Appendix 1 – Methods and Statistics in GenAlEx 6.5

By Rod Peakall and Peter Smouse

Overview

This appendix summarizes the methods and statistics available in GenAlEx 6.502 (Peakall and Smouse [1, 2]), along with supporting references. With regard to the references, we point (where possible) to one or more texts in population genetics, although some procedures provided in GenAlEx are not yet covered in such texts. The texts we have found particularly useful for teaching include: *Principles of Population Genetics* (3rd Ed) by Hartl and Clark [3], *Genetics of Populations* by Hedrick [4, 5], *Introduction to Conservation Genetics* by Frankham et al. [6], and *Conservation and the Genetics of Populations* by Allendorf and Gordon [7]. A series of very useful primers include: *A Primer of Population Genetics* (3rd Ed) by Hartl [8], *A Primer of Ecological Genetics* by Conner and Hartl [9], and *A Primer of Conservation Genetics* by Frankham et al. [10]. Invaluable statistical population genetic resources for advanced users include *Genetic Data Analysis* by Weir [11, 12] *Handbook of Statistical Genetics* [13], *The Evaluation of Forensic DNA Evidence* by the National Research Council [14], and *Forensic DNA Evidence Interpretation* [15].

GenAlEx Tutorials – An Overview of Topics

A series of self-paced tutorials on population genetic analysis that employ hand calculations and exercises within GenAlEx are freely available from the web site. These are drawn from the graduate workshops that we have offered (jointly and independently), around the world. We strongly recommend that this appendix be read in conjunction with these tutorials.

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

2 - *Genetic Distance and AMOVA*: Haploid, Codominant and Binary Genetic Distance, AMOVA and F-statistics

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

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

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

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

7 – *Hierarchical Shannon Diversity Analysis:* Introduces the new Shannon Diversity Partition options of Smouse et al (2015) as released in GenAlEx 6.502 in September 2015.

8 – *Troubleshooting:* A handy tutorial on the tools offered within GenAlEx for troubleshooting data files that fail to run.

Table 1: An overview of methods in GenAlEx 6.5

Methods	Overview	GenAlEx Option
AMOVA	The Analysis of Molecular Variance (AMOVA) procedure follows the methods of Excoffier et al. [16], Huff et al. [17], Peakall et al. [18], and Michalakis and Excoffier [19]. AMOVA 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 was a key development in population genetic analysis, because it allowed (for the first time) hierarchical population structure analysis for all types of genetic markers (haploid, haplotype, binary, codominant and sequence), as well as offering statistical testing options by random permutation. There have been a number of later variations on this same theme. Within GenAlEx, the data type and choice of distance calculation used as input for AMOVA lead to related but different analyses. To estimate F_{ST} for codominant data, choose the *Codom-Allelic* distance. To estimate R_{ST} for microsatellites, choose *Codom-Microsat*. Note, R_{ST} should not be your default option for microsatellites, because the underlying assumptions of simple step-wise mutation rarely hold in natural populations. Consequently, estimation of R_{ST} is often not useful. To suppress within population variance and simply calculate population differentiation based on the genotypic variance, choose *Codom-Genotypic* distance. This option produces an estimate of Φ_{PT}, an analogue of F_{ST}. Φ_{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 and other markers (such as binary AFLPs), Φ_{PT} should be used for all markers (see [18] and [20] for comparative studies) In all other cases involving codominant data, we recommend reporting F_{ST} rather than Φ_{PT}. GenAlEx 6.5 offers new AMOVA routines to estimate standardized F'_{ST}, following Meirmans [21], and we suggest this statistic should also be reported in all studies. The software packages Arlequin [22] and Genodive [23] also offer AMOVA. GenAlEx offers data export to both packages. See also, *F*-statistics, *G*-statistics and Genetic Distance below, and Tutorial 2.	AMOVA
AMOVA and Statistical Tests	A typical null hypothesis in biology is '*No Difference*'. For AMOVA: H_0 = No genetic difference among populations (F_{ST} = 0 or R_{ST} = 0 or Φ_{PT} = 0), H_1 = There is genetic difference among populations (F_{ST} >0 or R_{ST} >0 or Φpt > 0). Thus, for AMOVA, under H_0 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 a single gene pool. Thus, we should find little difference (other than minor sampling effects) between arbitrary subpopulations. It follows that if we shuffle (randomize) the samples in our data set, and calculate AMOVA for each shuffle, we should obtain values of the same magnitude as expected by random sampling from a single population. Because of sampling effects, the results will naturally vary from shuffle to shuffle. Moreover, if we perform multiple shuffles (say 999 or 9999 times), we can construct a good estimate of the range of values we would expect if the null hypothesis were true. This is the rationale for statistical testing by random permutation when performing AMOVA in GenAlEx. To determine whether 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 the results significant at the 5% level. Note, that in calculating the probability value *P*, GenAlEx always includes the observed value as 'just another permutation' adding this value to the 99, 999 or 9999 permutations. *P* is calculated as the *Number of Values ≥ Observed Value (Including Observed Value)* ÷ (*Number of Permutations + 1*). As a consequence, the smallest probability value *P* reported by GenAlEx will never be less than *1* ÷ (*Number of Permutations + 1*). Thus, for 99 permutations, the smallest *P* value will be 0.01; for 999, the smallest *P* value will be 0.001, etc. In GenAlEx, *P* values are reported with the caption *P(rand>=data)*, 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 GenAlEx, not just AMOVA. (In addition, such as for autocorrelation analysis, GenAlEx also offers tests for *P(rand<=data)* when the statistic in question can be negative). For regional analyses, GenAlEx offers two types of permutation: *Standard* and *Specialized*. Standard shuffles individuals across populations and regions, whereas the specialized option varies by statistic: To calculate the probability for F_{IS}, individuals are shuffled within populations, while for the probability of F_{SR}/*PhiPR* individuals are shuffled within regions. For F_{RT}/*PhiRT* whole populations are shuffled among regions to estimate the probability. The outcomes of standard and specialized permute are listed side-by-side to allow comparison. Note that when there are very few populations and regions, estimates of the probability for F_{RT}/*PhiRT* via specialized permute should be treated with caution, since there are very few different combinations to shuffle. Note that in Arlequin [22] the default permutational tests are equivalent to the specialized option of GenAlEx, but reported *P* values can equal zero (which may seem counterintuitive). Please refer to the Arlequin guide for more details.	AMOVA

F-statistics and G-statistics	Wright's *F*-statistics [24-26] are widely used to characterize population genetic structure. These statistics allow the partitioning of genetic diversity (~ heterozygosity) within and among populations. GenAlEx provides three pathways for the calculation of *F*-statistics: *Frequency, G-statistics* and *AMOVA*. The frequency based calculation of *F*-statistics follow [3]. 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 *F*-statistics by hand. In line with common usage in textbooks such as [3, 4, 27], F_{ST} is here calculated by Nei's [28] formula for G_{ST}, which represents a multiallelic expansion of Wrights F_{ST} [29]. When calculated in this fashion, F_{ST} and G_{ST} are frequently used interchangeably (e.g. [30]). For research purposes, the calculation of *F*-statistics via either *AMOVA* or *G-statistics* is recommended, since both allow for statistical testing. In GenAlEx 6.5 onwards, the *G-statistics* option offers a comprehensive range of new standardised estimators of genetic differentiation, including G'_{ST}, G''_{ST} and Jost's D_{est}, following the recommendations and formulae of Meirmans and Hedrick [29]. In line with [29], GenAlEx applies the corrections of Nei and Cheeser [30] in the calculations of H_T and H_S, denoted as cH_S and cH_T in the formulae that use these corrections (see below). In calculating G-statistics across multiple loci, cH_S and cH_T are first averaged over loci. In GenAlEx 6.5 onwards, we retain the notation F_{ST} when calculated without these corrections (via *Frequency* option), whereas we use the notation G_{ST} (via *G-statistics* option) when the above corrections are applied. The notation F_{ST} is also used when estimating genetic differentiation via AMOVA when allelic distances are used as the input for codominant data. New AMOVA routines also enable the estimation of standardized F'_{ST}, following Meirmans [21]. Note that in GenAlEx 6.1 onwards, we provided a minor modification of the *F*-statistics routine in AMOVA that brings the estimates for F_{ST} in line with the Weir-Cockerham estimates, following formulas and notation in Peakall et al. [18]. See also AMOVA above, relevant formulas in Table 2 and Tutorial Part 1.	Frequency, G-statistics AMOVA
Genetic Distance (Binary)	A pairwise, individual-by-individual (*N x N*) 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. [17], 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 GenAlEx for subsequent PCA, Mantel and all Spatial analyses involving binary data. This distance option is also be used to calculate Φ_{PT} via AMOVA, a measure of population genetic differentiation for binary data that is analogous to *Fst*. This is a Euclidean distance metric, unlike binary measures such as Nei's (1 – *F*), and is therefore appropriate for AMOVA, which requires a Euclidean metric [16-18]. 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.	Distance->Genetic, AMOVA
Genetic Distance (Codom-Genotypic)	A pairwise, individual-by-individual (*N x N*) genetic distance matrix is calculated for codominant data by this genetic distance option. For a single-locus analysis, with *ith*, *j*-th, *k*-th and *l*-th different alleles, a set of squared distances is defined as $d^2(ii, ii) = 0$, $d^2(ij, ij) = 0$, $d^2(ii, ij) = 1$, $d^2(ij, ik) = 1$, $d^2(ij, kl) = 2$, $d^2(ii, jk) = 3$, and $d^2(ii, jj) = 4$. See [18] and Smouse and Peakall [31] for graphical explanation of this method. This is the most important genetic distance option for codominant data, since the matrix generated is used in GenAlEx for subsequent PCA, Mantel and all Spatial analyses. This distance option can also be used to calculate Φ_{PT} 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 [20]), where no intra-individual variation (heterozygosity) is available. See also AMOVA above, Φ_{PT} in Table 2 and Tutorial 2.	Distance->Genetic, AMOVA
Genetic Distance (Codom Allelic)	This *Codom-Allelic* option generates a 2*N x* 2*N* 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 GenAlEx is calculating *Fst* 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, *Fst* in Table 2 and Tutorial 2.	AMOVA

