part of a single large random mating genetic population. If true, any subpopulation groups we define are arbitrary and merely represent a sample from the same gene pool. Thus, we should find little difference (other than minor sampling effects) between the arbitrary subpopulations. It follows that if we shuffle (randomize) the samples in our data set, and calculate AMOVA for each shuffle, we should get values close to that expected by chance in a randomly sampled population. Because of sampling effects, the results will naturally vary from shuffle to shuffle. On the other hand, if we perform multiple shuffles (say 99 or 999 times) we can obtain a good estimate of the value we would expect if the null hypothesis were true. This is the rationale for statistical testing by random permutation that is used in an AMOVA analysis within GenAlEx. To determine if the observed value is significantly greater than that expected by chance, we simply compare our observed value against the outcomes of the permutations. If our observed value is greater than the permuted values 95% or more of the time, we declare	AMOVA

	<p>the results significant at the 5% level. Note, that in calculating the probability value <math>P</math>, GenAEx always includes the observed value as ‘just another permutation’ adding this value to the 99, 999 or 9999 permutations. <math>P</math> is calculated as the <i>Number of Values <math>\geq</math> Observed Value (Including Observed Value) <math>\div</math> (Number of Permutations + 1)</i>. As a consequence, the smallest probability value <math>P</math> reported by GenAEx will never be less than <math>1 \div (\text{Number of Permutations} + 1)</math>. Thus, for 99 permutations, the smallest <math>P</math> value will be 0.01; for 999, the smallest <math>P</math> value will be 0.001, etc. In GenAEx 6.1 onwards, <math>P</math> values are more explicitly labelled. For AMOVA, probability values are reported with the caption <math>P(\text{rand} \geq \text{data})</math>, which is read as ‘the probability of a random value greater than equal to the observed data value’. These rules apply for all permutational tests in GenAEx, not just AMOVA. In Arlequin, permutational tests are performed differently and are designed to simultaneously test the probability of the Phi statistic (or equivalent) and the variance. Furthermore, the reported <math>P</math> values can equal zero. Please refer to the Arlequin guide for more details.</p>	
Binary Genetic Distance	See Genetic Distance Binary and Tutorial Part 2	
Codominant Genetic Distance	See Genetic Distance Codominant and Tutorial Part 2	
$F$ -statistics	<p>Wright’s <math>F</math>-statistics [18-20] are widely used to characterize population genetic structure. These statistics allow the partition of genetic diversity (<math>\sim</math> heterozygosity) within and among populations. GenAEx provides two separate pathways for the calculation of <math>F</math>-statistics: via the Frequency option and via AMOVA. In GenAEx, the frequency based calculation of <math>F</math>-statistics follows [1]. This option is provided largely for teaching purposes, given the wide coverage of formulas in population genetic texts, and the ease with which students can calculate the <math>F</math>-statistics by hand. For research purposes, the calculation of <math>F</math>-statistics by AMOVA is recommend, since the AMOVA framework allows for statistical testing by random permutation. Note that in GenAEx 6.1 onward, a minor modification of the <math>F</math>-statistics routine in AMOVA was applied that brings the estimates for <math>F_{st}</math> for codominant data in line with the Weir and Cockerham estimates, following formulas and notation in Peakall et al. [14]. See also AMOVA above, relevant formulas in Table 2 and Tutorial Part 1.</p>	Frequency, AMOVA

<p>Genetic Distance Binary (Diploid) and Binary (Haploid) (Binary Data only)</p>	<p>A pairwise, individual-by-individual (<math>N \times N</math>) genetic distance matrix is generated for binary data by this genetic distance option. This calculation of pairwise genetic distances for binary data follows the method of Huff et al. [13], in which any comparison with the same state yields a value of 0 (both 0 vs 0 comparisons and 1 vs 1 comparisons), while any comparison of different states (0 vs 1 or 1 vs 0) yields a value of 1. When calculated across multiple loci for a given pair of samples, this is equivalent to the tally of differences between the two genetic profiles. This genetic distance matrix is used in GenAIEx for subsequent PCA, Mantel and all Spatial analyses involving binary data. This distance option is also be used to calculate <math>\Phi_{pt}</math> via AMOVA, a measure of population genetic differentiation for binary data that is analogous to <math>F_{st}</math>. This is a Euclidean distance metric, unlike binary measures such as Nei's <math>(1 - F)</math>, and is therefore appropriate for AMOVA, which requires a Euclidean metric [12-14].</p> <p>Note that there is no difference between Binary (Diploid) and Binary (Haploid) genetic distance. The separate Diploid and Haploid options shown on the Genetic Distance Options dialog box is merely retained for continuity with the Allele Frequency Dialog box where the subsequent allele frequency calculations are different for diploid and binary data.</p>	<p>Distance-&gt;Genetic, AMOVA</p>
<p>Genetic Distance Codom-Genotypic (Codominant Data only)</p>	<p>A pairwise, individual-by-individual (<math>N \times N</math>) genetic distance matrix is calculated for codominant data by this genetic distance option. For a single-locus analysis, with <math>i</math>-th, <math>j</math>-th, <math>k</math>-th and <math>l</math>-th different alleles, a set of squared distances is defined as <math>d^2(ii, ii) = 0</math>, <math>d^2(ij, ij) = 0</math>, <math>d^2(ii, ij) = 1</math>, <math>d^2(ij, ik) = 1</math>, <math>d^2(ij, kl) = 2</math>, <math>d^2(ii, jk) = 3</math>, and <math>d^2(ii, jj) = 4</math>. See [14] and Smouse and Peakall [21] for graphical explanation of this method. This is the most important genetic distance option for codominant data, since the matrix generated is used in GenAIEx for subsequent PCA, Mantel and all Spatial analyses. This distance option can also be used to calculate <math>\Phi_{pt}</math> via AMOVA, a measure of population genetic differentiation that suppresses intra-individual variation and is therefore ideal for comparisons between codominant and haploid or binary data (see [16]), where no intra-individual variation (heterozygosity) is available. See also AMOVA above, <math>\Phi_{pt}</math> in Table 2 and Tutorial 2.</p>	<p>Distance-&gt;Genetic, AMOVA</p>
<p>Genetic Distance Codom-Allelic (Codominant Data only)</p>	<p>This <i>Codom-Allelic</i> option generates a <math>2N \times 2N</math> genetic distance matrix, representing the pairwise distances between all alleles. The first allele of individual 1 is presented, followed by the second allele of individual 1, then the first allele of individual 2, and so on. The distance between alleles is either 0 (alleles are identical) or 1 (alleles are different). Values are summed across loci. This genetic distance option can only be generated when GenAIEx is calculating <math>F_{st}</math> via AMOVA. It is not necessary to output this distance matrix for the AMOVA analysis, and since this matrix cannot be used for other analyses, its output is not generally recommended, except for advanced users. See also AMOVA above, <math>F_{st}</math> in Table 2 and Tutorial Part 2.</p>	<p>AMOVA</p>

<p>Genetic Distance Codom-Microsat</p> <p>(Codominant Data only)</p>	<p>This <i>Codom-Microsat</i> distance option produces a <math>2N \times 2N</math> distance matrix. The first allele of individual 1 is presented, followed by the second allele of individual 1, then the first allele of individual 2, and so on. Alleles must be coded by size, either the inferred number of repeats or the size of the allele in base pairs (bp). The genetic distance is calculated as the sum of the squared size difference between the two alleles in the comparisons: <math>(S1 - S2)^2</math>, where <math>S1</math> is the size of allele 1 and <math>S2</math> the size of allele 2. Distances are summed across loci. This genetic distance option can only be generated when GenAEx is calculating <math>Rst</math> via AMOVA. <math>Rst</math> is an estimator of genetic differentiation for microsatellite loci that assumes a stepwise mutation model [15, 22]. It is not necessary to output this distance matrix for the AMOVA analysis, and since this matrix cannot be used for other analyses, its output is not generally recommended, except for advanced users. See AMOVA above, <math>Rst</math> in Table 2 and Tutorial Part 2.</p>	<p>AMOVA</p>
<p>Genetic Distance Haploid</p> <p>(Haploid Data only)</p>	<p>A pairwise, individual-by-individual (<math>N \times N</math>) genetic distance matrix is generated for haploid data by this genetic distance option. The calculation of pairwise individual genetic distances for haploid data is similar to that for binary data, since any two alleles that are the same yield a distance of 0, while any pair of alleles that are different yield a distance of 1. These distances are summed over loci to give a total genetic distance. This genetic distance matrix is used in GenAEx for subsequent PCA, Mantel and all Spatial analyses involving haploid data. This distance option is also used with haploid data to calculate <math>\Phi_{pt}</math> via AMOVA, a measure of population genetic differentiation that is analogous to <math>F_{st}</math>.</p>	<p>Distance- &gt;Genetic, AMOVA</p>
<p>Genetic Distance Haploid-SSR</p> <p>(Haploid SSR Data only)</p>	<p>A pairwise, individual-by-individual (<math>N \times N</math>) genetic distance matrix is generated for haploid-SSR data by this genetic distance option. The calculation is similar to Codom-Microsat. Alleles must be coded by size, either the inferred number of repeats or the size of the allele in base pairs (bp). The genetic distance is calculated as the sum of the squared size difference between the two alleles in the comparisons: <math>(S1 - S2)^2</math>, where <math>S1</math> is the size of allele 1 and <math>S2</math> the size of allele 2. Distances are summed across loci. This option is provided for haploid microsatellite or simple sequence repeat (SSR) data only. Note that this distance estimate assumes a step-wise mutation model that may not be applicable for many data sets. Furthermore, even if a step-wise mutation model holds, SSRs in chloroplast DNA may be highly homoplasic (see Ebert and Peakall 2009), thus the estimate may yield spurious patterns in a similar way to that found with the Codom-Microsat distance and down stream estimates of <math>Rst</math>. Consequently, studies using this genetic distance estimate should also report the outcomes using the standard Haploid genetic distance (see also related comments on <math>Rst</math> in Table 2). Provided users are mindful of the underlying assumptions and risks of spurious patterns, this genetic distance matrix can be used in GenAEx for subsequent PCA, Mantel and all Spatial analyses involving haploid-SSR data. Note that the <math>\Phi_{pt}</math> generated via AMOVA with this option provides a measure of population genetic differentiation that is analogous to <math>Rst</math>. However, the current outputs from GenAEx do not note this connection. That is, AMOVA outputs from both Haploid and Haploid-SSR are called <math>\Phi_{pt}</math>, unlike their codominant counterparts. This may change in future versions of GenAEx.</p>	<p>Distance- &gt;Genetic, AMOVA</p>

