Heredity (2009) 102, 506–513

www.nature.com/hdy

ORIGINAL ARTICLE

How well do evolutionary trees describe genetic relationships among populations?

ST Kalinowski
Department of Ecology, Montana State University, Lewis Hall, Bozeman, MT, USA

Bifurcating evolutionary trees are commonly used to describe genetic relationships between populations, but may not be appropriate for populations that did not evolve in a hierarchical manner. The degree to which bifurcating trees distort genetic relationships between populations can be quantified with R^2, the proportion the variation in a matrix of genetic distances between populations that is explained by a tree. Computer simulations were used to measure how well the unweighted pair group method with arithmetic mean (UPGMA) and neighbor-joining (NJ) trees depicted population structure for three evolutionary models: a hierarchical model of population fragmentation, a linear stepping-stone model of gene flow and a two-dimensional stepping-stone model of gene flow. These simulations showed that the UPGMA did an excellent job of describing population structure when populations had a bifurcating history of fragmentation, but severely distorted genetic relationships for the linear and two-dimensional stepping-stone models. The NJ algorithm worked well in a broader range of evolutionary histories, including the linear stepping-stone model. A computer program for performing the calculations described in this study is available for download at www.montana.edu/kalinowski.
Heredity (2009) **102,** 506–513; doi:10.1038/hdy.2008.136; published online 28 January 2009

Keywords: tree; population; UPGMA; neighbor joining; cophenetic correlation

Introduction

One of the primary goals of population genetics is to describe genetic differences among populations of the same species. This can be challenging because populations often have complex evolutionary histories. We can often safely assume that species have a tree-like phylogeny, and, therefore can be clustered into hierarchies. However, when gene flow and genetic drift shape intraspecific population structure, the resulting genetic relationships can be decidedly nonhierarchical. For example, if the rate of gene flow between populations is inversely related to the geographic distance separating populations, an isolation-by-distance pattern may evolve, in which populations will be genetically similar to their closest neighbors. In this circumstance, populations cannot easily be clustered into groups, because each population is similar to a different set of populations.

Despite the fact that populations are frequently not expected to cluster nicely into hierarchical sets, bifurcating trees are one of the most commonly used statistical tools to summarize genetic relationships between populations. For example, in an earlier investigation (Gutiérrez-Espeleta *et al*., 2000), I used a unweighted pair group method with arithmetic mean (UPGMA) tree to describe genetic relationships between populations of desert bighorn sheep populations in the American Southwest—even though there was a strong isolation-by-distance relationship in the data. It is reasonable to ask whether this UPGMA tree accurately depicted genetic relationships between these populations or whether the tree imposed a hierarchical pattern upon data that did not have one, and thereby distorted relationships between the populations. Trees have been used to describe genetic relationships among populations for decades (for example, Edwards and Cavalli-Sforza, 1964), and their usefulness is beyond question. However, there may be some populations for which bifurcating trees cannot accurately describe genetic relationships. It would be useful to identify such populations.

The goals of this investigation are threefold. First, I show how R^2 values for evolutionary trees can be used to assess how faithfully trees depict genetic differences between populations. Next, I use computer simulation to explore how well bifurcating trees describe genetic relationships among populations having three different types of evolutionary histories. Last, I introduce a computer program to perform the calculations described in this study.

Correspondence: Dr ST Kalinowski, Department of Ecology, Montana State University, 310 Lewis Hall, Bozeman, MT 59717, USA.
E-mail: skalinowski@montana.edu
Received 2 September 2008; revised 10 November 2008; accepted 23 December 2008; published online 28 January 2009

Tree construction

A brief review of how trees are read and how they are constructed will illustrate why some trees may not be able to accurately summarize genetic relationships among populations. Trees are interpreted by summing up the length of the branches separating populations. If the lengths of the branches are short, the tree describes the populations as being genetically similar; if the branches separating two populations are long, the tree describes the populations as being genetically different.

Population structure is usually inferred from allele frequency data, and distance-based approaches are normally used to make trees (see Felsenstein (2004) for a readable introduction). Distance-based tree construction algorithms begin by calculating a genetic distance, such as Weir and Cockerham's (1984) F_{ST} or Nei's (1978) standard distance (1978), between each pair of populations. This matrix of genetic differences is the raw data from which a tree is made. The goal of a tree construction algorithm is to construct a tree, so that the distance between each pair of populations along the branches of the tree is equal to the genetic distance between the populations in the genetic distance matrix.