Genetic Distance (Codom Microsat)	This *Codom-Microsat* distance option produces a 2*N x* 2*N* 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: $(S1 – S2)^2$, where *S*1 is the size of allele 1 and *S*2 the size of allele 2. Distances are summed across loci. This genetic distance option can only be generated when GenAlEx is calculating R_{ST} via AMOVA. R_{ST} is an estimator of genetic differentiation for microsatellite loci that assumes a stepwise mutation model [19, 32]. 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, R_{ST} in Table 2 and Tutorial 2.	AMOVA
Genetic Distance (Haploid)	A pairwise, individual-by-individual (*N x N*) 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 GenAlEx for subsequent PCA, Mantel and all Spatial analyses involving haploid data. This distance option is also used with haploid data to calculate Φ_{PT} via AMOVA, a measure of population genetic differentiation that is analogous to F_{ST}.	Distance->Genetic, AMOVA
Genetic Distance (Haploid SSR)	A pairwise, individual-by-individual (*N x N*) 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: $(S1 – S2)^2$, where *S1* is the size of allele 1 and *S*2 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 R_{ST}. Consequently, studies using this genetic distance estimate should also report the outcomes using the standard Haploid genetic distance (see also related comments on R_{ST} in Table 2). Provided users are mindful of the underlying assumptions and risks of spurious patterns, this genetic distance matrix can be used for subsequent PCA, Mantel and all Spatial analyses involving haploid-SSR data. Note that the Φ_{PT} generated via AMOVA with this option provides a measure of population genetic differentiation that is analogous to R_{ST}. However, the current outputs from GenAlEx do not note this connection. That is, AMOVA outputs from both Haploid and Haploid-SSR are called Φ_{PT}, unlike their codominant counterparts. This may change in future versions of GenAlEx.	Distance->Genetic, AMOVA
Geographic Distance	A pairwise, individual-by-individual (*N x N*) linear geographic distance matrix is generated from X and Y coordinates by this distance option. See also GGD below for background to the calculation of distances from Latitude and Longitude.	Distance->Geographic
HWE – Tests for Hardy-Weinberg Equilibrium (Codom Data)	The HWE procedures follow Hedrick [27], but is similar to many texts [e.g., 1-6]. For codominant genotypes at a single locus, and for a single population, we can determine whether observed tallies of genotypes are consistent with expectations, as follows: 1. Determine the number of samples, 2. Determine the number of alleles, *Na*. 3. Count the numbers of each genotype. 4. Calculate allele frequencies. 5. Estimate the expected genotype frequencies, given the sample size of the population. [p^2 homozygotes, $2pq$ for heterozygotes 6. Test for conformity with HWE expectations by calculating X^2. 7. Determine the degrees of freedom. 8. Given the calculated Chi-squared value and the degrees of freedom, determine 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 GenAlEx is provided primarily for teaching purposes and for data exploration. An alternative statistical test for assessing an overall departure from random mating expectations is provided in GenAlEx via the AMOVA framework, where permutational tests provide an assessment of whether or not the inbreeding coefficient F_{IS} is equal to zero. Other programs such as GenePop [33] and Arlequin [22] provide Exact Tests which are recommended for research purposes (but note there remain some technical issues when employing Exact Tests [34]). GenAlEx offers data export to these programs and other relevant programs. See also ChiSquare in Table 2 and Tutorial 1.	HWE

Linkage Disequil. (Codom Biallelic)	Despite its importance, there is no universal test for linkage disequilibrium [35]. GenAlEx 6.5 offers pairwise tests for disequilibrium between biallelic markers such as SNPs. When phase is known, this includes the calculation of *D, D’, r, and* r^2*,* following Hedrick [4]. Maximum likelihood estimation is used to calculate *D* and *r* when phase is unknown (Weir [11], p. 310). For large SNP sets, or multiallelic data, GenAlEx users are encouraged to take advantage of the options to export their data to other packages such as Arlequin 3.5 [22]. Note that GenAlEx 6.5 requires all biallelic data for pairwise linkage disequilibrium analysis to be coded with alleles as either ‘1’ or ‘2’. GenAlEx provides a set of tools for recoding genetic data under the *Edit Raw Data* menu, if recoding is required. In regards to the data format required for biallelic data of known phase, consider the case of Locus A and Locus B with alleles coded as either 1 or 2 for both loci. The maternal gametic haplotype is specified by the allele in the first column of locus A and the first column of locus B, while the paternal gamete haplotype (or vice versa) is represent in the second column for each locus. For example, the two locus genotype ‘1 2 1 2’ represents gametic haplotypes 11 and 22, while the two locus genotype ‘2 1 2 1’ represents gametes 22 and 11. Note that for data of known phase, the genotype ‘2 1’ at locus A does represent a valid GenAlEx data format, although for data of unknown phase it is usual to present genotypes with alleles sorted (e.g. 1 2, not 2 1). Note that if exporting GenAlEx data of known phase to Arlequin, the haplotype phases are correctly retained, however, you will need to manually specify in the exported Arlequin data header that the phase is known.	Disequil
Linkage Disequil. (Haploid)	The test for haploid disequilibrium follows Gordon [36], based on the theory developed by Brown et al. [37] and Souza et al. [38]. The index of linkage disequilibrium is *Vo/Ve*, where *Ve* 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 *Ve*. To test whether the ratio of *Vo/Ve* is significantly greater than one, GenAlEx employs a randomisation test, following [36, 38]. *Vr* is calculated for each random sample as the variance of the randomized data set and the probability of observing a *Vr* value as extreme as that measured for the original data (*Vo*). This option is only available in GenAlEx when a haploid data set consists of a single population.	Frequency
Mantel Tests	Mantel tests for Matrix Correspondence [39] follow the methods of Smouse and co-workers [40, 41], 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., [18, 20]). While it is easy to plot a graph of the relationship between elements from any two matrices, we cannot use the *P*-values of standard regression analysis, because the *N x* (*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 [42, 43] 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 [42]. Both the biased and unbiased estimates of Nei’s genetic distance [43] are offered. 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 [44, 45] is available within GenAlEx. See also [46, 47] 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 the population with the smallest value. We recommend the default *Leave one out* option, which uses an allele frequency estimate that leaves the sample to be assigned out of the frequency estimate. The *As is* option is primarily provided for teaching purposes and for compatibility with the *Sex Bias* option in GenAlEx. For research purposes, the program *GeneClass 2* [48] is recommended. The program *Structure* [49] provides alternative statistical options for population assignment, using different methods from *GeneClass*. GenAlEx provides data export options to both these programs. See also Tutorial 4.	Assignment->Pop Assign
Principal Coordinates Analysis	Principal Coordinate Analysis (PCoA) 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, PCoA 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 two or three axes will typically reveal most of the separation among them. The procedure in GenAlEx is based on an algorithm published by Orloci [50]. 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. Note that in GenAlEx 6.5 onwards, we have changed the previous notation of PCA to PCoA, to bring the notation into line with common usage for Principal Coordinates Analysis. There are no changes to the procedure itself.	PCoA
Phi Statistics	The estimation of ***Φ**-s*tatistics parallels the same logic as for *F*-statistics for codominant data, except that ***Φ**-s*tatistics are also estimable from binary and haploid data [18]. See AMOVA above. Formulas for the various ***Φ**-s*tatistics 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) [51-54]. For a summary of the formulas, see Ritland [53]. 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 [55] as a indication of the statistical power of a specific set of marker loci. This is also used for genetic tagging in molecular ecology [56, 57], 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 [56, 58], 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 [59].	Multilocus->Prob. Ident.
Probability of Exclusion	GenAlEx offers the calculation of three probability estimates for parentage exclusion, following Jamieson and Taylor [60]. 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 GenAlEx, following [61-63]. Formulae for these probability estimates are provided in Table 2. GenAlEx 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 GenAlEx. 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 [62].	Clonal->Prob. Clone