Geographic Distance	A pairwise, individual-by-individual ( $N \times N$ ) linear geographic distance matrix is generated from X and Y coordinates by this distance option. See Tutorial Part 3 for background to the calculation of distances from Latitude and Longitude.	Distance- >Geographic
Haploid Disequilibrium  (Haploid Data only for a single population)	The test for Haploid Disequilibrium follows Gordon [23], based on the theory developed by Brown et al. [24] and Souza et al. [25]. The index of linkage disequilibrium is $V_o/V_e$ , where $V_e$ is the expected variance of $K$ - the number of loci for which two individuals differ. In the absence of linkage disequilibrium, the expected variance is given by $V_e$ . To test whether the ratio of $V_o/V_e$ is significantly greater than one, GenAEx employs a randomisation test, following [23, 25]. $V_r$ is calculated for each random sample as the variance of the randomized data set and the probability of observing a $V_r$ value as extreme as that measured for the original data ( $V_o$ ). This option is only available in GenAEx when a haploid data set consists of a single population.	Frequency
HWE – Tests for Hardy-Weinberg Equilibrium  (Codominant Data only)	The HWE procedures follow Hedrick [26], but similar Chi-square testing procedures for HWE are covered in most standard population genetic texts [e.g., 1-6]. In brief, for codominant genotypes at a single locus, and for a single population, we can determine whether the observed tallies of genotypes are consistent with Hardy-Weinberg expectations, as follows: 1. Determine the number of samples, 2. Determine the number of alleles, $N_a$ . 3. Count the numbers of each genotype. 4. Calculate allele frequencies. 5. Estimate the expected genotype frequencies, given the sample size of the population, either as $p^2$ for a homozygous genotype or as $2pq$ for a heterozygous genotype. 6. Test for conformity with HWE expectations by calculating the Chi-squared statistic $X^2$ . 7. Determine the degrees of freedom. 8. Given the calculated Chi-squared value and the degrees of freedom, determine from a Chi-squared table whether the observed numbers would deviate as far from the expected numbers by chance alone. If the probability of obtaining the observed Chi-squared value (given the degrees of freedom) is greater than 0.05 ( $P$ in the range 0.05 to 1.0), the result is NOT statistically significant and we accept the null hypothesis $H_0$ = Population is mating randomly. If the probability of obtaining the observed Chi-squared value (given the degrees of freedom) is less than 0.05 (in the range $0 < P < 0.05$ ), we conclude that the result is statistically significant, and we reject the null hypothesis $H_0$ , in favour of $H_1$ = Population is not mating randomly. Hedrick [2] notes that results from Chi-Square tests for HWE should be treated with caution when samples sizes $< 50$ and when the expected numbers are $< 5$ in some classes. Note that the HWE option in GenAEx is provided primarily for teaching purposes and for data exploration. Other programs such as GenePop [27] and Arlequin [17] provide superior tests which are recommended for research purposes. GenAEx offers data export to these programs. See also Chi in Table 2 and Tutorial 1.	HWE

Mantel Tests	Mantel tests for Matrix Correspondence [28] follow the methods of Smouse and coworkers [29-30], with the option for statistical testing by random permutation. The Mantel option allows tests for a statistical relationship between the elements of any two distance matrices with matching entries. Typical applications include testing for isolation-by-distance, for which one might compare a Nei genetic distance matrix (or log of the genetic distance) with the geographic distance matrix for the respective populations. Alternatively, one can test for a correlation of individual-by-individual genetic distances calculated from two different genetic markers sets, such as SSRs and AFLPs (e.g., [14, 16]). While it is easy to plot a graph of the relationship between elements from any two matrices, we cannot use the <i>P</i> -values of standard regression analysis, because the $N \times (N - 1)$ elements within each matrix cannot be independent. Consequently, we need another way to test the significance of two matrices, and the Mantel test provides such an option. This method yields a correlation coefficient for the two data matrices, with a range from $-1$ to $+1$ , with a test for a significant relationship by random permutation. The null hypothesis is that there is no significant relationship, in which case a random shuffle of the data set should yield a similar result to the observed value. On the other hand, if there is significant relationship between the two data sets, a random correlation will be more extreme (closer to $+1$ or $-1$ ) than the data value less than 5% of the time. See also Tutorial 3.	Mantel
Nearest Neighbor Distance	The calculation of nearest neighbor distances for a user-specified number of neighbors is offered by GenAlEx. Frequency distributions of nearest neighbor distances can also be generated. This option is provided to support the 2D local spatial autocorrelation option, but users may find this option useful in other contexts as well.	Spatial-> NN Dist
Nei Genetic Distance	GenAlEx offers the calculation of Nei's standard genetic distance [31-32] between pairs of populations for codominant, binary and haploid data sets. This measure is one of the most widely used for estimating genetic distance among populations. For neutral markers, under an infinite-allele-model, this genetic distance is predicted to increase linearly with time [31]. Both the biased and unbiased estimates of Nei's genetic distance [32] are calculated in GenAlEx 6.1 onwards. Hedrick [2] suggests, however, that the unbiased correction may give spurious results when homozygosity and sample size are small. See also Nei I and Nei D in Table 2 and Tutorial 1.	Frequency
Population Assignment	The frequency-based assignment test of Paetkau [33-34] is available within GenAlEx. See also [35-36] for reviews. In brief, for each sample, the expected genotype frequency at each locus is calculated, assuming random mating in the population in question, multiplied across loci and log-transformed to give a log likelihood value. For each sample, a log likelihood value is calculated for each population, using the allele frequencies of the respective population. If an allele frequency value of zero is encountered for a given allele (i.e., if the allele is absent from one of the represented populations), GenAlEx uses the value 0.01 or another (user-specified) value. A sample is assigned to the population with the highest log likelihood (i.e., the population with the least negative log-likelihood value). Alternatively, GenAlEx offers the option to convert negative log-likelihood values to positive numbers (multiplying by $-1$ ), in which case the sample is assigned to	Assignment->Pop Assign

	the population with the smallest value. We recommend the default <i>Leave one out</i> option, which uses an allele frequency estimate that leaves the sample to be assigned out of the frequency estimate. The <i>As is</i> option is primarily provided for teaching purposes and for compatibility with the <i>Sex Bias</i> option in GenAlEx. For research purposes, the program <i>GeneClass 2</i> [37] is recommended. The program <i>Structure</i> [38] provides alternative statistical options for population assignment, using different methods from <i>GeneClass</i> . GenAlEx provides data export options to both these programs. See also Tutorial 4.	
Principal Coordinates Analysis	Principal Coordinate Analysis (PCA) is a multivariate technique that allows one to find and plot the major patterns within a multivariate data set (e.g., multiple loci and multiple samples). The mathematics is complex, but in essence, PCA is a process by which the major axes of variation are located within a multidimensional data set. For multidimensional data sets, each successive axis explains proportionately less of the total variation, such that when there are distinct groups, the first 2 or 3 axes will typically reveal most of the separation among them. The procedure in GenAlEx is based on an algorithm published by Orloci [39]. Four different options are provided, two based on the conversion of the distance matrix to a covariance matrix, and two working directly from the input distance matrix. The two standardization options divide the respective distance or covariance inputs by the square root of $n-1$ . See also Tutorials 1 and 3.	PCA
Phi Statistics	The estimation of $\Phi$ -statistics parallels the same logic as for $F$ -statistics for codominant data, except that $\Phi$ -statistics are also estimable from binary and haploid data [14]. See AMOVA above. Formulas for the various $\Phi$ -statistics are provided in Table 2. See also Tutorial 2.	AMOVA
Pairwise Relatedness	The calculations for several pairwise relatedness estimators are provided by GenAlEx: (1) Ritland (1996), (2) Lynch and Ritland (1999) and (3) the estimator of Queller and Goodnight (1989) [40-43]. For a summary of the formulas, see Ritland [42]. The algorithm for these calculations follows the publicly available code in the software program MaRQ by K. Ritland. Note that as in MaRQ the Lynch and Ritland (1999) estimate of relatedness in GenAlEx has a default range of 0 to 0.5. Some other programs report this value as 2x the MaRQ/GenAlEx estimate (range 0 to 1). You can choose the option 2x to give this range from 0 to 1.	Relatedness->Pairwise
Probability of Identity	The Probability of Identity $PI$ provides an estimate of the average probability that two unrelated individuals, drawn from the same randomly mating population, will by chance have the same multilocus genotype. Also called Population Match Probability. $PI$ is widely used in DNA forensics [44] as a indication of the statistical power of a specific set of marker loci. This is also used for genetic tagging in molecular ecology [45-46], an indication of the minimum number of loci required for reliable genetic tagging. GenAlEx provides both estimates of $PI$ and $PIsibs$ . The latter statistic is calculated, following [45, 47], and takes into account the genetic similarity among siblings. When additional information is known about likely levels of inbreeding and population substructure, more complex estimators of $PI$ are available [48].	Multilocus->Prob. Ident.