The two most commonly used distance-based methods for building trees are the UPGMA, and the neighbor-joining (NJ) method (Saitou and Nei, 1987; see Felsenstein (2004) for a detailed description of both methods). Both methods build trees by searching the genetic distance matrix for the most similar populations, and then connecting these populations at a node. The main difference between the algorithms is how branch lengths are defined. UPGMA trees implicitly assume an equal rate of evolution in all populations, and assign equal branch lengths from a node to all the populations connected to the node. When displayed as a rooted tree, the tips of a UPGMA tree are all at the same distance from the root (which gives the characteristic right-aligned appearance for the rooted UPGMA trees when display in a rectangular manner). The NJ algorithm does not assume equal rates of evolution, and allows branch lengths to vary. If two populations are joined to a node, the branch connecting one population to the node may be shorter than the other. Because NJ trees have more parameters, they should be able to more faithfully match the genetic distance matrix.

Calculating R^2 for trees

The degree of fit of a tree to a matrix of genetic distances can be quantified with R^2, the proportion of variation in the genetic distance matrix that is explained by the tree. This familiar statistic is calculated for trees as follows. Let $\hat{D}_{ij}$ represent the estimated genetic distance between populations i and j in the genetic distance matrix. Let $\hat{d}_{ij}$ represent the distance between populations i and j through the branches of a tree. If the tree fits observed data well, the genetic distances between each pair of populations in the tree will be approximately equal to the observed genetic distance between them. R^2 is calculated in the usual manner

$$R^2 = 1 - \frac{\Sigma(\hat{D}_{ij} - \hat{d}_{ij})^2}{\Sigma(\hat{D}_{ij} - \bar{\hat{D}})^2} \quad (1)$$

where summation is taken over all pairs of populations and $\bar{\hat{D}}$ is the average of $\hat{D}$. If R^2 is near 1.0, the tree represents a good summary of the genetic relationships shown in the distance matrix. If R^2 is not near 1.0, the tree does not represent a good summary of the genetic relationships among populations and probably should not be used to depict population structure.

Sokal and Rohlf (1962) introduced the method described above for quantifying how well evolutionary trees fit genetic distance data. Their calculation measured genetic distance in the tree differently, and instead of reporting R^2 values, they calculated a correlation coefficient (which they called the cophenetic correlation coefficient) between observed and fitted distances, but the concept is identical. The cophenetic correlation coefficient was widely used by numerical taxonomists in the 1960s, but is not frequently used by contemporary molecular ecologists.

An empirical example

An empirical example shows the usefulness of calculating R^2 values for trees. Let us consider the bighorn sheep data of Gutiérrez-Espeleta *et al*. (2000) mentioned above. Gutiérrez-Espeleta *et al*. (2000) genotyped 10 microsatellite loci for 279 sheep distributed among 13 locations in North America, most of which were in the deserts of Southern California and Arizona. The standard distance of Nei (1978) was used as a genetic distance. The R^2 value for the UPGMA tree published by Gutiérrez-Espeleta *et al*., 2000, calculated according to Equation (1), is 74%. R^2 for an NJ tree was higher (89%) which shows that the authors' choice to use a UPGMA tree imposed a moderate amount of distortion of the relationships among the populations of sheep. The biggest difference between the two trees is the placement of the population