Spatial Auto-correlation	GenAlEx provides an extensive series of spatial autocorrelation analyses, following the methods of Smouse and co-workers [31, 64-68]. GenAlEx 6.2 onwards provides users with access to new spatial genetic analysis procedures developed and described in Smouse et al., Beck et al. and Gonzales et al. [66-68]. From GenAlEx 6.5 onwards, the Heterogeneity test procedure of Smouse et al. [66] is offered within the standard spatial menu options (thus the 'Adv Spatial' menu options are no longer required). Banks and Peakall [69] employed simulations to assess statistical power and the Type I and Type II error rates of the Heterogeneity Test. They found that under some circumstances, Type I error rates maybe inflated above 5%, leading them to suggest that a more stringent critical cut-off of 1% should be applied. Consequently, GenAlEx 6.5 onwards only declares significance of the Heterogeneity Tests at the 1% cut-off. GenAlEx users are strongly advised to complete the Tutorial 3 before conducting spatial genetic analysis. Banks and Peakall [69] also offer important advice and recommendations for spatial analysis.	Spatial
Spatial Auto-correlation 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, GenAlEx 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 1000th permutation), the values of the 25th and 975th 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, GenAlEx 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}) 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 $n(n\text{-}1)/2$ 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
Spatial Auto-correlation and Heterogeneity Testing	From GenAlEx 6.5 onwards, the Heterogeneity test procedure of Smouse et al. [66] is offered within the standard spatial menu options. In the case of a single population spatial analysis, the *Omega* value and probability are provided above the spatial correlogram graph. In this case, the heterogeneity test provides a test of correlogram significance. More often, applications of the heterogeneity tests will be of more interest when you have suitable data from two or more populations. This test is also useful for assessing difference in spatial genetic structure patterns between sexes (here sexes are treated as different populations for the purpose of the analysis). See Banks and Peakall [69] for a comprehensive overview of how spatial autocorrelation analysis can be applied to detect sex-biased dispersal. To illustrate how this test works, assume we are testing for sex-biased dispersal. After standard spatial autocorrelation analysis of males and females as separate populations, GenAlEx computes the nonparametric heterogeneity test as follows: First, the pooled within-population autocorrelation (across both sexes) is estimated, representing the base autocorrelation levels under the null hypothesis of no difference between the sexes. Next, the distribution of random departure from this average is tested by bootstrap resampling. The bootstrapping is achieved by randomly drawing paired samples from across the two populations, but maintaining the original samples sizes within each distance class. Next a squared paired-sample t-test statistic *T2* for each distance class is computed to evaluate the upper tail probability that the observed *T2* value is larger than expected under the null hypothesis. In the final step, the two correlograms are compared, drawing on the *P* values for the *T2* statistic at each distance class, across both sexes (populations), to compute the correlogram wide *Omega*. Finally, the probability that observed *Omega* is larger than expected under the null hypothesis of homogeneous correlograms is determined. The null hypothesis for this test predicts homogeneity between the spatial correlograms of the two sexes, while the alternative hypothesis predicts heterogeneity. See Tutorial 3 for further details.	Spatial

Sex Biased Dispersal	GenAlEx implements the sex-biased assignment test procedure developed by Favre et al. [70] and extended by Mossman and Waser [71]. For each individual, GenAlEx calculates a log likelihood assignment test value, as described under population assignment, except that the *As is* allele frequency estimate is used instead of the *Leave one* out option. Next, an Assignment Index correction (*AIc*) for each individual is calculated as: Individual (log likelihood – mean log likelihood of the population). AIc values will average zero for each population, while negative values will characterize individuals with a higher probability of being immigrants. In GenAlEx a plot of the mean *AIc* for males versus females is provided, as well as a plot of the frequency distribution of corrected assignment indices (*AIc*) for the males and females. The genetic signal of sex-biased dispersal is indicated when there is a difference in the frequency distribution of *AIc* values among males and females. Note that GenAlEx does not yet offer the recommended non-parametric test of this difference.	Assignment->Sex Bias
Shannon Pairwise	Shannon information indices have been widely employed in ecology but largely overlooked in genetics. Sherwin et al [72]. assessed the performance, power and theoretical expectation of Shannon indices for estimating genetic diversity. They concluded that the Shannon information framework offers an alternative method of quantifying biological diversity across multiple scales (genes to landscapes). GenAlEx 6.3 onwards offers the calculation of a series of pairwise population Shannon indices, including the mutual information index ${}^{S}H_{\mathrm{UA}}$, an alternative estimator of population structure, and locus-by-locus *G*-tests of mutual information following Sherwin et al [72]. Note that while we list the G-test and chi-square probability values, statistical testing by random permutation is recommended, because the *G*-test may exhibit high type I error rates (false rejection of the null hypothesis). Test via permutation are offered via the Shannon Partition option. 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.	Shannon->Pairwise
Shannon Diversity Partition	GenAlEx 6.5 offers a new Shannon Diversity Partition option that extends Shannon Indices to multiple hierarchical levels, following [73] Smouse and Ward (1978), with updates in 6.502 following Smouse et al. [74]. The Diversity Partition option allows estimates of *Alpha*, *Gamma* and *Beta Diversity*, as well as [0,1] Scaled *Divergence* and *Overlap*. A unique three level partition option for apportioning diversity among regions, among populations, and within populations is presented in the Shannon Statistics Summary Table, which is analogous to an AMOVA Summary Table. Although, traditional statistical testing is by means of the log-likelihood ratio G-test, which is approximately chi-square-distributed for large sample sizes, here, we offer an alternative (random permutation) test. If G-test and chi-square probability values are required they can be obtained via the Shannon Pairwise option that implements the closely related methods of Sherwin et al. [72] (see above).	Shannon->Diversity Partition
TwoGener	Two generational analysis of pollen flow following [75-78] is provided for codominant data. Note that GenAlEx 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 [79], or the program POLDISP. Data export to these software packages is provided by GenAlEx 6.3 onwards. For details see Tutorial 6. Note that first time users of TwoGener will need to enable the TwoGener menu via the Options-> Menus.	TwoGener

Table 2: A summary of the statistics used in GenAlEx 6.5

GenAlEx Notation	Measure	Formula	Range	Notes	GenAlEx Worksheet	Ref
	Allele Frequency (Codom Data)	$F_x = \frac{2N_{xx} + N_{xy}}{2N}$	[0,1]	Calculated for a single locus. Determined for each allele. *Nxx* = # of XX homozygous individuals, and *Nxy* = # 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	[3, 6, 8-10, 27]
	Allele Frequency (Binary Data)	Assuming random mating: Presence = *AA* or *Aa* Absence = *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 = √(Freq. of absence)	[0,1]	With dominant binary markers (e.g. AFLPs), it is not possible to directly calculate allele frequencies. If, 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 GenAlEx estimate is shown to the left. Following Lynch and Milligan [80], it assumes complete outcrossing, but does not impose the recommended pruning of low frequency bands. Note: Zhivotovsky [81] has developed an alternative Bayesian allele frequency estimation procedure that is available in other programs such as FAMD. GenAlEx 6.3 onwards offers data export to this program.	BAFL BAFP BAPT	[80]
	Allele Frequency (Haploid Data)	$F_x = \frac{N_x}{N}$	[0,1]	Calculated for single loci, and determined for each allele, where *Nx* = number of the *x* alleles and *N* = the number of samples. Can also be determined by direct count of the proportion of different alleles.	HAFL HAFP HAPT	
AIc	Assignment Index	Individual log-likelihood – mean log-likelihood of the population.		See also Sex Biased Dispersal in Table 1	SB, FDSB	[70]
Chi	Chi-Squared Test for HWE	$X^2 = \sum_{i=1}^{k} \frac{(O - E)^2}{E}$	[0,∞]	Where O_i is the observed number of individuals of the *i*th genotype, and E_i the expected number with *DF* = [*Na*(*Na*-1)]/2, where *Na* is the number of alleles at the locus. See also HWE in Table 1	HW HWS	[27]
D	Linkage Disequilibrium (Phase Known) (Biallelic Codom Data)	$D = x_{11} - p_1 q_1$ $D = x_{11}x_{22} - x_{12}x_{21}$ $r^2 = \frac{D^2}{p_1 p_2 q_1 q_2}$	[-1,+1] [0,1] [-1,+1]	When phase is known *D, D'*, *r, and* r^2 are calculated for pairs of biallelic loci, following [4]. Here with alleles at each locus coded as 1 or 2, x_{11} is the observed frequency of the gametic haplotype (11) [Allele 1 both at locus A and locus B], with expected frequency p_1q_1 where p_1 is the observed frequency of allele 1 at locus A, and q_1 the observed frequency of allele 1 at locus B. *D* is calculated as the observed minus the expected frequency of the gametic haplotype, for each combination of the 4 gametic haplotypes (11, 12, 21, 22). The square of the correlation coefficient (r^2) is a transformation of *D*. The correlation coefficient (*r*) is calculated as the square root of r^2, with the same sign as *D*.	LDK	[4]