Probability of Exclusion	GenAIEx offers the calculation of three probability estimates for parentage exclusion, following Jamieson and Taylor [49]. <i>P1</i> estimates the probability of exclusion when both parents are known. <i>P2</i> estimates the probability of exclusion when only one parent is known, and <i>P3</i> estimates the probability of excluding two putative parents. Formulae are provided in Table 2.	Multilocus->Prob. Excl.
Probability of Clonality	Several different probability estimates for inferring clonality in plants from codominant data are provided in GenAIEx, following [50-52]. Formulae for these probability estimates are provided in Table 2. GenAIEx also offers tools for finding repeated matching genotypes that may represent ramets of the same clone/genet. Data subsets of genotypes repeated more than once, and the converse of data sets without repeated genotypes, can also be extracted by GenAIEx. An option to estimate the size of putative clones is also provided. This option requires the clonal coordinates output as a starting point. Note that programs such as MLGsim provide alternative simulation approaches to the detection of clones [51].	Clonal->Prob. Clone
Spatial Autocorrelation	GenAIEx provides an extensive series of spatial autocorrelation analyses, following the methods of Smouse and Peakall, Peakall et al., Double et al., Smouse et. al. [21, 53-57]. See below for further information on statistical testing, Tutorial Part 3 and Tutorial Appendix 3.1 for advanced spatial autocorrelation options.	Spatial
Spatial Autocorrelation and Statistical Testing	For spatial autocorrelation, the null ( $H_0$ ) and alternative hypotheses ( $H_1$ ) are: $H_0 =$ A random distribution of genotypes in space ( $r = 0$ ), $H_1 =$ a non-random distribution of genotypes in space ( $r \neq 0$ ). In order to distinguish between these hypotheses, GenAIEx offers statistical testing for spatial autocorrelation, based on two methods: (i) random permutation, similar to that used for AMOVA and Mantel, and (ii) bootstrap estimates of $r$ . Random permutation allows us to generate a distribution of permuted ( $r_p$ ) values under the assumption of no spatial structure, by the random shuffling of all individuals among the geographic locations. From 999 such random shuffles (plus the observed value as the 1000 <sup>th</sup> permutation), the values of the 25 <sup>th</sup> and 975 <sup>th</sup> ranked $r_p$ values are taken to define the upper and lower bounds of the 95% confidence interval. If the calculated $r$ -value falls outside this confidence belt, significant spatial genetic structure is inferred. This is the classic two-tailed test. When one's interest is in the detection of positive autocorrelation, as predicted under restricted dispersal, GenAIEx also computes a one-tailed probability. In this case, the individual $r_p$ values are compared with the observed $r$ -value, to estimate the probability of randomly achieving a value greater than or equal to the observed $r$ . If this probability is less than 0.05, the alternative hypothesis of positive spatial genetic structure is accepted. Bootstrap estimates allow us to place a confidence interval around the observed estimate of $r$ by drawing (with replacement) from within the set of pairwise comparisons for a specific distance class. For each of 1,000 bootstrap trials, the bootstrap autocorrelation coefficient ( $r_{bs}$ )	Spatial

	is calculated for each distance class. The 25th and 975th ranked $r_p$ are then taken to define 95% confidence interval. When the bootstrap confidence interval does not straddle $r = 0$ , significant spatial genetic structure is inferred. Note that while providing an alternative statistical test, this bootstrap test is less powerful than permutational tests, since the number of samples per distance class is much smaller than the $n(n-1)/2$ comparisons used during permutation. Thus, for small sample sizes, bootstrap errors tend to be larger than the permutational errors. The bootstrap test is conservative, favouring the null hypothesis to a greater extent than does the permutational test. Despite this limitation, the calculation of bootstrap errors enable a graphical test of statistical significance among different $r$ values, using the respective 95% confidence intervals.	
Spatial Autocorrelation Options (Advanced)	GenAIEx 6.2 onwards provides users with access to new spatial genetic analysis procedures that were developed and described in Smouse et al., Beck et al. and Gonzales et al.[55-57]. GenAIEx users are strongly advised to complete the Tutorial 3 fully and read Tutorial Appendix 3.1 before attempting to use these Advanced Spatial options.	Spatial->Adv Single Pop Spatial->Adv Multiple Pop
Sex Biased Dispersal	GenAIEx implements the sex-biased assignment test procedure developed by Favre et al. [58] and extended by Mossman and Waser [59]. For each individual, GenAIEx calculates a log likelihood assignment test value, as described under population assignment, except that the <i>As is</i> allele frequency estimate is used instead of the <i>Leave one</i> out option. Next, an Assignment Index correction ( <i>Aic</i> ) for each individual is calculated as: Individual (log likelihood – mean log likelihood of the population). <i>Aic</i> values will average zero for each population, while negative values will characterize individuals with a higher probability of being immigrants. In GenAIEx a plot of the mean <i>Aic</i> for males versus females is provided, as well as a plot of the frequency distribution of corrected assignment indices ( <i>Aic</i> ) for the males and females. The genetic signal of sex-biased dispersal is indicated when there is a difference in the frequency distribution of <i>Aic</i> values among males and females. Note that GenAIEx does not yet offer the recommended non-parametric test of this difference.	Assignment->Sex Bias
Shannon Diversity	GenAIEx 6.3 onwards provides options for calculating Shannon Diversity estimates among multiple populations and locus-by-locus <i>G</i> -tests of mutual information following Sherwin et al [60]. Further options associated with Shannon Diversity analysis are planned for future releases of GenAIEx. For further information see Tutorial 1 and the accompanying Tutorial Appendix 1.1 by Bill Sherwin that provides an extensive overview on Shannon Diversity statistics.	
TwoGener	Two generational analysis of pollen flow following [61-64] is provided for codominant data. Note that GenAIEx does not duplicate some of the features offered in the software programs FAMOZ, the server based program GENER and in GENETIC STUDIO provided by Dyer [65], or the program POLDISP. Data export to these software packages is provided by GenAIEx 6.3 onwards. For details see Tutorial 6.	Two Gener

**Table 2: A summary of the statistics used in GenAIEx 6.3**

GenAIEx Notation	Measure	Formula	Range	Notes	GenAIEx Worksheet	Ref
	Allele Frequency (Codominant Data)	$FreqAllele_x = \frac{2N_{xx} + N_{xy}}{2N}$	0 to 1	Calculated for a single locus. Determined for each allele. $N_{xx}$ = # of XX homozygous individuals, and $N_{xy}$ = # of XY heterozygous individuals, where Y can be any other allele. $N$ = the number of samples. Can also be determined simply by direct count of the proportion of different alleles.	AFL AFP APT	[1, 4, 6-8, 26]
	Allele Frequency (Binary Data)	Assuming random mating: - Presence represents both genotypes $AA$ or $Aa$ - Absence represents the genotype $aa$ - Allele $A$ has Freq. $p = 1 - q$ - Allele $a$ has Freq. $q = 1 - p$ - Freq. of genotype $aa$ = $q^2$ = Freq. of band absence = $1 -$ Freq. of band presence, so $q = \sqrt{\text{Freq. of absence}}$		With dominant binary markers such as AFLPs, it is not possible to directly calculate allele frequencies, as for codominant or haploid markers. If, however, we can assume either complete outcrossing (most animals and some plants) or obligate selfing (some agricultural plants), we can still estimate the allele frequencies. The basis of the GenAIEx estimate is shown to the left. Following Lynch and Milligan [66], it assumes complete outcrossing, but does not impose the pruning of low frequency bands recommended in this reference. Note: Zhivotovsky [67] has developed an alternative Bayesian allele frequency estimation procedure that is available in other programs such as FAMD. GenAIEx 6.3 offers data export to this program.	AFL AFP APT	[66]
	Allele Frequency (Haploid Data)	$FreqAllele_x = \frac{N_x}{N}$		Calculated for single loci, and determined for each allele, where $N_x$ = number of the $x$ alleles and $N$ = the number of samples. Can also be determined by direct count of the proportion of different alleles.	AFL AFP APT	
Aic	Assignment Index	Individual (log-likelihood – mean log-likelihood) score of the population.		See also Sex Biased Dispersal in Table 1	SB, FDSB	[58]