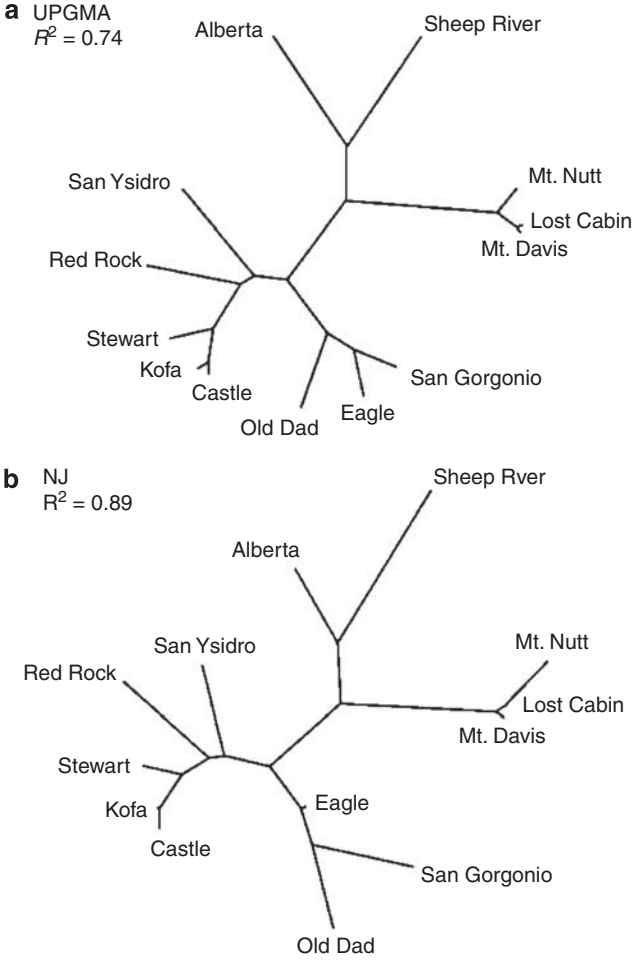

Figure 1 UPGMA (**a**) and NJ (**b**) trees for bighorn sheep populations studied by Gutiérrez-Espeleta *et al*., 2000. The standard genetic distance of Nei (1978) was used. NJ, neighbor joining; UPGMA, the unweighted pair group method with arithmetic mean.

from the Eagle Mountains (Figure 1). The UPGMA tree clustered this population with populations from San Gorgonio and the Old Dad Mountains, whereas the NJ tree placed the Eagle population as intermediate between a San Gorgonia/Old Dad cluster and the rest of the populations. Had F_{ST} been used a genetic distance, the difference between UPGMA and NJ trees would have been more extreme. A UPGMA tree for this data using pairwise F_{ST} as a genetic distance has an $R^2$of 0.61, and an NJ tree has an R^2 of 0.91.

Methods

Simulated genetic data were used to investigate how well evolutionary trees depict genetic relationships between populations. Three evolutionary models were examined: a hierarchical model of population fragmentation (Figure 2a), a linear stepping-stone model of gene flow (Figure 2b), and a two-dimensional stepping-stone model of gene flow (Figure 2c). Two general questions were explored. First, can evolutionary trees accurately reflect the genetic structure of these populations? This was explored by assuming that the true genetic distances between populations was known and constructing trees from this matrix of genetic distances. R^2 values for these trees were then calculated to evaluate how much trees distorted the genetic relationships among populations.