D or cD	Linkage Disequilibrium Phase Unknown (Biallelic Codom Data)	See Weir [11] for details of the estimation of *D, cD*. $r = \frac{D}{\sqrt{p_1 p_2 q_1 q_2}}$	[-1,+1] [-1,+1]	For the more usual case of biallelic data of unknown phase, the maximum likelihood method of Weir [11] is used to estimate *D* and *r.* Two different methods for estimating *D* are employed, the first method assumes HWE, the second with outcome denoted *cD*, is estimated via the composite disequilibrium coefficient approach. These calculations are achieved in GenAlEx via the conversion of Weir’s (1990, pp. 310) LD79 Fortran program to VBA. Validation of the disequilibrium tests was made against outcomes in the program GDA [82] that implements the same analysis. The correlation coefficient *r* is calculated for the first method of *D* estimation, as shown.	LDU	[11]
D’	Standardized Linkage Disequilibrium	$D' = \frac{D}{D_{\max}}$	[-1,+1]	*D’* represents the standardized *D* value. If *D* > 0, D_{max} = min(p_1q_2 , p_2q_1). For *D* < 0, D_{max} = min(p_1q_1, or p_2q_2). See [4]. See also r and rSq.	LDK	[4]
Diversity D’	[0,1] Scaled Diversity D’	(1-(1/Beta Diversity))/(1-(1/DivWt))	[0,1]	[0,1] Scaled Diversity calculated via Shannon->Diversity Partition. See also *sH*, *sH(WP)*, *sH(GT)*, *sH(AP).*	SHT SH	[74]
DivWt	Weighted Diversity	$\exp\left(-wt_1 \ln(wt_1) - wt_2 \ln(wt_2)\right)$		Weighted Diversity used in the calculation of [0,1] Scaled Diversity via Shannon->Diversity Partition. See also *sH*, *sH(WP)*, *sH(GT)*, *sH(AP).*	SHT SH	[74]
Dest	Jost’s *D* via G-Statistics	$D_{est} = \left(\frac{k}{k-1}\right)\left(\frac{{}_cH_T - {}_cH_S}{1 - {}_cH_s}\right)$	[0,1]	Here, Jost’s estimate of differentiation (*Dest*) [83] is calculated following Meirmans and Hedrick eq 2.[29]. Their recommendation to average cH_S and cH_T for estimating *Dest* across loci is also used. See H_S, H_T and G_{ST} below for further details. Note that some software packages estimate *Dest* over loci as the harmonic mean of individual locus *Dest* values [29].	Gst GstG GstS DestP	[29]
e^sH(WP)	Alpha Diversity	*exp(sH(WP))*	[>0]	See *sH(WP)* via Shannon-> Diversity Partition	SHT	[74]
e^sH(GT)	Gamma Diversity	*exp(sH(GT))*	[>0]	See *sH(GT)* via Shannon->Diversity Partition	SHT	[74]
e^sH(AP)	Beta Diversity	*exp(sH(AP))*	[>0]	See *sH(AP)* via Shannon->Diversity Partition	SHT	[74]
F	Fixation Index (Codom Data)	$F = \frac{H_E - H_O}{H_E}$	[–1,1]	Calculated on a per locus basis. GenAlEx 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	[3]
Fis	F_{IS} via Frequency F_{IS} via G-Statistics (Codom Data)	$F_{IS} = 1 - \frac{H_O}{H_S}$	[-1,+1]	The inbreeding coefficient within individuals, relative to the population. F_{IS} measures the reduction in heterozygosity, due to non- random mating within each subpopulation. Note that the notation (F_{IS}) for subpopulations is equivalent to (F_{IP}) for populations in GenAlEx. Note that G_{IS} offered via G-statistics is closely related, differing only by a bias correction for H_S.	HFL HFP Gst GstG GstS	[3]

Fst	F_{ST} via Frequency F_{ST} via G-Statistics (Codom Data)	$F_{ST}=\frac{H_T-\bar{H}_e}{H_T}$ $F_{ST}=\frac{H_T-H_S}{H_T}$ $G_{ST}=\frac{H_T-H_S}{H_T}$	[<0,1] [<0,1] [<0,1]	The inbreeding coefficient within subpopulations, relative to the total. F_{ST} provides a measure of the genetic differentiation among populations. That is, the proportion of the total genetic divergence that separates the populations. F_{ST} is typically greater than or equal to zero (but can be slightly negative hence the bounds are shown as [<0,1]. If all subpopulations are in Hardy-Weinberg equilibrium with the same allele frequencies, $F_{ST} \approx 0$. (Note that the *s* used for subpopulations in the notation for F_{ST} is equivalent to F_{PT} for populations in GenAlEx). Although ranging between 0 and 1 for biallelic data, the max F_{ST} decreases quickly with increasing numbers of alleles. This property has led to the development of new standardized estimators, offered in GenAlEx 6.5 onwards via *G-statistics* and AMOVA. See also Gst.	Fst FstP FstL FstT Gst GstG GstS	[3, 6, 10]
Fst	Linearized F_{ST}	$LinF_{ST}=\frac{F_{ST}}{(1-F_{ST})}$		A transformation for pairwise population Fst values recommended by Slatkin [84]. GenAlEx also offers the analogous transformation for PhiPT.	LinFst	[84]
F'st	Standarized F_{ST} via AMOVA		[<0,1]	Calculated via AMOVA, F'_{ST} is analogous to G'_{ST}, and is a [0,1]-scaled estimator of differentiation, following [21]. GenAlEx also offers calculation of the analogous Φ'_{ST} via AMOVA for haploid data. See [29] for additional background on standardized *F*- and *G*-statistics. See [21] and Tutorial Part 7 for details on the calculation of standardized *F*- and *Φ*statistics via AMOVA.	Fst FstP FstL FstT	[21, 29]
F'rt	Standarized F_{RT} via regional AMOVA		[0,1]	A [0,1]-scaled estimator of differentiation among regions, as estimated via a regional AMOVA of codominant genetic distance, as described by [21]. Analogous to Φ'_{RT} for haploid data.	Fst FstP FstL FstT	[21, 29]
F'sr	Standarized F_{SR} via regional AMOVA		[<0,1]	A [0,1]-scaled estimator of differentiation among populations within regions, estimated via a regional AMOVA of codominant genetic distance [17]. Analogous to Φ'_{PR} for haploid data.	Fst FstP FstL FstT	[21, 29]
G	Log-likelihood *G*-statistic Via Shannon-Partition	$G = 2\,{}^{S}H_{UA}\left(ct_1+ct_2\right)$	[0,n]	The formula shown for *G* applies when the natural logarithm is used in the calculation of ${}^{S}H_A$ and ${}^{S}H_U$. For research purposes, statistical testing by random permutation is recommended, because there are reports that the log-likelihood *G*-test and associated chi-square probabilities may have elevated type I error rates (false rejection of the null hypothesis). GenAlEx 6.5 onwards, offers testing by random permutation via Shannon-Partition option.	SHa SHuaP SHuaL SHT SH	[72]
GD	Genetic Distance Binary	$D = n\left[1-\frac{2n_{xy}}{2n}\right]$	[0,n]	Here, 2*nxy* = number of shared character states, *n* = 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	[17, 20]

GGD	Geographic Distance	$D = \sqrt{(xi - xj)^2 + (yi - yj)^2}$	[0,n]	Here, *xi and yi* are the coordinates for the *i*th sample and *xj* and *yj* are the coordinates for the *j*-th sample.	GGD	
GGD	Geographic Distance (Via Lat/Long)		[0,n]	GenAlEx uses a modification of the Haversine Formula developed by R.W. Sinnott (Virtues of the Haversine (1984) *Sky and Telescope* 68,159) following computer code published online by Bob Chamberlain from JPL, NASA. (http://www.usenet-replayer.com/faq/comp.infosystems.gis.html still available on 12/12/12). Distances calculated via Lat/Long coordinators are returned in km.	GGD	
Gis	G_{IS} via G-statistics (Codom Data)	$G_{IS} = 1 - \frac{H_O}{{}_cH_S}$	[-1,+1]	Analogous to F_{IS}, except that G_{IS} is calculated with the correction of Nei and Cheeser [30] for small population size and inbreeding, applied in the calculation of cH_S. Note that Nei [85] p. 164], among others, uses the notation F_{IS} for this formula. However, for consistency with textbooks, we retain that notation F_{IS} for the uncorrected estimate and use G_{IS} here inline with our notation for G-statistics generally. See also cHs and Fis.	Gst GstG GstS	[29, 85]
GP	Genotype Probability (Codom Data)	$\mathrm{GP} = \prod \mathrm{p}_\mathrm{i}^2 \mathrm{x} \prod 2\mathrm{p}_\mathrm{i}\mathrm{p}_\mathrm{j}$	[0,1]	Π indicates chain multiplication across each locus, *pi* is the frequency of the allele at homozygous loci, *pi* and *pj* are the frequencies of alleles at heterozygous loci. Also called *DNA Profile Probability* and *Random Match Probability*, 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]
Gst	G_{ST} via G-statistics (Codom Data)	$G_{ST} = \frac{{}_cH_T - {}_cH_S}{{}_cH_T}$	[<0,1]	G_{ST} is calculated following Meirmans and Hedrick [29] with the correction of Nei and Cheeser [30] and Nei [85] for small population size and inbreeding applied in the calculations of H_T and H_S. The notation cH_S and cH_T is used to indicate these corrections. In calculating multi-locus average *G*-statistics, cH_S and cH_T are averaged over loci, before use in the formula. See also *F-statistics* and Fst on the exchangeability of F_{ST} and G_{ST} and the notation applied in GenAlEx 6.5 onwards.	Gst GstG GstS	[29, 85]
GstM Gst max	$G_{ST(max)}$ via G-Statistics (Codom Data)	$G_{ST\,(\max)} = \frac{(k-1)(1-{}_cH_S)}{k-1+{}_cH_S}$	[<0,1]	$G_{ST(max)}$ provides an estimate of the maximum possible G_{ST} , given the data over all *k* populations. It is used in the calculation of $G'_{ST(Hed)}$.	Gst GstG GstS	[29]
G'stH G'st (Hed)	G'_{ST} Hedrick via G-Statistics (Codom Data)	$G'_{ST\,(Hed)} = \frac{G_{ST}}{G_{ST\,(\max)}}$	[<0,1]	Hedrick's standardized $G'_{ST(Hed)}$ ensures an upper limit of 1, is reached when populations have non overlapping sets of alleles.	Gst GstG GstS, GstH	[29]
G'stN G'st (Nei)	G'_{ST} Nei via G-Statistics (Codom Data)	$G'_{ST\,(Nei)} = \frac{k({}_cH_T - {}_cH_S)}{k{}_cH_T - {}_cH_S}$	[<0,1]	Nei's standardized $G'_{ST(Nei)}$ corrects for bias when the number of populations *k* is small. Used in the calculation of G''_{ST} .	Gst GstG GstS, GstN	[29]
G"st	G''_{ST} via G-Statistics (Codom Data)	$G''_{ST} = \frac{G'_{ST\,(Nei)}}{1 - {}_cH_S}$	[<0,1]	Hedrick's standardized G_{ST} further corrected for bias when the number of populations *k* is small.	Gst GstG GstS, GstC	[29]
h	Haploid Genetic Diversity (Haploid Binary	$h = 1 - \sum p_i^2$	[0,1]	Here, *pi* 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]