Chi	Chi-Squared Test for HWE	$X^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$	0 to $\infty$	$O_i$ is the observed # of individuals of the $i$ th genotype, and $E_i$ the expected number. Degrees of freedom for the Chi-Squared test are calculated as $DF = [Na(Na-1)]/2$ , where $Na$ is the number of alleles at the locus. See also HWE in Table 1	HW, HWS	[26]
F	Fixation Index (Codominant Data only)	$F = \frac{H_e - H_o}{H_e}$	-1 to 1	Calculated on a per locus basis. GenAEx also provides the arithmetic mean across loci. Values close to zero are expected under random mating, while substantial positive values indicate inbreeding or undetected null alleles. Negative values indicate excess of heterozygosity, due to negative assortative mating, or heterotic selection.	HFL, HFP	[1]
Fis	Fis via Frequency (Codominant Data only)	$F_{IS} = \frac{\bar{H}_e - \bar{H}_o}{\bar{H}_e}$	-1 to 1	The inbreeding coefficient within individuals, relative to the population. <i>Fis</i> measures the reduction in heterozygosity of an individual, due to non random mating within each population (subdivision). (Note that the <i>s</i> for subpopulations in the notation for <i>Fis</i> is equivalent to populations in GenAEx.)	HFL, HFP	[1]
Fis	Fis via AMOVA (Codominant Data only)	$F_{IS} = \frac{V_{AI}}{(V_{WI} + V_{AI})}$	-1 to 1	$V_{WI}$ is the variance within individuals, and $V_{AI}$ the variance among individuals. When the optional <i>Specialized Permute</i> option is chosen, GenAEx 6.3 performs an additional permutational procedure that shuffles only individuals within populations to calculate the probability for <i>Fis</i> . In practice, unless there is exceptionally strong population structure, the <i>Standard</i> and <i>Specialized Permute</i> options yield similar probability estimates.	Fst	
Fit	Fit via Frequency (Codominant Data only)	$F_{IT} = \frac{H_T - \bar{H}_o}{H_T}$	-1 to 1	The inbreeding coefficient within individuals, relative to the total. This statistic takes into account the effects of both non random mating within and genetic differentiation among populations.	HFL, HFP	[1]
Fit	Fit via AMOVA (Codominant Data only)	$(1 - F_{IT}) = (1 - F_{IS})(1 - F_{ST})$	-1 to 1	See above.	Fst	[14]

Fit	Fit via AMOVA, with regional data structure  (Codominant Data only)	$\left(1 - F_{IT}\right) = \left(1 - F_{IS}\right)\left(1 - F_{SR}\right)\left(1 - F_{RT}\right)$	-1 to 1	See above.	Fst	[14]
Frt	<i>Frt</i> via AMOVA, with regional data structure.  (Codominant Data only)	<p>Default option</p> $F_{RT} = \frac{V_{AR}}{(V_{WI} + V_{AI} + V_{AP} + V_{AR})}$ <p><i>Suppress within individual analysis</i> option</p> $F_{RT} = \frac{V_{AR}}{(V_{WP} + V_{AP} + V_{AR})}$	-1 to 1	For the default option: $V_{AR}$ is variance among regions; $V_{AP}$ is the variance among populations within regions; $V_{WI}$ is the variance within individuals; and $V_{AI}$ is variance among individuals. When the option <i>Suppress within individual analysis</i> is checked in the AMOVA dialog box: $V_{AR}$ is variance among regions, $V_{AP}$ is the variance among populations within regions, and $V_{WP}$ is the variance within populations. <i>Frt</i> is calculated as a proportion of the variance among regions, relative to the total, and represents the correlation within a region, relative to the total, analogous to $\Phi_{rt}$ when the data are Haploid or Binary.	Fst	[14]

Fsr	Fsr via AMOVA with regional data structure.  (Codominant Data Only)	<p>Default option</p> $F_{SR} = \frac{V_{AP}}{(V_{WI} + V_{AI} + V_{AP})}$ <p><i>Suppress within individual analysis</i> option</p> $F_{SR} = \frac{V_{AP}}{(V_{AP} + V_{WP})}$	-1 to 1	<p>For the default option: <math>V_{AP}</math> is the variance among populations within regions; <math>V_{WI}</math> is the variance within individuals; <math>V_{AI}</math> is the variance among individuals within populations. When the option <i>Suppress within individual analysis</i> is checked in the AMOVA dialog box, <math>V_{AP}</math> is the variance among populations within regions and <math>V_{WP}</math> is the variance within populations. <math>F_{sr}</math> is calculated as the proportion of variance among populations within regions, relative to the variance both within and among populations, and represents the correlation between individuals within a population, relative to the individuals from the same region. When the optional <i>Specialized Permute</i> option is chosen, GenAEx 6.3 performs an additional permutation procedure that shuffles only within regions to calculate the probability for <math>F_{sr}</math>. In practice, unless there is exceptionally strong regional structure, the <i>Standard</i> and <i>Specialized Permute</i> options result in similar probability estimates.</p>	Fst	[14]
Fst	Fst via Frequency  (Codominant Data only)	$F_{ST} = \frac{H_T - \bar{H}_e}{H_T}$	0 to 1	<p>The inbreeding coefficient within subpopulations, relative to the total. <math>F_{st}</math> provides a measure of the genetic differentiation among populations. That is, the proportion of the total genetic diversity (~ heterozygosity) that separates the populations. <math>F_{st}</math> is typically greater than or equal to zero. If all subpopulations are in Hardy-Weinberg equilibrium with the same allele frequencies, <math>F_{st} = 0</math>. <math>F_{st}</math> is more or less equivalent to <math>G_{st}</math>. (Note that the <math>s</math> used for subpopulations in the notation for <math>F_{st}</math> is equivalent to populations in GenAEx.). Estimates of <math>F_{st}</math> are also calculated in GenAEx via the AMOVA option, the recommend course for research purposes.</p>	Fst FstP FstL FstT	[1, 4, 8]

Fst	<p><i>Fst</i> via AMOVA with no regional data structure.  (Codominant Data only)</p>	<p>Default option</p> $F_{ST} = \frac{V_{AP}}{(V_{WI} + V_{AI} + V_{AP})}$ <p><i>Suppress within individual analysis</i> option</p> $F_{ST} = \frac{V_{AP}}{(V_{AP} + V_{WP})}$	0 to 1	<p>For the default option: <math>V_{AP}</math> is the variance among populations; <math>V_{WI}</math> is the variance within individuals; and <math>V_{AI}</math> is the variance among individuals within populations. When the option <i>Suppress within individual analysis</i> is checked in the AMOVA dialog box; <math>V_{AP}</math> is the variance among populations and <math>V_{WP}</math> is the variance within populations. <i>Fst</i> is calculated as the proportion of the variance among populations, relative to the total variance. The estimation of <i>Fst</i> parallels that for <math>\Phi_{pt}</math> (see below), except that <i>Fst</i> can only be estimated via AMOVA for codominant data, using the option <i>Codom-Allelic</i> genetic distance. In GenAlEx, the respective variance components of the <i>Fst</i> statistic (e.g., <math>V_{AP}</math> &amp; <math>V_{WP}</math>) are output in the AMOVA table under the heading “Est. Var.” <i>Fst</i> can be calculated directly by hand from this table, using the formulas shown here. Subpopulations in the notation for <i>Fst</i> are equivalent to populations in GenAlEx, so we use <i>P</i> for populations in the variance notation <math>V_{WP}</math>, etc. <i>Fst</i> is typically greater than or equal to zero, but with the AMOVA routine can sometimes be negative. See also AMOVA, <math>\Phi</math>-statistics and <i>Rst</i>.</p>	Fst FstP FstL FstT	[14]
Fst	<p><i>Fst</i> via AMOVA, with regional data structure.  (Codominant Data only)</p>	<p>Default option</p> $F_{ST} = \frac{V_{AP} + V_{AR}}{(V_{WI} + V_{AI} + V_{AP} + V_{AR})}$ <p><i>Suppress within individual analysis</i> option</p> $F_{ST} = \frac{V_{AP} + V_{AR}}{(V_{WP} + V_{AP} + V_{AR})}$	0 to 1	<p>For the default option: <math>V_{AR}</math> is the variance among regions; <math>V_{AP}</math> is the variance among populations; <math>V_{WI}</math> is the variance within individuals; <math>V_{AI}</math> is the variance among individuals within populations. When the option <i>Suppress within individual analysis</i> is checked in the AMOVA dialog box: <math>V_{AR}</math> is the variance among regions; <math>V_{AP}</math> is the variance among populations; <math>V_{WP}</math> is the variance within populations.</p>	Fst FstP FstL FstT	[14]