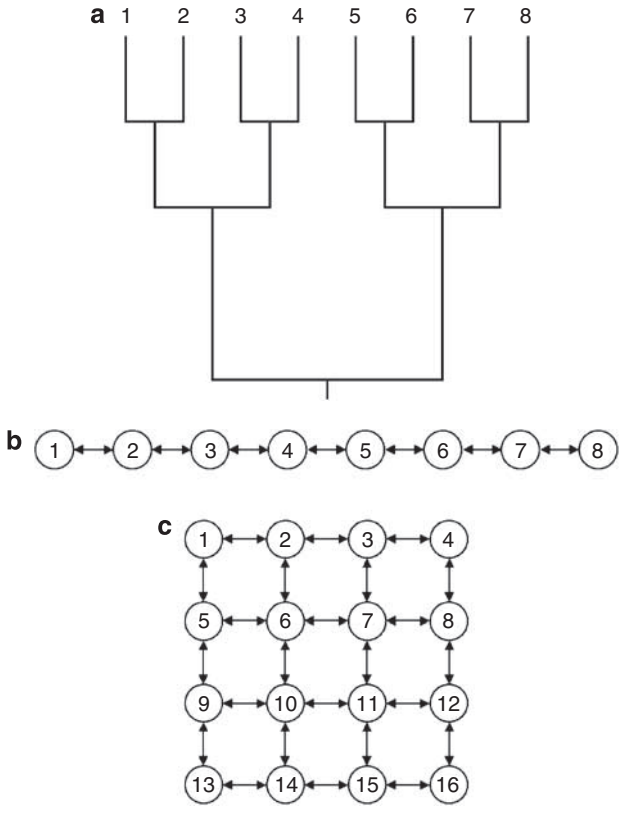

Figure 2 The three evolutionary models used in this investigation (**a**–**c**). The top figure (**a**) shows a hierarchical history of population fragmentation in which an ancestral population is repeatedly split into descendant populations that do not exchange migrants. The middle figure (**b**) shows a linear stepping-stone model of gene flow in which only adjacent populations exchange migrants. The bottom figure (**c**) shows a two-dimensional stepping-stone model of gene flow.

The true genetic distances among natural populations will seldom be known. Therefore, the second question that was explored was how sampling error (of the genetic distance matrix) and tree construction interact.

In the hierarchical model of population fragmentation, genotypes were simulated for eight populations descended from an ancestral population that split into two populations, each of which later simultaneously split into two additional populations, which, in turn, each later split into two populations (Figure 2a). I assumed that population fragmentation was instantaneous and that after fragmentation, populations were completely isolated from each other. The following parameters were used in these simulations. The effective population size for all populations (including each ancestral and descendant populations) was 2000. Divergence times for the populations were (50, 100 and 200), (200, 400 and 800) and (800, 1600 and 3200) generations before sampling, where these numbers refer to the timing of the three fragmentation events that gave rise to the eight populations (Figure 2a).

For the linear stepping-stone model of evolution, genotypes were simulated for eight populations arranged in a line (Figure 2b). Each population received immigrants from adjacent populations (and only adjacent populations), so that the rate of gene flow into populations was m from each neighbor. The two populations at the end of the line received half as many immigrants as the six populations in the middle. The effective population size for each of the eight populations was 1000. The migration rate into each population was varied from 0.01 to 0.0001.

The two-dimensional stepping-stone model (Figure 2c) had a similar pattern of gene flow among 16 populations arranged in a four-by-four grid. The rate of gene flow was varied from 0.01 to 0.001. The effective population size among the populations was 1000.

For all three models, coalescent methods (for example, Hudson, 1991) were used to simulate genotypes for microsatellite loci having a stepwise mutation rate of 2×10^{-4}. Samples were simulated for 100 diploid genotypes at 6, 12 or 24 unlinked loci. Once the data were simulated, Weir and Cockerham's (1984) F_{ST} and Nei's (1978) standard distance (D_S) were used as a pairwise genetic distances. The true genetic distance between each pair of populations in each model was estimated by simulating 100 diploid genotypes for 10 000 independent loci and calculating F_{ST} and D_S for these large data sets. These estimates can be used as true values because with 10 000 loci, sampling error will be negligible.

The UPGMA and NJ trees were constructed for each evolutionary model using the computer programs *TreeFit* (described below) and *TreeView* (Page, 1996). R^2 values were calculated as above (Equation (1)). One thousand samples were simulated for each combination of three parameters (evolutionary model, degree of divergence and number of loci). Lastly, a two-way error decomposition was performed as described below.

When evolutionary trees are constructed from estimates of genetic distances between populations, there are two sources of error: sampling error in the genetic distance matrix, and distortion of this matrix by the tree construction algorithm. The amount of error attributable to each source can be measured when the true genetic

distances between populations is known (for example, in computer simulations). This can be done as follows: let D_{ij} represent the true genetic distance between populations i and j; the sum of squared errors caused by estimating genetic distances ($SSE_{Distance}$) is calculated as

$$SSE_{Distance} = \Sigma_i \Sigma_{j \neq i} (D_{ij} - \hat{D}_{ij})^2 \quad (2)$$

where summation is taken over all pairs of populations. The total sum of squared errors, which contains error caused by estimating genetic distances and making a tree from them, SSE_{Total}, can be calculated as

$$SSE_{Total} = \Sigma_i \Sigma_{j \neq i} (D_{ij} - \hat{d}_{ij})^2 \quad (3)$$

Once $SSE_{Distance}$ and SSE_{Total} have been calculated, the amount of error attributable to the tree construction algorithm, SSE_{Tree}, can be calculated from the difference

$$SSE_{Tree} = SSE_{Total} - SSE_{Distance} \quad (4)$$

If a tree produces estimates of genetic distance that are more accurate than the matrix of genetic distance estimates, SSE_{Tree} will be negative. This would be expected to happen when tree construction successfully 'smoothed' out sampling error in the genetic distance matrix. SSE_{Tree} values are more easily interpreted when they are expressed relative to the genetic distance sampling error.

$$\varepsilon_{Relative} = \frac{SSE_{Tree}}{SSE_{Distance}} \quad (5)$$

The quantities in this error partition were estimated by averaging results across 1000 simulated data sets.