	& Haploid Data)					
uh	Unbiased h	$uh = \frac{n}{n-1}\left(1-\sum p_i^2\right)$	[0,1]	An unbiased estimate of h, where *pi* is the frequency of the *i*th allele and *n* is the sample size.	AFP AFL	[14]
H	Mean haploid genetic diversity (Haploid Data)	$H = \sum_{i=1}^{k}\frac{h}{k}$	[0,1]	The mean haploid genetic diversity calculated as the arithmetic mean of *h* across *k* loci. Equivalent to H_S for codominant data.	HDL, HDP	[14]
HD	Haploid Genetic Distance			See Genetic Distance Haploid		
He	Expected heterozygosity (Codom Data)	$H_E = 1-\sum p_i^2$	[0,1]	H_E is the Expected Heterozygosity or Genetic Diversity within a population. Calculated per locus as 1 minus the sum of the squared allele frequencies, p_i^2. See also Mean He and Hs.	HFL, HFP, APT	[3]
uHe	Unbiased Heterozygosity (Codom Data)	$uH_E = \frac{2n}{2n-1}\left(1-\sum p_i^2\right)$	[0,1]	An unbiased estimate of H_E where p_i is the frequency of the *i*th allele and *n* is the sample size. Often the only estimate of H_E reported in other packages and should be the one reported from GenAlEx for research. H_E is retained for teaching purposes.	HFL HFP	[1]
Mean He	Expected heterozygosity averaged across populations (Codom Data)	$H_S = \bar{H}_E = \frac{\sum H_{Es}}{k}$	[0,1]	The average H_E or genetic diversity per population, also called H_S and used in the calculation of *F*- and *G*-statistics. Where H_{Es} is the expected heterozygosity in the *s*-th population; *k* is the number of populations. Output via *G-statistics* and also via *Frequency* when the Step-by-Step option is chosen.	HFL, HFP, Gst GstG GstS	[3]
H-indiv	Individual Heterozygosity (Codom Data)	$H-indiv = \frac{nH}{nL}$	[0,1]	*H-indiv* = the proportion of loci that are heterozygous across an individual, where *nH* is the number of heterozygous loci, and *nL* is the number of loci. When compared across individuals *H-indiv* can offer important clues about the amount and distribution of inbreeding in populations.	IH IHP	
Mean Ho	Observed heterozygosity, averaged across populations (Codom Data)	$\bar{H}_O = \frac{\sum H_{Os}}{k}$	[0,1]	The average observed heterozygosity of a collection of populations, used in the calculation of *F*-statistics and *G*-statistics alike. Here, H_{Os} is the observed heterozygosity in the *s*-th population; *k* is the number of populations. Output via *Frequency* when the Step-by-Step option is chosen, and routinely via *G-Statistics*.	HFL HFP Gst GstG GstS	[3, 29]
Ho	Observed Heterozygosity (Codom Data)	$Ho = \frac{No.\_of\_Hets}{N}$	[0,1]	Observed heterozygosity for a single locus within a population, where the number of heterozygotes is determined by direct count, *N* = sample size.	HFL, HFP Gst GstG, GstS	[3]

Hs	Average within population heterozygosity	$H_E = 1 - \sum p_i^2$ $H_S = \bar{H}_E = \frac{\sum H_{Es}}{k}$	[0,1]	Identical to the mean *He*, being the average of the within population expected heterozygosity across populations.	Gst GstG GstS	[30]
cHs	Corrected Hs	${}_cH_S = \frac{\hat{n}}{\hat{n}-1}\left[H_S - \frac{\bar{H}_O}{2\hat{n}}\right]$	[0,1]	Here H_S for a given locus is adjusted for small population size and inbreeding by the correction of Nei and Cheeser [30], where $\hat{n}$ is the harmonic mean population size for *k* populations, and $\bar{H}_o$ is the average observed within-population heterozygosity for the populations. Following [29], cH_S is used to calculate *G*-statistics.	Gst GstG GstS	[30]
Ht	Total expected heterozygosity (Codom Data)	$H_T = 1 - \sum_{i=1}^{h} \bar{p}_i^2$	[0,1]	H_T is the expected heterozygosity if all populations were pooled (no subdivision). Calculated as 1 minus the sum of the average allele frequencies over populations. Used in the calculation of *F*- and *G*-statistics. When calculating *F*-statistics via *Frequency*, H_T is only output if the Step-by-Step option is chosen in the options dialog box. H_T is routine output for *G-statistics*, along with corrected cH_T.	HFL HFP Gst GstG GstS	[3]
cHt	Corrected Ht	${}_cH_T = H_T + \frac{{}_cH_S}{\hat{n}k} - \frac{\bar{H}_O}{2\hat{n}k}$	[0,1]	H_T for a given locus is adjusted for small population size and inbreeding, using the correction of Nei and Cheeser [30]. The harmonic mean of population size over the *k* populations is $\hat{n}$; $\bar{H}_o$ is the average of the observed heterozygosity. As in Meirmans and Hedrick [29], this correction is used to calculate the *G*-statistics. Note that when calculating *G*-statistics via the Raw Frequency input data option, GenAlEx substitutes H_E for H_O in this correction (assuming HWE), because H_O is unknown.	Gst GstG GstS	[30]
I	Information index (Codom Data)	$I = \sum p_i \ln p_i$	[>0]	Calculated on a single-locus basis, where *ln* = the natural logarithm and p_i is the frequency of the *i*th allele. Equivalent to the Shannon-Weaver Index of ecology. Unlike *He*, not bounded by 1 and may therefore be a better measure of allelic and genetic diversity, though largely overlooked in genetic studies. GenAlEx 6.3 onwards also offers calculation of this and other Shannon indices for haploid and codominant data types via the Shannon options (see below).	HFL HFP HDP HDL	[86]
Log-L	Log likelihood for Population Assignment (Codom Data)	$Log(\prod p_i^2 \mathrm{x} \prod 2 p_i p_j)$	[<0]	Calculated for a given genotype, where p_i is the frequency of the *i*th allele, and p_j the frequency of the *j*-th allele at each locus in a multilocus genotype. See also Population Assignment. Log-likelihood values are negative. However, for graphing purposes, GenAlEx provides an option to convert –ve to +ve.	PI	[44-46, 87]
Na	No. of alleles (Codom and Haploid Data)		[1,n]	Determined by direct count. GenAlEx also provides the arithmetic mean across loci.	HFL, HFP, APT	
Na Freq. > 0.05	Na Freq. > 0.05		[0+]	Number of alleles with frequency greater than 5%. Calculated for Codominant and Haploid Data.	APT	[20]

	No. LComm Alleles (<=50%)		[0+]	Number of locally common alleles (Freq. > 5%) occurring in 50% or less of the populations.	APT	[20]
	No. LComm Alleles (<=25%)		[0+]	Number of locally common alleles (Freq. > 5%) occurring in 25% or less of the populations.	APT	[20]
	No. of private alleles		[0+]	Equivalent to the number of alleles unique to a single population in the data set.	APT, PAL, PAS	[20]
Ne, cNe	Effective number of alleles (Haploid Data) (Codom Data)	$N_e = \dfrac{1}{1-H_E}$ $cN_e = \dfrac{1}{1-H_S}$	[1,n]	Here *Ne* represents an estimate of the number of equally frequent alleles in an ideal population. Enables meaningful comparisons of allelic diversity across loci with diverse allele frequency distributions. GenAlEx provides two slightly different estimates. The first (*Ne*) via *Frequency* is calculated by locus from H_E for each population. The second (*cNe)* via *G-statistics* is calculated by locus over populations based on H_S.	HFL HFP Gst	[86]
Nei D	Nei's Genetic Distance	$D = -\ln(I)$	[>0]	Nei's genetic distance *D*, where *I* is Nei's Genetic Identity (see below for details).	NeiP NeiL NeiT	[27]
Nei uD	Nei's Unbiased Genetic Distance	$uD = -\ln(uI)$	[0+]	Nei's unbiased genetic distance *uD*, where *uI* is the Unbiased Genetic Identity (see below for details)	UNeiP UNeiL UNeiT	[27]
Nei I	Nei's Genetic Identity	$I = \dfrac{J_{xy}}{\sqrt{(J_x J_y)}}$ $J_{xy} = \sum_{i=1}^{k} p_{ix} p_{iy}$, $J_x = \sum_{i=1}^{k} p_{ix}^2 , J_y = \sum_{i=1}^{k} p_{iy}^2$	[0,1]	Here, p_{ix} and p_{iy} are the frequencies of the *i*th allele in populations *x* and *y.* For multiple loci, *Jxy*, *Jx* and *Jy* 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.* GenAlEx provides a step-by-step option to illustrate the calculation of Nei's *GD* and *ID*. See also Tutorial 1.	NeiP NeiL NeiT	[27]