Fst	Linearized Fst	$Linearized F_{ST} = \frac{F_{ST}}{(1 - F_{ST})}$		A transformation for pairwise population Fst values recommended by Slatkin [68]. GenAlEx also offers the analogous transformation for PhiPT.	LinFst	
GD	Genetic Distance Binary	$D = n \left[ 1 - \frac{2n_{xy}}{2n} \right]$	0 – n	Here, $2n_{xy}$ = the number of shared character states and $n$ equals the total number of binary characters. When calculated across multiple loci for a given pair of samples, this is equivalent to the tally of state differences among the two DNA profiles. See Table 1 for details on other genetic distance options.	GD	[13, 16]
GGD	Geographic Distance	$D = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$	0-n	Here, $x_i$ and $y_i$ are the coordinates for the $i$ th sample and $x_j$ and $y_j$ are the coordinates for the $j$ -th sample.	GGD	
GP	Genotype Probability (Codominant Data only)	$GP = \prod p_i^2 \times \prod 2p_i p_j$	0 to 1	$\Pi$ indicates chain multiplication across each locus, $p_i$ is the frequency of the allele at homozygous loci, $p_i$ and $p_j$ are the frequencies of alleles at heterozygous loci. Also called the <i>DNA Profile Probability</i> and the <i>Random Match Probability</i> , the chance of a random match to a given specific genotype or DNA profile. Widely used in DNA forensics. See also Pgen.	GP	[1, 7]
h	Haploid Genetic Diversity (Haploid Binary and Haploid Data only)	$h = 1 - \sum p_i^2$	0 to 1	Here, $p_i$ is the frequency of the $i$ th allele. Haploid genetic diversity provides an indication of the probability that two individuals will be different (e.g., 2 haploid strains of bacteria).	HDL, HDP	[14]
H	Mean haploid genetic diversity across loci (Haploid Data only)	$H = \sum_{i=1}^k h/k$	0 to 1	Calculated as the arithmetic mean of $h$ across $k$ loci.	HDL, HDP	[14]
HD	Haploid Genetic Distance			See Genetic Distance Haploid		
He	Expected Heterozygosity or Genetic Diversity (Codominant)	$He = 1 - \sum p_i^2$	0 to 1	Calculated on a single locus basis. GenAlEx also provides the arithmetic mean across loci. Here, $p_i$ is the frequency of the $i$ th allele.	HFL, HFP	[1]

He	Expected Heterozygosity (Diploid Binary Data only)	$He = 1 - \sum p^2$	0 to 1	Calculated as for <i>He</i> for codominant data, once allele frequencies, <i>p</i> have been calculated as described under Allele Frequency for Binary Data (see above).	HFL, HFP	[16, 66]
Ho	Observed Heterozygosity (Codominant Data only)	$Ho = \frac{\text{No. of Hets}}{N}$	0 to 1	This is calculated on a per locus basis. GenAEx also provides the arithmetic mean across loci. Where the number of heterozygotes is determined by direct count and <i>N</i> = the number of samples.	HFL, HFP	[1]
Ht	Total expected heterozygosity (Codominant Data only)	$H_T = 1 - \sum_{i=1}^h \bar{p}_i^2$	0 to 1	A partition of the total genetic diversity among populations, used in the calculation of <i>F</i> -statistics. Here, <i>p<sub>i</sub></i> is the frequency of the <i>i</i> th allele, averaged over populations. <i>H<sub>T</sub></i> is the expected heterozygosity if all the populations were pooled (i.e., if there were no population subdivision). Only output by GenAEx if the Step-by-Step option is chosen in the <i>Frequency</i> options dialog box.	HFL, HFP	[1]
I	Information index	$I = -\sum p_i \ln p_i$	> 0	Calculated on a single-locus basis. GenAEx also provides the arithmetic mean across loci. <i>ln</i> = the natural logarithm and <i>p<sub>i</sub></i> is the frequency of the <i>i</i> th allele. Same as Shannon-Weaver Index of ecology. Unlike <i>He</i> , not bounded by 1 and may therefore be a better measure of allelic and genetic diversity, though seldom used in genetic studies. GenAEx 6.3 now offers this calculation for all data types.	HFL, HFP HDP, HDL	[69]
	Linearized Fst			See Fst Linearized		
Log-L	Log likelihood for Population Assignment (Codominant Data only)	$Log(\prod p_i^2 \times \prod 2p_i p_j)$		Calculated for a given genotype, where <i>p<sub>i</sub></i> is the frequency of the <i>i</i> th allele, and <i>p<sub>j</sub></i> the frequency of the <i>j</i> -th allele at each locus in a multilocus genotype. See also Population Assignment.	PI	[33-35, 70]

Mean $H_e$	Expected heterozygosity averaged across populations (Codominant Data only)	$\overline{H_e} = \sum_{i=1}^k H_e(i)/k$	0 to 1	The average per-population genetic diversity, used in the calculation of $F$ -statistics. Where $H_e(i)$ is the expected heterozygosity in population $i$ , and $k$ is the number of populations. Only output by GenAlEx when the Step-by-Step option is chosen under in the Allele Frequency Options dialog box.	HFL, HFP, APT	[1]
Mean $H_o$	Observed heterozygosity, averaged across populations (Codominant Data only)	$\overline{H_o} = \sum_{i=1}^k H_o(i)/k$	0 to 1	The average observed heterozygosity of a collection of populations, used in the calculation of $F$ -statistics. Here, $H_o(i)$ is the observed heterozygosity in population $i$ , $k$ is the number of populations. Only output by GenAlEx when the Step-by-Step option is chosen in the Allele Frequency Options dialog box.	HFL, HFP	[1]
Na	No. of alleles (Codominant and Haploid Data)		> 1	Determined by direct count, ranging from 1 to $N$ , where $N$ is the sample size. GenAlEx also provides the arithmetic mean across loci.	HFL, HFP, APT	
Na Freq. > 0.05	Na Freq. > 0.05 (Codominant and Haploid Data)			Number of alleles with frequency greater than 5%.	APT	[16]
	No. LComm Alleles (<=50%)			Number of locally common alleles (Freq. > 5%) occurring in 50% or less of the populations.	APT	[16]
	No. LComm Alleles (<=25%)			Number of locally common alleles (Freq. > 5%) occurring in 25% or less of the populations.	APT	[16]
	No. private alleles			Equivalent to the number of alleles unique to a single population in the data set.	APT, PAL, PAS	[16]
Ne	Effective number of alleles (Diploid Binary and Codominant Data only)	$N_e = \frac{1}{1 - H_e}$	1 - n alleles	Calculated for a single locus. GenAlEx also provides the arithmetic mean across loci. This measure enables meaningful comparisons of allelic diversity to be made across loci with diverse allele frequency distributions. Provides an estimate of the number of equally frequent alleles in an ideal population with homozygosity equivalent to the actual population	HFL, HFP	[69]

Nei GD	Nei's Genetic Distance	$Nei\ GD = -\ln(I)$	> 0	$I$ is Nei's Genetic Identity (see below, and also Nei's Unbiased Genetic Identity). GenAEx 6.3 provides a step-by-step option to illustrate this calculation.	NeiP NeiL NeiT	[26]
Nei ID	Nei's Genetic Identity	$Nei\ ID = \frac{J_{xy}}{\sqrt{(J_x J_y)}}$ $J_{xy} = \sum_{i=1}^k p_{ix} p_{iy},$ $J_x = \sum_{i=1}^k p_{ix}^2, \quad J_y = \sum_{i=1}^k p_{iy}^2$	0 to 1	Here, $p_{ix}$ and $p_{iy}$ are the frequencies of the $i$ th allele in populations $x$ and $y$ . For multiple loci, $J_{xy}$ , $J_x$ and $J_y$ are calculated by summing over all loci and alleles and dividing by the number of loci. These average values are then used to calculate $I$ . GenAEx 6.3 provides a step-by-step option to illustrate this calculation. See also Tutorial 1.	NeiP NeiL NeiT	[26]
Nm	Number of Migrants (Codominant Data only)	$Nm = [(1 / F_{ST}) - 1] / 4$		Where $F_{st}$ represents the degree of population genetic differentiation.	Fst FstT	[4, 8]
Nm	Haploid Number of Migrants (Haploid Data only)	$Nm = [(1 / \phi_{PT}) - 1] / 2$		The haploid equivalent of Nm, where $\phi_{PT}$ the haploid analog of $F_{st}$ represents the degree of population genetic differentiation. Note division by 2 rather than 4 for this haploid case.	Fst FstT	[4, 8]
Nm	Nm via Shannon's Mutual Information index (Diploid)	$Nm = \left( \frac{0.156}{{}^s H_{UA}} \right)^2$		For diploid species with effective population sizes > 500 estimates of Nm among pairs of populations can be computed via ${}^s H_{UA}$ as shown. See also sHua in this table, Tutorial 1 and Tutorial Appendix 1.1 for further details.	SH	[60]
Nm	Nm via Shannon's Mutual Information index (Haploid)	$Nm = \left( \frac{0.22}{{}^s H_{UA}} \right)^2$		For haploid species with effective population sizes > 1000 estimates of Nm among pairs of populations can be computed via ${}^s H_{UA}$ as shown. See also sHua in this table, Tutorial 1 and Tutorial Appendix 1.1 for further details.	SH	[60]