Results

The degree to which evolutionary trees could describe genetic relationships between populations varied greatly, depending on the method used to construct the tree and the evolutionary history of the populations (Table 1 and Figure 3). When the true genetic distance between populations was used to construct trees (Figure 3), and the populations had a strictly hierarchal history of population fragmentation with constant population size, both NJ and UPGMA algorithms produced trees with a perfect R^2 of 1.00. This is not surprising, evolutionary trees describe population structure in a hierarchical manner, and this assumption was met in this model of population fragmentation. The two-dimensional stepping-stone model produced a similarly predictable result. When true genetic distances were used to construct evolutionary trees for the two-dimensional stepping-stone model, R^2 values were low—in the neighborhood of 0.35 for UPGMA trees and 0.80 for NJ trees. This is not surprising because the two-dimensional stepping-stone model is a decidedly unhierarchical model.

The linear stepping-stone model revealed an important difference between the UPGMA and NJ algorithms (Table 1 and Figure 3), and provides a good example of why caution must be used when using evolutionary trees to describe the genetic structure of populations. Populations in the linear stepping-stone model had a nearly perfect isolation-by-distance pattern (not shown), and intuition might suggest that hierarchical evolutionary trees would distort these relationships beyond recognition. For UPGMA trees, this was true. UPGMA trees constructed from true genetic distances had low R^2 values (in the range of 0.45–0.60, Table 1) and topologies that severely distorted the parametric genetic structure of the populations (Figure 3). In contrast, NJ trees showed an almost a perfect fit to the true genetic distance matrix for the linear stepping-stone model with R^2 values usually greater than 0.96. This surprising result is readily explained. Branch lengths in NJ trees are free to vary, and if the terminal branches of the interior populations have lengths of near 0, an NJ tree can successfully depict an isolation-by-distance genetic structure for populations in a linear stepping-stone arrangement (Figure 3).

Similar results were obtained for trees estimated from realistic amounts of data (6, 12 or 24 loci) as for trees constructed from true genetic distances (Table 2). NJ trees constructed from samples had high R^2 values for the hierarchical model and the linear stepping-stone model, and low R^2 values for the two-dimensional stepping-stone model—just as they did when trees were constructed from

Table 1 Parameters for the three evolutionary models used in this investigation

Model	H_{exp}			R^2 *for tree*			
		Range of genetic distances		*UPGMA*		*NJ*	
		F_{ST}	D_S	F_{ST}	D_S	F_{ST}	D_S
Hierarchical fragmentation							
$t = 50, 100, 200$	0.51	(0.01, 0.05)	(0.01, 0.05)	1.00	1.00	1.00	1.00
$t = 200, 400, 800$	0.51	(0.05, 0.14)	(0.05, 0.19)	1.00	1.00	1.00	1.00
$t = 800, 1600, 3200$	0.51	(0.14, 0.29)	(0.19, 0.56)	1.00	1.00	1.00	1.00
Linear stepping stones							
$m = 0.01$	0.71	(0.01, 0.07)	(0.03, 0.20)	0.45	0.46	1.00	1.00
$m = 0.003$	0.69	(0.04, 0.17)	(0.08, 0.50)	0.45	0.51	0.99	1.00
$m = 0.001$	0.64	(0.09, 0.30)	(0.18, 0.91)	0.45	0.55	0.96	0.98
$m = 0.0001$	0.49	(0.34, 0.52)	(0.69, 2.01)	0.46	0.59	0.82	0.89
Two-dimensional stepping stones							
$m = 0.01$	0.80	(0.007, 0.02)	(0.05, 0.12)	0.32	0.30	0.82	0.81
$m = 0.003$	0.78	(0.02, 0.07)	(0.10, 0.27)	0.33	0.32	0.82	0.80
$m = 0.001$	0.74	(0.06, 0.15)	(0.23, 0.60)	0.34	0.37	0.83	0.77

The average expected heterozygosity within populations is denoted by H_{exp}, and the genetic distance of Nei 1978 is denoted by D_S.