Nei uI	Nei's Genetic Identity	$uI = \frac{J_{xy}}{\sqrt{(uJ_x uJ_y)}}$ $J_{xy} = \sum_{i=1}^{k} p_{ix} p_{iy}$ $uJ_x = \frac{2n(\sum_{i=1}^{k} p_{ix}^2 - 1)}{2n - 1}$	[0,1]	Here, p_{ix} and p_{iy} are the frequencies of the *i*th allele in populations *x* and *y*. For multiple loci, *Jxy*, *uJx* and the analogous *uJy* 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*. 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 provides a step-by-step option to illustrate the process of calculation.	UNeiP UNeiL UNeiT	[27]
Nm	Number of Migrants (Codom Data)	$Nm = \left[(1/F_{ST}) - 1\right]/4$	[0+]	Where F_{ST} represents the degree of population genetic differentiation. Estimation of *Nm* via F_{ST} and related methods is now generally considered problematic. Nevertheless, retained for teaching purposes.	Fst FstT	[6, 10]
Nm	Haploid Number of Migrants (Haploid Data)	$Nm = \left[(1/\phi_{PT}) - 1\right]/2$	[0+]	The haploid equivalent of Nm, where ϕ_{PT} the haploid analog of *Fst* represents the degree of population genetic differentiation. Note division by 2 rather than 4 for this hapoid case.	Fst FstT	[6, 10]
Nm	*Nm* via Shannon (Diploid)	$Nm = \left(\frac{0.156}{{}^{S}H_{UA}}\right)^2$	[0+]	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.	SHuaP SHuaL SH	[72]
Nm	*Nm* via Shannon (Haploid)	$Nm = \left(\frac{0.22}{{}^{S}H_{UA}}\right)^2$	[0+]	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.	SHuaP SHuaL SH	[72]
Omega				The spatial correlogram wide test statistic, *Omega*. See Spatial in Table 1 and Tutorial 3 for details.	RC MPOS	[66]
Overlap O'	[0,1] Scaled Overlap O'=1-D'	1-(*[0,1] Scaled Diversity D'*)	[0,1]	Calculated as 1-*Scaled Diversity D*' via Shannon->Diversity Partition. See also *sH*, *sH(WP)*, *sH(GT)*, *sH(AP)* and Scaled Diversity.	SHT SH	[74]
P	Polymorphism (Codom Data) (Haploid Data)	Calculated as percentage of polymorphic loci across loci.	[0,100 %]	Once frequently reported in allozyme studies, where the type and number of loci were 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 high degree of polymorphism. May be useful for multi-locus DNA markers such as AFLPs.	HFL HFP	[73]

P1	Probability of Exclusion (Codom Data)	$P1 = 1 - 2\sum p_i^2 + \sum p_i^3$ $+2\sum p_i^4 - 3\sum p_i^5$ $-2(\sum p_i^2)^2 + 3\sum p_i^2 \sum p_i^3$	[0,1]	*P1* estimates the probability of exclusion when the other parent is known (following equation 1a in [60] for 'One parent' exclusion). Labelled in GenAlEx 6.5 onwards as 'P1-When the other parent is known'.	PX1	[60]
P2	Probability of Exclusion (Codom Data)	$P2 = 1 - 4\sum p_i^2 + 2(\sum p_i^2)^2$ $+4\sum p_i^3 - 3\sum p_i^4$	[0,1]	*P2* estimates the probability of exclusion when one parent is known but the other genotype is unavailable (following Eq. 2a in [60] for 'Missing parent' exclusion). Labelled in GenAlEx 6.5 onwards as 'P2-When genotype of one parent is missing'.	PX2	[60]
P3	Probability of Exclusion (Codom Data)	$P3 = 1 + 4\sum p_i^4 - 4\sum p_i^5$ $-3\sum p_i^6 - 8(\sum p_i^2)^2$ $+8(\sum p_i^2)(\sum p_i^3) + 2(\sum P_i^3)^2$	[0,1]	*P3* estimates the probability of excluding a putative parent pair (following equation 3a in [60] for 'Both parents exclusion'). Labelled in GenAlEx 6.5 onwards as 'P3-Excluding a putative parent pair'.	PX3	[60]
Phi'PT	Standardized ***Φ'***$_{\mathrm{PT}}$ via AMOVA (Haploid Data)		[<0,1]	Analogous to standardised F'_{ST}, for haploid data. Represents a [0,1]-scaled estimator of differentiation, following [21]. See [29] for additional background on standardized *F*- and *G*-statistics. See [21] and Tutorial Part 7 for details on the calculation of standardized *F*- and ***Φ***statistics via AMOVA.	PhiPT PhiPTP PhiPTL PhiPTT	[18, 21]
Phi'PR	Standardized ***Φ'***$_{\mathrm{PR}}$ via AMOVA (Haploid Data)		[<0,1]	A [0,1]-scaled estimator of differentiation among populations within regions, as estimated via a regional AMOVA of haploid data, following [21]. Analogous to F'_{SR} for codominant data.	PhiPT PhiPTP PhiPTL PhiPTT	[18, 21]
Phi'RT	Standardized ***Φ'***$_{\mathrm{RT}}$ via AMOVA (Haploid Data)		[<0,1]	A [0,1]-scaled estimator of differentiation among regions, as estimated via a regional AMOVA of haploid data, following [21]. Analogous to F'_{RT} for codominant data.	PhiPT PhiPTP PhiPTL PhiPTT	[18, 21]
PI	Probability of Identity (Codom Data)	$PI = 2(\sum p_i^2)^2 - \sum p_i^4$ for each locus.	[0,1]	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 [15, 55, 88] where it is also called the *Match Probability*, *Matching Probability* and *Power of Inclusion*. *1-PI* is called the *Exclusion Power*, or *Power of Discrimination* [14, 15, 88]. *PI* is also used for assessing the number of loci required for genetic tagging [56-58].	PI	[55-58]

PIsibs	Probability of Identity Sibs (Codom Data)	$PIsibs =$ $0.25 + (0.5\sum p_i^2)$ $+[0.5(\sum p_i^2)^2]$ $-(0.25\sum p_i^4)$	[0,1]	In addition to *PI* GenAlEx also calculates the more conservative *PIsibs* that estimates the probability of identity among siblings [56, 57]. 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. [58] for an example application of *PISibs*.	PI	[55-58]
Pgen	Probability of Genotype (Codom Data)	$Pgen = (\prod p_i)2^h$	[0,1]	Identical to the genotype probability. *Pgen* provides an estimate of the probability of identical genotypes arising from sexual reproduction and random mating, where *pi* is the frequency of each allele (two per locus) observed in the multilocus genotype and *h* the number of loci that are heterozygous (see [61, 63]).	CLP	[61, 63]
Pse	Probability of Second Encounter (Codom Data)	$Pse = 1-(1-Pgen)^N$	[0,1]	*Pse* provides an estimate of the probability of a second encounter of a specific multilocus genotype generated by sexual reproduction under random mating. GenAlEx calculates *Pse* using *N* = total no. of samples, irrespective of the number of different genotypes. The estimate *Psex Ngen* is calculated using *n* = no. of different genotypes following [61]	CLP	[61, 62]
Psex	Probability of Sex (Codom Data)	$Psex =$ $\sum_n^N \frac{N!}{n!(N-n)!}$ $*(Pgen)^i$ $*(1-Pgen)^{N-n}$	[0,1]	*Psex* 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 *Psex* using *N* = total no. of samples, irrespective of the number of different genotypes. The estimate *Psex Ngen* is calculated using *N* = no. of different genotypes.	CLP	[61, 62]
r	Autocorrelation coefficient		[−1,1]	See Tutorial 3 for a detailed overview of the spatial autocorrelation procedures in GenAlEx.		[31, 64, 65]
r	Pairwise Relatedness		[−1,1]	For a summary of the formulas, see Ritland [53]. See also Pairwise Relatedness in Table 1 and Tutorial 4.	RI, LR, QG, PSA	[53]
r	Linkage Disequilibrium	$r = \sqrt{r^2}$ with same sign as *D*.		See rSq below, *D* and also Pairwise Linkage Disequilibrium.	LDK	[4]
rc	Autocorrelation coefficient (multiple pops)		[−1,1]	See Tutorial 3 for a detailed overview of the spatial autocorrelation procedures in GenAlEx.		[31, 64, 65]