P	Polymorphism (Codominant, Haploid)	Calculated as percentage of polymorphic loci across loci.	0 to 100 %	Frequently reported in allozyme studies, where the type and number of loci investigated are similar across studies. Of limited value for DNA based markers such as SSRs, where comparisons make little sense, because the selection of markers is often based on their degree of polymorphism. May be useful for multi-locus DNA markers such as AFLPs.	HFL, HFP	
PhiPR	$\Phi_{pr}$ via AMOVA with regional data structure.	$\phi_{PR} = \frac{V_{AP}}{(V_{AP} + V_{WP})}$	-1 to 1	$V_{AP}$ is the variance among populations within regions and $V_{WP}$ is the variance within populations. $\Phi_{PR}$ is calculated as the proportion of variance among populations within regions, relative to the variance among and within populations. $\Phi_{PR}$ represents the correlation between individuals within a population, relative to the individuals from the same region. Typically greater than zero. When the optional <i>Specialized Permute</i> option is chosen, GenAlEx 6.3 performs an additional permutational procedure that shuffles only within regions to calculate the probability for <i>PhiPR</i> . In practice, unless there is exceptionally strong regional structure, the <i>Standard</i> and <i>Specialized Permute</i> options yield similar probability estimates.	PhiPT PhiPTP PhiPTL PhiPTT	[14]

PhiPT	$\Phi_{PT}$ via AMOVA without regional data structure.	$\phi_{PT} = \frac{V_{AP}}{(V_{AP} + V_{WP})}$	-1 to 1	$V_{AP}$ is the variance among populations and $V_{WP}$ the variance within populations. $\Phi_{PT}$ is calculated as the proportion of the variance among populations, relative to the total variance. $\Phi_{PT}$ represents the correlation between individuals within a population, relative to the total. Analogous to $Fst$ when the data are Haploid or Binary. May also be calculated via AMOVA for codominant data when <i>Codom-Genotypic</i> genetic distance is used. In this latter case, the partition of variation within individuals is suppressed. For codominant data, $\Phi_{PT}$ is best for comparisons between Codominant data and Binary or Haploid data [16]; otherwise $Fst$ via AMOVA is recommended, because $Fst$ is more widely used. Typically greater than zero.	PhiPT PhiPTP PhiPTL PhiPTT	[14]
PhiPT	$\Phi_{PT}$ via AMOVA with regional data structure.	$\phi_{PT} = \frac{V_{AP} + V_{AR}}{(V_{WP} + V_{AP} + V_{AR})}$	-1 to 1	$V_{AR}$ is the variance among regions, $V_{AP}$ the variance among populations, and $V_{WP}$ the variance within populations. See above.	PhiPT PhiPTP PhiPTL PhiPTT	[14]
PhiRT	$\Phi_{RT}$ via AMOVA with regional data structure.	$\phi_{RT} = \frac{V_{AR}}{(V_{WP} + V_{AP} + V_{AR})}$	-1 to 1	$V_{AR}$ is the variance among regions, $V_{AP}$ the variance among populations, within regions, and $V_{WP}$ is the variance within populations. $\Phi_{RT}$ is calculated as the proportion of the variance among regions, relative to the total variance. $\Phi_{RT}$ represents the correlation within a region, relative to the total. Analogous to $Frt$ for codominant data.	PhiPT PhiPTP PhiPTL PhiPTT	[14]

PI	Probability of Identity (Codominant Data only)	$PI = 2(\sum p_i^2)^2 - \sum p_i^4$ for each locus.	0 to 1	Here, $p_i$ is the frequency of the $i$ th allele at a locus. For multiple loci calculated as the product of individual locus $PI$ 's. $PI$ represents the average probability of a match for any genotype, rather than for a specific genotype, as in Genotype Probability. $PI$ is widely used in DNA forensic analysis [11, 44, 71] where it is also called the <i>Match Probability</i> , <i>Matching Probability</i> and <i>Power of Inclusion</i> . $1-PI$ is called the <i>Exclusion Power</i> , or <i>Power of Discrimination</i> [10-11, 71]. $PI$ is also used for assessing the number of loci required for genetic tagging in molecular ecology [45-47].	PI	[44-47]
PIsibs	Probability of Identity Sibs (Codominant Data only)	$PIsibs = 0.25 + (0.5\sum p_i^2) + [0.5(\sum p_i^2)^2] - (0.25\sum p_i^4)$	0 to 1	In addition to $PI$ GenA1Ex also calculates the more conservative $PIsibs$ that estimates the probability of identity among siblings [45-46]. As in $PI$ , $p_i$ is the frequency of the $i$ th allele at a locus. For multilocus genotypes, the $PIsibs$ of the genotype is calculated as the product of individual locus $PIsibs$ . See Peakall et al. [47] for an example application of $PIsibs$ .	PI	[44-47]
P1	Probability of Exclusion (Codominant Data only)	$P1 = 1 - 2\sum p_i^2 + \sum p_i^3 + 2\sum p_i^4 - 3\sum p_i^5 - 2(\sum p_i^2)^2 + 3\sum p_i^2 \sum p_i^3$	0 to 1	$P1$ estimates the probability of exclusion when both parents are known (following equation 1a in [49]).	PX1	[49]
P2	Probability of Exclusion (Codominant Data only)	$P2 = 1 - 4\sum p_i^2 + 2(\sum p_i^2)^2 + 4\sum p_i^3 - 3\sum p_i^4$	0 to 1	$P2$ estimates the probability of exclusion when only one parent is known (following equation 2a in [49]).	PX2	[49]
P3	Probability of Exclusion (Codominant Data only)	$P3 = 1 + 4\sum p_i^4 - 4\sum p_i^5 - 3\sum p_i^6 - 8(\sum p_i^2)^2 + 8(\sum p_i^2)(\sum p_i^3) + 2(\sum p_i^3)^2$	0 to 1	$P3$ estimates the probability of excluding two putative parents (following equation 3a in [49]).	PX3	[49]

$P_{gen}$	Probability of Genotype (Codominant Data only)	$P_{gen} = (\prod p_i)2^h$	0 to 1	Identical to the genotype probability. $P_{gen}$ provides an estimate of the probability of identical genotypes arising from sexual reproduction and random mating, where $p_i$ is the frequency of each allele (two per locus) observed in the multilocus genotype and $h$ the number of loci that are heterozygous (see [50, 52])	CLP	[50, 52]
$P_{se}$	Probability of Second Encounter (Codominant Data only)	$P_{se} = 1 - (1 - P_{gen})^N$	0 to 1	$P_{se}$ provides an estimate of the probability of a second encounter of a specific multilocus genotype generated by sexual reproduction under random mating. GenAlEx calculates $P_{se}$ using $N$ = total no. of samples, irrespective of the number of different genotypes. The estimate $P_{sex}$ $N_{gen}$ is calculated using $n$ = no. of different genotypes following [50]	CLP	[50-51]
$P_{sex}$	Probability of Sex (Codominant Data only)	$P_{sex} = \sum_n^N \frac{N!}{n!(N-n)!} (P_{gen})^n (1 - P_{gen})^{N-n}$	0 to 1	$P_{sex}$ provides an estimate of the probability of obtaining $n$ repeated multilocus genotypes in a sample of size $N$ by sexual reproduction under random mating. GenAlEx calculates $P_{sex}$ using $N$ = total no. of samples, irrespective of the number of different genotypes. The estimate $P_{sex}$ $N_{gen}$ is calculated using $N$ = no. of different genotypes.	CLP	[50-51]
$r$	Autocorrelation coefficient		-1 to 1	See Tutorial 3 for a detailed overview of the spatial autocorrelation procedures in GenAlEx.		[21, 53-54]
$r$	Pairwise Relatedness		-1 to 1	For a summary of the formulas, see Ritland [42]. See also Pairwise Relatedness in Table 1 and Tutorial 4.	RI, LR, QG, PSA	[42]
$rc$	Autocorrelation coefficient across multiple populations		-1 to 1	See Tutorial 3 for a detailed overview of the spatial autocorrelation procedures in GenAlEx.		[21, 53-54]

Rst	Rst via AMOVA with no regional data structure.  (Codominant Data only)	<p>Default option</p> $R_{ST} = \frac{V_{AP}}{(V_{WI} + V_{AI} + V_{AP})}$ <p>Suppress within individual analysis option</p> $R_{ST} = \frac{V_{AP}}{(V_{AP} + V_{WP})}$	-1 to 1	The estimation of <i>Rst</i> parallels that for <i>Fst</i> and $\Phi_{PT}$ (see above) except that <i>Rst</i> can only be estimated via AMOVA for SSR data for which alleles are coded in either base pair size (bp) or number of repeats with the option <i>Codom-Microsat</i> genetic distance. <i>Rst</i> was introduced by Slatkin [22] as an <i>Fst</i> analogue that uses the stepwise mutation model (SMM) to characterize microsatellite loci. In practice, despite initial enthusiasm for this statistic, variation at microsatellites is rarely as simple as assumed by the SMM model, and the statistic is less informative than <i>Fst</i> , as a consequence. See also AMOVA, <i>Phi</i> Statistics, <i>Fst</i> and Genetic Distance.	Rst RstP RstL RstT	[15, 22]
rx	Mantel correlation coefficient	$r_{xy} = \frac{SPxy}{\sqrt{[SSxSSy]}}$ $SSx = \sum_{i \neq j}^N (x_{ij} - \bar{x})^2$ $SSy = \sum_{i \neq j}^N (y_{ij} - \bar{y})^2$ $SPxy = \sum_{i \neq j}^N (x_{ij} - \bar{x})(y_{ij} - \bar{y})$	-1 to 1	<i>SPxy</i> is the sum of cross products of corresponding elements of the <b>X</b> and <b>Y</b> Matrices; <i>SSx</i> is the sum of products of <b>X</b> matrix elements and <i>SSy</i> that of <b>Y</b> matrix elements. In addition to listing <i>rx</i> , GenAlEx also outputs the variance/covariance components, from which one can easily calculate the statistic by hand. Mantel matrix correlations are very widely useful for comparing different distance matrices.	MT	[29-30]
SE	Standard Error	$SE = \frac{s}{\sqrt{n}}$	>0	<i>SE</i> is the standard error of the mean and is widely reported by GenAlEx when the arithmetic mean and other summary statistics are reported. Where <i>s</i> is the standard deviation and <i>n</i> is the sample size.		