Rst	R_{ST} via AMOVA with no regional data structure. (Codom Data)	*Default option* $R_{ST} = \frac{V_{AP}}{(V_{WI}+V_{AI}+V_{AP})}$ *Suppress within individual analysis* option $R_{ST} = \frac{V_{AP}}{(V_{AP}+V_{WP})}$	[−1,1]	The estimation of R_{ST} parallels that for F_{ST} and Φ_{PT} (see above) except that R_{ST} 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 *Codom-Microsat* genetic distance. R_{ST} was introduced by Slatkin [32] as an F_{ST} 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 F_{ST}, as a consequence. See also AMOVA, *Phi* Statistics, F_{ST} and Genetic Distance.	Rst RstP RstL RstT	[19, 32]
rSq	Linkage Disequilibrium	$r^2 = \frac{D^2}{p_1 p_2 q_1 q_2}$	[0,1]	See *D* above for details, and also Pairwise Linkage Disequilibrium.	LDK	[4]
rxy	Mantel correlation coefficient	$r_{xy} = \frac{SPxy}{\sqrt{[SSxSSy]}}$ $SSx = \sum_{i \neq j}^{N}\left(x_{ij}-\bar{x}\right)^2$ $SSy = \sum_{i \neq j}^{N}\left(y_{ij}-\bar{y}\right)^2$ $SPxy = \sum_{i \neq j}^{N}\left(x_{ij}-\bar{x}\right)\left(y_{ij}-\bar{y}\right)$	[−1,1]	*SPxy* is the sum of cross products of corresponding elements of the **X** and **Y** Matrices; *SSx* is the sum of products of **X** matrix elements and *SSy* that of **Y** matrix elements. In addition to listing *rxy*, 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	[40, 41]
SE	Standard Error	$SE = \frac{s}{\sqrt{n}}$	[>0]	*SE* is the standard error of the mean and is widely reported by GenAlEx when the arithmetic mean and other summary statistics are reported. Where *s* is the standard deviation and *n* is the sample size.		
sHa	Shannon's *Allele Information* index via Shannon-Pairwise	${}^{S}H_{A} = -\sum p_i \log_2 p_i$ ${}^{S}H_{A1} = -\sum p_{i1} \log_2 p_{i1}$ and ${}^{S}H_{A2} = -\sum p_{i2} \log_2 p_{i2}$	[>0]	One of 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 ${}^{S}H_{A}$. At each specific locus across multiple populations, we consider each pairwise combination of populations in turn, calculating ${}^{S}H_{A1}$ and ${}^{S}H_{A2}$ for each pair of 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). The formulae shown here use log base-2, following Sherwin et al. [70], who recommend the use of log base-2, because the Shannon Index readily translates into heterozygosity. Note that sHa is also output, via the Shannon Partition option, where the user has the option of using the natural log (ln), which leads naturally to the log-likelihood test criteria. See Sherwin et al. [72], Tutorial 1 and Tutorial Appendix 1.1 for further details.	SHa SHuaP SHuaL SH	[72]

sHu	Shannon's *Total Information* index via Shannon-Pairwise	$^{S}H_{U} = -\sum \overline{p}_i \log_2 \overline{p}_i$ $\overline{p}_i = wt_1 p_{i2} + wt_2 p_{i2}$ $wt_1 = \frac{ct_1}{ct_1 + ct_2}$ $wt_2 = \frac{ct_2}{ct_1 + ct_2}$	[>0]	Another of the Shannon's Information Indices. For a specific locus in a given population, Shannon's Total Allele Information index is calculated by the formula for $^{S}H_U$. Where $\overline{p}_i$ is the weighted average frequency of the *i*th allele for each pair of populations, with the weights calculated as shown, and the *ct* values are the total allele counts at the locus for the respective populations. See sHa (above) for comments on the choice of log base, Sherwin et al. [72], Tutorial 1 and Tutorial Appendix 1.1 for further details.	SHa SHuaP SHuaL SH	[72]
sHua	Shannon's *Mutual Information* index via Shannon-Pairwise	$^{S}H_{UA} = {}^{S}H_{U}$ $-wt_1 {}^{S}H_{A1}$ $-wt_2 {}^{S}H_{A2}$ $G = 1.3863 {}^{S}H_{UA}(ct_1 + ct_2)$	[0+]	Shannon's *Mutual Information* index $^{S}H_{UA}$ is calculated for each pair of populations via the Shannon-Pairwise option, as shown here. Sherwin et al. [70] further showed that $^{S}H_{UA}$ can be readily converted to a log-likelihood contingency test statistic *G*. The formula shown for *G* applies when log base 2 is used in the calculation of $^{S}H_A$ and $^{S}H_U$. The degrees of freedom *DF* are calculated as the (number of populations compared - 1) x (number of alleles compared - 1). For research purposes, we recommend statistical testing by random permutation, because there are reports that the log-likelihood *G*-test and associated chi-square probabilities may exhibit high type I error rates (false rejection of the null hypothesis). GenAlEx 6.5 onwards, offers testing by random permutation via the Shannon-Partition option (see below). See also Nm in this table for calculations of Nm via $^{S}H_{UA}$, Tutorial 1 and Tutorial Appendix 1.1 for further details.	SHa SHuaP SHuaL SH	[72]
sH(AP)	*Among Population Information* via Shannon-Diversity Partition	$sH(AP) = sH(GT) - sH(WP)$	[0+]	*sH(AP)* is calculated for each pair of populations via the Shannon->Diversity Partition option, with statistical testing by random permutation following Smouse et al. [74]. This permutation test is recommended over the log-likelihood *G*-test (offered via Shannon->Pairwise Pops), to avoid the risk of elevated type I error rates (false rejection of the null hypothesis). See also *sH(GT)* and *sH(WP)* via Shannon-Partition and the equivalent *sHua* calculated via the Shannon->Pairwise Pops option.	SHAPP SHAPL SHAPT SH SHT	[74]

sH(WP)	*Within Population Information* via Shannon-Diversity Partition	$sH(WP) = wt_1 \times sH(WP1) + wt_2 \times sH(WP2)$ $sH(WP1) = -\sum p_{i1} \ln p_{i1}$ and $sH(WP2) = -\sum p_{i2} \ln p_{i2}$	[>0]	Here the calculation is shown, using the natural log (ln), following Smouse et al. [74]. It is standard to use the natural log in ecological applications of Shannon Diversity and for this reason use of the natural log may be appropriate when making comparisons between different levels of diversity. Nevertheless, GenAlEx does allow users to change the log base, as required, offering log base-2, log base-10 and the log base-e. The scaling of the Shannon indices will change with the base selected, but the standardized diversities *alpha*, *beta* and *gamma*, are invariant with respect to the choice of log base. The exponential of *sH(WP)* equals the *alpha diversity*. See also the equivalent *sHa* calculated via the Shannon->Pairwise Pops option.	SHT SH	[74]
sH(GT)	*Grand Total Information* via Shannon-Diversity Partition	$sH(GT) = -\sum \overline{p}_i \ln \overline{p}_i$	[>0]	Here the calculation of is shown, using the natural log (ln). See *sH(WP)* for comments on the choice of log base. The exponential of *sH(GT)* equals the *Gamma Diversity*. See also the equivalent *sHu* calculated via the Shannon->Pairwise Pops option.	SHT SH	[74]
sH	Shannon Statistics Summary Table via Shannon-Diversity Partition	*sH* Among Pops = *sH(AP)* *sH* Within Pops = *sH(WP)* *sH* Grand Total = *sH(GT)* Alpha Diversity = *exp(sH(WP)* Beta Diversity = *exp(sH(AP))* Gamma Diversity = *exp(sH(GT))* [0,1] Scaled Diversity = *D'* = (1-(1/Div))/(1-(1/DivWt)) [0,1] Scaled Overlap = 1 – *D'*		A locus-by-locus *Shannon Statistics Summary Table* is an output option accessible via the Shannon->Diversity Partition option. In this AMOVA-like table, the label *sH* represents a generic label for the relevant Shannon statistic calculated using the natural logarithm. The corresponding *Diversity* values are calculated as the exponential of the relevant *sH* values. [0,1] Scaled *Diversity* and *Overlap* are calculated as shown.	SHT	[74]
t	Outcrossing rate (Codom Data)	$t = \frac{(1-F)}{(1+F))}$	[0,1+]	Not output in GenAlEx, but can easily be calculated from the *F*-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.	HFL HFP	[86]
T2				The squared paired-sample t-test statistic *T2* output during spatial heterogeneity tests. See Spatial in Table 1 and Tutorial 3 for details.	MPTS MPT	[66]

Ve	Expected variance (Haploid Data)	$V_e = \sum h(1-h)$, $K = \Sigma h$		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 Ve. See also Haploid Disequilibrium.	HDE	[36]
Vo	Observed variance (Haploid Data)			The observed variance of *K*. The disequilibrium index is *Vo/Ve*. To test if this ratio is significantly greater than one, GenAlEx employs a randomisation test, following [36, 38]. For each random sample, *Vr* is calculated as the variance of the randomized data set and the probability of observing a *Vr* value as extreme as that observed (Vo) is calculated.	HDE	[36]