sHa	Shannon's <i>Allele Information</i> index	${}^sH_A = -\sum p_i \log_2 p_i$ ${}^sH_{A1} = -\sum p_{i1} \log_2 p_{i1}$ <p style="text-align: center;">and</p> ${}^sH_{A2} = -\sum p_{i2} \log_2 p_{i2}$	>0	One of the Shannon's Information Indices. For a specific locus in a given population the Shannon's Allele Information index is calculated by the general formula for ${}^sH_A$ . At each specific locus across multiple populations we consider each pairwise combination of populations in turn calculating ${}^sH_{A1}$ and ${}^sH_{A2}$ for each of the two populations. Where $p_i$ is the allele frequency of the $i$ th allele at the locus in question for the specified population (1 or 2). See Sherwin et al. [60], Tutorial 1 and Tutorial Appendix 1.1 for further details.	SH	[60]
sHu	Shannon's <i>Total Information</i> index	${}^sH_U = -\sum \bar{p}_i \log_2 \bar{p}_i$ $\bar{p}_i = p_{i1} \times wt_1 + p_{i2} \times wt_2$ $wt_1 = \frac{ct_1}{ct_1 + ct_2}$ $wt_2 = \frac{ct_2}{ct_1 + ct_2}$	>0	One of the Shannon's Information Indices. For a specific locus in a given population Shannon's Allele Total Information index is calculated by the formula for ${}^sH_U$ . Where $\bar{p}_i$ is the average weighted frequency of the $i$ th allele for each pair of populations with the weights calculated as shown. $ct$ is the total allele count at the locus for the respective populations. See Sherwin et al. [60], Tutorial 1 and Tutorial Appendix 1.1 for further details.	SH	[60]
sHua	Shannon's <i>Mutual Information</i> index	${}^sH_{UA} = {}^sH_U - wt_1 {}^sH_{A1} - wt_2 {}^sH_{A2}$ $G = 1.3863 {}^sH_{UA} (ct_1 + ct_2)$	>0	Shannon's <i>Mutual Information</i> index ${}^sH_{UA}$ is calculated for each pair of populations as shown. This statistic can then be used to compute the log-likelihood contingency test statistic $G$ . With degrees of freedom $DF$ calculated as the (number of populations compared - 1) x (number of alleles compared - 1). See Nm in this table for calculations of Nm via ${}^sH_{UA}$ , Tutorial 1 and Tutorial Appendix 1.1 for further details.	SH	[60]

t	Outcrossing rate (Codominant Data only)	$t = \frac{(1-F)}{(1+F)}$	0 to 1+	Not output in GenAlEx, but can easily be calculated from the <i>F</i> -values output by GenAlEx, using Excel functions. A useful transformation of the Fixation index for plant population that provides an estimate of the outcrossing rate. Assumes no selection between fertilisation and the stage at which the samples were analysed genetically.	via HFL, HFP	[69]
UHe	Unbiased Expected Heterozygosity or Genetic Diversity (Diploid Binary and Codominant Data only)	$UHe = \frac{2n}{2n-1} \left(1 - \sum p_i^2\right)$	0 to 1	Calculated on a single locus basis. GenAlEx also provides the arithmetic mean across loci. Here, <i>p<sub>i</sub></i> is the frequency of the <i>i</i> th allele and <i>n</i> is the sample size. See also Expected Heterozygosity.	HFL HFP	[1]
uh	Unbiased Haploid Genetic Diversity (Haploid Binary & Haploid Data)	$uh = \frac{n}{n-1} \left(1 - \sum p_i^2\right)$	0 to 1	Here, <i>p<sub>i</sub></i> is the frequency of the <i>i</i> th allele and <i>n</i> is the sample size. See also Haploid Genetic Diversity.	AFP AFL	[14]
UNei GD	Nei's Unbiased Genetic Distance	$UNei GD = -\ln(UI)$	> 0	<i>UI</i> is Nei's Unbiased Genetic Identity (see below). See also Nei's Genetic Identity. GenAlEx 6.3 provides a step-by-step option to illustrate this calculation.	UNeiP UNeiL UNeiT	[26]
UNei ID	Nei's Genetic Identity	$UNei ID = \frac{J_{xy}}{\sqrt{(uJ_x uJ_y)}}$ $J_{xy} = \sum_{i=1}^k p_{ix} p_{iy},$ $uJ_x = \frac{2n \left(\sum_{i=1}^k p_{ix}^2 - 1\right)}{2n - 1},$	0 to 1	Here, <i>p<sub>ix</sub></i> and <i>p<sub>iy</sub></i> are the frequencies of the <i>i</i> th allele in populations <i>x</i> and <i>y</i> . For multiple loci, <i>J<sub>xy</sub></i> , <i>uJ<sub>x</sub></i> and the analogous <i>uJ<sub>y</sub></i> are calculated by summing over all loci and alleles and dividing by the number of loci. These average values are then used to calculate <i>I</i> . Note that unlike Nei's Genetic Identity, the unbiased correction can yield slightly negative values. As recommended by Nei (1978), negative values are converted by GenAlEx to zero with associated warnings in subsequent outputs! GenAlEx 6.3 provides a step-by-step option to illustrate this calculation.	UNeiP UNeiL UNeiT	[26]

Ve	Expected variance (Haploid Data only)	$V_e = \sum h(1-h),$ $K = \sum h$		The expected variance of $K$ , where $K$ is the number of loci for which two individuals differ. In the absence of linkage disequilibrium, the expected variance is given by $V_e$ .	HDE	[23]
Vo	Observed variance (Haploid Data only)			The observed variance of $K$ . The disequilibrium index is $V_o/V_e$ . To test if this ratio is significantly greater than one, GenAlEx employs a randomisation test, following [23, 25]. For each random sample, $V_r$ is calculated as the variance of the randomized data set and the probability of observing a $V_r$ value as extreme as that observed ( $V_o$ ) is calculated.	HDE	[23]

## References

1. Hartl, DL and Clark, AG. 1997, *Principles of Population Genetics 3rd Ed.* Sunderland, Massachusetts: Sinauer Associates, Inc.
2. Hedrick, PW. 2005, *Genetics of Populations. Third Edition.* Sudbury, Massachusetts: Jones and Bartlett Publishers.
3. Hedrick, PW. 2009, *Genetics of Populations (4th Ed):* Jones and Bartlett Publishers.
4. Frankham, R, et al., *Introduction to Conservation Genetics.* 2002, Cambridge University Press: Cambridge.
5. Allendorf, FW and Luikart, G. 2006, *Conservation and Genetics of Populations:* Wiley-Blackwell. 642.
6. Hartl, DL. 2000, *A Primer of Population Genetics 3rd Ed.* Sunderland, Massachusetts: Sinauer Associates, Inc.
7. Conner, JK and Hartl, DL. 2004, *A Primer of Ecological Genetics.* Sunderland, Massachusetts: Sinauer Associates, Inc.
8. Frankham, R, et al. 2004, *A Primer of Conservation Genetics.* Cambridge: Cambridge University Press.
9. Weir, BS. 1990, *Genetic Data Analysis.* Sunderland, Massachusetts: Sinauer Associates, Inc.
10. Anon, A. 1996, *The Evaluation of Forensic DNA Evidence.* Washington, DC: National Academy Press.
11. Buckleton, J, et al. 2005, *Forensic DNA Evidence Interpretation.* New York: CRC Press.
12. Excoffier, L, et al. 1992. *Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction sites.* Genetics **131**, 479-491.
13. Huff, DR, et al. 1993. *RAPD variation within and among natural populations of outcrossing buffalograss *Buchloe dactyloides* (Nutt) Engelm.* Theoretical and Applied Genetics **86**, 927-934.
14. Peakall, R, et al. 1995. *Evolutionary implications of allozyme and RAPD Variation in diploid populations of dioecious buffalograss *Buchloe dactyloides*.* Molecular Ecology **4**, 135-147.
15. Michalakis, Y and Excoffier, L. 1996. *A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci.* Genetics **142**, 1061-1064.
16. Maguire, TL, et al. 2002. *Comparative analysis of genetic diversity in the mangrove species *Avicennia marina* (Forsk.) Vierh. (Avicenniaceae) detected by AFLPs and SSRs.* Theoretical and Applied Genetics **104**, 388-398.
17. Schneider, S, et al. 1997, *A software for population genetic data analysis.* University of Geneva, Switzerland: Genetics and Biometry Laboratory.
18. Wright, S. 1946. *Isolation by distance under diverse systems of mating.* Genetics **31**, 39-59.
19. Wright, S. 1951. *The genetical structure of populations.* Ann. Eugenics **15**, 323-354.
20. Wright, S. 1965. *The interpretation of population structure by F-Statistics with special regard to systems of mating.* Evol. **19**, 395-420.
21. Smouse, PE and Peakall, R. 1999. *Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure.* Heredity **82**, 561-573.
22. Slatkin, M. 1995. *A measure of population subdivision based on microsatellite allele frequencies.* Genetics **139**, 1463.
23. Gordon, DM. 1997. *The genetic structure of *Escherichia coli* populations in feral house mice.* Microbiology **143**, 2039-2046.
24. Brown, ADH, et al. 1980. *Multilocus structure of natural populations of *Hordeum spontaneum*.* Genetics **96**, 523-536.
25. Souza, V, et al. 1993. *Hierarchical analysis of linkage disequilibrium in *Rhizobium* populations: evidence for sex?* Proceedings of the National Academy of Sciences (USA) **89**, 8389-8393.

26. Hedrick, PW. 2000. *Genetics of Populations 2nd Ed.* Boston: Jones and Bartlett.
27. Raymond, M and Rousset, F. 1995. *Genepop (version 1.2) - population genetics software for exact tests and ecumenicism.* Journal of Heredity **86**, 248-249.
28. Mantel, N. 1967. *The detection of disease clustering and a generalized regression approach.* Cancer Res. **27**, 209-220.
29. Smouse, PE and Long, JC. 1992. *Matrix correlation analysis in anthropology and genetics.* Yearbook Phys. Anthropol. **35**, 187-213.
30. Smouse, PE, et al. 1986. *Multiple regression and correlation extensions of the Mantel test of matrix correspondence.* Systematic Zoology **35**, 627-632.
31. Nei, M. 1972. *Genetic distance between populations.* American Naturalist **106**, 283-392.
32. Nei, M. 1978. *Estimation of average heterozygosity and genetic distance from a small number of individuals.* Genetics **89**, 583-590.
33. Paetkau, D, et al. 1995. *Microsatellite analysis of population structure in canadian polar bears.* Molecular Ecology **4**, 347-354.
34. Paetkau, D, et al. 2004. *Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power.* Molecular Ecology **13**, 55-65.
35. Waser, PM and Strobeck, C. 1998. *Genetic signatures of interpopulation dispersal.* Trends in Ecology and Evolution **13**, 43-44.
36. Cornuet, JM, et al. 1999. *New methods employing multilocus genotypes to select or exclude populations as origins of individuals.* Genetics **153**, 1989-2000.
37. Piry, S, et al. 2004. *GENECLASS2: A software for genetic assignment and first-generation migrant detection.* Journal of Heredity **95**, 536-539.
38. Pritchard, JK, et al. 2000. *Inference of population structure using multilocus genotype data.* Genetics **155**, 945-959.
39. Orloci, L. 1978, *Multivariate analysis in vegetation research.* The Hague: Dr W. Junk B. V.
40. Lynch, M and Ritland, K. 1999. *Estimation of pairwise relatedness with molecular markers.* Genetics **152**, 1753-1766.
41. Ritland, K. 1996. *Estimators for pairwise relatedness and individual inbreeding coefficients.* Genetical Research **67**, 175-185.
42. Ritland, K. 2000. *Marker-inferred relatedness as a tool for detecting heritability in nature.* Molecular Ecology **9**, 1195-1204.
43. Queller, DC and Goodnight, KF. 1989. *Estimating relatedness using genetic markers.* Evolution **43**, 258-275.
44. Peakall, R and Sydes, MA. 1996, *Defining priorities for achieving practical outcomes from the genetic studies of rare plants*, in *Back from the Brink: refining the threatened species recovery process.*, S Stephens and S Maxwell, Editors. Surrey Beatty and Sons: Sydney.
45. Taberlet, P and Luikart, G. 1999. *Non-invasive genetic sampling and individual identification.* Biological Journal of the Linnean Society **68**, 41-55.
46. Waits, LP, et al. 2001. *Estimating the probability of identity among genotypes in natural populations: cautions and guidelines.* Molecular Ecology **10**, 249-256.
47. Peakall, R, et al. 2006. *Mark-recapture by genetic tagging reveals restricted movements by bush rats, Rattus fuscipes, in a fragmented landscape.* Journal of Zoology **268**, 207-216.
48. Ayres, KL and Overall, ADJ. 2004. *API-CALC 1.0: a computer program for calculating the average probability of identity allowing for substructure, inbreeding and the presence of close relatives.* Molecular Ecology Notes **4**, 315-318.
49. Jamieson, A and Taylor, SCS. 1997. *Comparisons of three probability formulae for parentage exclusion.* Animal Genetics **28**, 397-400.
50. Parks, JC and Werth, CR. 1993. *A study of spatial features of clones in a population of bracken fern, Pteridium aquilinum (Dennstaedtiaceae).* American Journal of Botany **80**, 537-544.

51. Stenberg, P, et al. 2003. *MLGsim: a program for detecting clones using a simulation approach*. Molecular Ecology Notes **3**, 329-331.
52. Sydes, MA and Peakall, R. 1998. *Extensive clonality in the endangered shrub Haloragodendron lucasii (Haloragaceae) revealed by allozymes and RAPDs*. Molecular Ecology **7**, 87-93.
53. Double, MC, et al. 2005. *Dispersal, philopatry and infidelity: dissecting local genetic structure in superb fairy-wrens (Malurus cyaneus)*. Evolution **59**, 625-635.
54. Peakall, R, et al. 2003. *Spatial autocorrelation analysis offers new insights into gene flow in the Australian bush rat, Rattus fuscipes*. Evolution **57**, 1182-1195.
55. Smouse, PE, et al. 2008. *A heterogeneity test for fine-scale genetic structure*. Molecular Ecology **17**, 3389-3400.
56. Gonzales, E, et al. 2010. *The impact of landscape disturbance on spatial genetic structure in the Guanacaste tree, Enterolobium cyclocarpum (Fabaceae)*. Journal of Heredity **101**, 133-143.
57. Beck, N, et al. 2008. *Social constraint and an absence of sex-biased dispersal drive fine-scale genetic structure in white-winged choughs*. Molecular Ecology **17**, 4346-4358.
58. Favre, L, Balloux, F., Goudet, J., and Perrin, N. 1997. *Female-biased dispersal in the monogamous mammal Crocidura russula: evidence from field data and microsatellite patterns*. Proceedings of the royal Society of London, Biological Series B **264**, 127-132.
59. Mossman, CA and Waser, PM. 1999. *Genetic detection of sex-biased dispersal*. Molecular Ecology **8**, 1063-1067.
60. Sherwin, W, et al. 2006. *Measurement of biological information with applications from genes to landscapes* Molecular Ecology **15**, 2857-2869.
61. Austerlitz, F and Smouse, PE. 2001. *Two-generation analysis of pollen flow across a landscape. II. Relation between  $\Phi_i(t)$ , pollen dispersal and interfemale distance*. Genetics **157**, 851-857.
62. Austerlitz, F and Smouse, PE. 2001. *Two-generation analysis of pollen flow across a landscape. III. Impact of adult population structure*. Genetical Research **78**, 271-280.
63. Austerlitz, F and Smouse, PE. 2002. *Two-generation analysis of pollen flow across a landscape. IV. Estimating the dispersal parameter*. Genetics **161**, 355-363.
64. Smouse, PE, et al. 2001. *Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females*. Evolution **55**, 260-271.
65. Dyer, RJ. 2005. *GENER: a server-based analysis of pollen pool structure*. Molecular Ecology Notes.
66. Lynch, M and Milligan, BG. 1994. *Analysis of population genetic structure with RAPD markers*. Mol. Ecol. **3**, 91-99.
67. Zhivotovsky, LA. 1999. *Estimating population structure in diploids with multilocus dominant DNA markers*. Molecular Ecology **8**, 907.
68. Slatkin, M. 1995. *A measure of population subdivision based on microsatellite allele frequencies (vol 139, pg 457, 1995)*. Genetics **139**, 1463.
69. Brown, AHD and Weir, BS. 1983, *Measuring genetic variability in plant populations*, in *Isozymes in Plant Genetics and Breeding, Part A.* , SD Tanksley and TJ Orton, Editors. Elsevier Science Publ.: Amsterdam. p. 219-239.
70. Davies, N, et al. 1999. *Determining the source of individuals: multilocus genotyping in nonequilibrium population genetics*. Trends in Ecology & Evolution **14**, 17-21.
71. Butler, JM. 2005, *Forensic DNA Typing: Biology, Technology and Genetics of STR Markers. 2nd Ed.* Oxford: Elsevier.