References

1. Peakall, R and Smouse, PE. 2012. *GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update.* Bioinformatics **28**, 2537-9.
2. Peakall, R and Smouse, PE. 2006. *GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research.* Molecular Ecology Notes **6**, 288-295.
3. Hartl, DL and Clark, AG. 1997. *Principles of Population Genetics 3rd Ed*. Sunderland, Massachusetts: Sinauer Associates, Inc.
4. Hedrick, PW. 2005. *Genetics of Populations. Third Edition.* Sudbury, Massachusetts: Jones and Bartlett Publishers.
5. Hedrick, PW. 2009. *Genetics of Populations (4th Ed)*: Jones and Bartlett Publishers.
6. Frankham, R, et al., *Introduction to Conservation Genetics*. 2002, Cambridge University Press: Cambridge.
7. Allendorf, FW and Luikart, G. 2006. *Conservation and Genetics of Populations*: Wiley-Blackwell. 642.
8. Hartl, DL. 2000. *A Primer of Population Genetics 3rd Ed*. Sunderland, Massachusetts: Sinauer Associates, Inc.
9. Conner, JK and Hartl, DL. 2004. *A Primer of Ecological Genetics*. Sunderland, Massachusetts: Sinauer Associates, Inc.
10. Frankham, R, et al. 2004. *A Primer of Conservation Genetics*. Cambridge: Cambridge University Press.
11. Weir, BS. 1990. *Genetic Data Analysis*. Sunderland, Massachusetts: Sinauer Associates, Inc.
12. Weir, BS. 1996. *Genetic Data Analysis II*. Sunderland: Sinauer Associates Inc.
13. Balding, DJ, et al., eds. *Handbook of Statistical Genetics*. Wiley series in probability and statistics. 2001, John Wiley & Sons, Ltd: Chichester.
14. Anon, A. 1996. *The Evaluation of Forensic DNA Evidence.* Washington, DC: National Academy Press.
15. Buckleton, J, et al. 2005. *Forensic DNA Evidence Interpretation*. New York: CRC Press.
16. Excoffier, L, et al. 1992. *Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitocondrial DNA restriction sites.* Genetics **131**, 479-491.
17. 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.
18. 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.
19. 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.
20. 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.
21. Meirmans, PG. 2006. *Using the AMOVA framework to estimate a standardized genetic differentiation measure.* Evolution **60**, 2399-2402.
22. Excoffier, L and Lischer, HEL. 2010. *Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows.* Molecular Ecology Resources **10**, 564-567.
23. Meirmans, PG and Van Tienderen, PH. 2004. *GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms.* Molecular Ecology Notes **4**, 792-794.
24. Wright, S. 1946. *Isolation by distance under diverse systems of mating.* Genetics **31**, 39-59.

25. Wright, S. 1951. *The genetical structure of populations.* Ann. Eugenics **15**, 323-354.
26. Wright, S. 1965. *The interpretation of population structure by F-Statistics with special regard to systems of mating.* Evol. **19**, 395-420.
27. Hedrick, PW. 2000. *Genetics of Populations 2nd Ed*. Boston: Jones and Bartlett.
28. Nei, M. 1977. *F-statistics and analysis of gene diversity in subdivided populations.* Annals of Human Genetics **41**, 225-233.
29. Meirmans, PG and Hedrick, PW. 2011. *Assessing population structure: FST and related measures.* Molecular Ecology Resources **11**, 5-18.
30. Nei, M and Chesser, RK. 1983. *Estimation of fixation indexes and gene diversities.* Annals of Human Genetics **47**, 253-259.
31. Smouse, PE and Peakall, R. 1999. *Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure.* Heredity **82**, 561-573.
32. Slatkin, M. 1995. *A measure of population subdivision based on microsatellite allele frequencies.* Genetics **139**, 1463.
33. Raymond, M and Rousset, F. 1995. *Genepop (version 1.2) - population genetics software for exact tests and ecumenicism.* Journal of Heredity **86**, 248-249.
34. Engels, WR. 2009. *Exact tests for Hardy-Weinberg proportions.* Genetics **183**, 1431-1441.
35. Slatkin, M. 2008. *Linkage disequilibrium - understanding the evolutionary past and mapping the medical future.* Nature Reviews Genetics **9**, 477-485.
36. Gordon, DM. 1997. *The genetic structure of Escherichia coli populations in feral house mice.* Microbiology **143**, 2039-2046.
37. Brown, ADH, et al. 1980. *Multilocus structure of natural populations of Hordeum spontaneum.* Genetics **96**, 523-536.
38. 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.
39. Mantel, N. 1967. *The detection of disease clustering and a generalized regression approach.* Cancer Res. **27**, 209-220.
40. Smouse, PE and Long, JC. 1992. *Matrix correlation analysis in anthropology and genetics.* Yearbook Phys. Anthropol. **35**, 187-213.
41. Smouse, PE, et al. 1986. *Multiple regression and correlation extensions of the Mantel test of matrix correspondence.* Systematic Zoology **35**, 627-632.
42. Nei, M. 1972. *Genetic distance between populations.* American Naturalist **106**, 283-392.
43. Nei, M. 1978. *Estimation of average heterozygosity and genetic distance from a small number of individuals.* Genetics **89**, 583-590.
44. Paetkau, D, et al. 1995. *Microsatellite analysis of population structure in canadian polar bears.* Molecular Ecology **4**, 347-354.
45. 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.
46. Waser, PM and Strobeck, C. 1998. *Genetic signatures of interpopulation dispersal.* Trends in Ecology and Evolution **13**, 43-44.
47. Cornuet, JM, et al. 1999. *New methods employing multilocus genotypes to select or exclude populations as origins of individuals.* Genetics **153**, 1989-2000.
48. Piry, S, et al. 2004. *GENECLASS2: A software for genetic assignment and first-generation migrant detection.* Journal of Heredity **95**, 536-539.
49. Pritchard, JK, et al. 2000. *Inference of population structure using multilocus genotype data.* Genetics **155**, 945-959.
50. Orloci, L. 1978. *Multivariate analysis in vegetation research.* The Hague: Dr W. Junk B. V.
51. Lynch, M and Ritland, K. 1999. *Estimation of pairwise relatedness with molecular markers.* Genetics **152**, 1753-1766.
52. Ritland, K. 1996. *Estimators for pairwise relatedness and individual inbreeding coefficients.* Genetical Research **67**, 175-185.

53. Ritland, K. 2000. *Marker-inferred relatedness as a tool for detecting heritability in nature.* Molecular Ecology **9**, 1195-1204.
54. Queller, DC and Goodnight, KF. 1989. *Estimating relatedness using genetic markers.* Evolution **43**, 258-275.
55. 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.
56. Taberlet, P and Luikart, G. 1999. *Non-invasive genetic sampling and individual identification.* Biological Journal of the Linnean Society **68**, 41-55.
57. Waits, LP, et al. 2001. *Estimating the probability of identity among genotypes in natural populations: cautions and guidelines.* Molecular Ecology **10**, 249-256.
58. 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.
59. 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.
60. Jamieson, A and Taylor, SCS. 1997. *Comparisons of three probability formulae for parentage exclusion.* Animal Genetics **28**, 397-400.
61. 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.
62. Stenberg, P, et al. 2003. *MLGsim: a program for detecting clones using a simulation approach.* Molecular Ecology Notes **3**, 329-331.
63. 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.
64. Double, MC, et al. 2005. *Dispersal, philopatry and infidelity: dissecting local genetic structure in superb fairy-wrens (Malurus cyaneus).* Evolution **59**, 625-635.
65. 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.
66. Smouse, PE, et al. 2008. *A heterogeneity test for fine-scale genetic structure.* Molecular Ecology **17**, 3389-3400.
67. 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.
68. 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.
69. Banks, SC and Peakall, R. 2012. *Genetic spatial autocorrelation can readily detect sex-biased dispersal.* Molecular Ecology **21**, 2092-2105.
70. 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.
71. Mossman, CA and Waser, PM. 1999. *Genetic detection of sex-biased dispersal.* Molecular Ecology **8**, 1063-1067.
72. Sherwin, W, et al. 2006. *Measurement of biological information with applications from genes to landscapes* Molecular Ecology **15**, 2857-2869.
73. Smouse, PE and Ward, RH. 1978. *A comparison of the genetic infrastructure of the Ye'cuana and Yanomama: a likelihood analysis of genotypic variation among populations.* Genetics **88**, 611-631.

74. Smouse, PE, et al. 2015. *An informational diversity framework, illustrated with sexually deceptive orchids in early stages of speciation.* Molecular Ecology Resources, In press: DOI 10.1111/1755-0998.12422.

75. Austerlitz, F and Smouse, PE. 2001. *Two-generation analysis of pollen flow across a landscape. II. Relation between Phi(ft), pollen dispersal and interfemale distance.* Genetics **157**, 851-857.

76. 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.

77. Austerlitz, F and Smouse, PE. 2002. *Two-generation analysis of pollen flow across a landscape. IV. Estimating the dispersal parameter.* Genetics **161**, 355-363.

78. Smouse, PE, et al. 2001. *Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females.* Evolution **55**, 260-271.

79. Dyer, RJ. 2005. *GENER: a server-based analysis of pollen pool structure.* Molecular Ecology Notes.

80. Lynch, M and Milligan, BG. 1994. *Analysis of population genetic structure with RAPD markers.* Mol. Ecol. **3**, 91-99.

81. Zhivotovsky, LA. 1999. *Estimatng population structure in diploids with multilocus dominant DNA markers.* Molecular Ecology **8**, 907.

82. Lewis, PO and Zaykin, D. 2001. *Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.1* http://www.eeb.uconn.edu/people/plewis/software.php.

83. Jost, L. 2008. *GST and its relatives do not measure differentiation.* Molecular Ecology **17**, 4015-4026.

84. Slatkin, M. 1995. *A measure of population subdivision based on microsatellite allele frequencies (vol 139, pg 457, 1995).* Genetics **139**, 1463.

85. Nei, M. 1987. *Molecular Evolutionary Genetics*. New York: Columbia University Press.

86. 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.

87. Davies, N, et al. 1999. *Determining the source of individuals: multilocus genotyping in nonequilibrium population genetics.* Trends in Ecology & Evolution **14**, 17-21.

88. Butler, JM. 2005. *Forensic DNA Typing: Biology, Technology and Genetics of STR Markers. 2nd Ed.* Oxford: Elsevier.