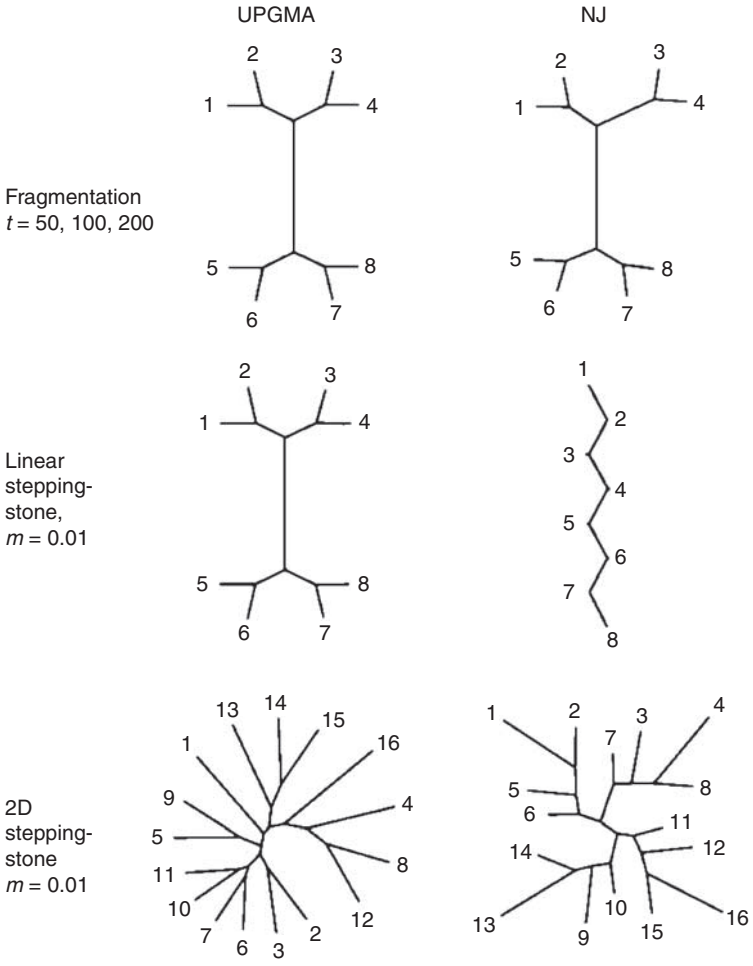

Figure 3 UPGMA and NJ trees constructed from the actual genetic distances (Nei, 1978) between populations in the three evolutionary models examined in this study. NJ, neighbor joining; UPGMA, the unweighted pair group method with arithmetic mean.

Table 2 Average values of R^2 for UPGMA and NJ trees fit to simulated data sets having 6, 12 or 24 loci

Model	*UPGMA*		*NJ*	
	F_{ST}	D_S	F_{ST}	D_S
Hierarchical fragmentation				
$t = 50, 100, 200$	(0.63, 0.82)	(0.62, 0.82)	(0.90, 0.95)	(0.89, 0.95)
$t = 200, 400, 800$	(0.66, 0.85)	(0.64, 0.84)	(0.91, 0.96)	(0.90, 0.96)
$t = 800, 1600, 3200$	(0.66, 0.83)	(0.62, 0.81)	(0.90, 0.95)	(0.86, 0.93)
Linear stepping stones				
$m = 0.01$	(0.54, 0.50)	(0.54, 0.51)	(0.95, 0.98)	(0.94, 0.98)
$m = 0.003$	(0.56, 0.51)	(0.55, 0.52)	(0.94, 0.98)	(0.93, 0.98)
$m = 0.001$	(0.59, 0.53)	(0.56, 0.55)	(0.92, 0.94)	(0.90, 0.95)
Two-dimensional stepping stones				
$m = 0.01$	(0.41, 0.45)	(0.44, 0.40)	(0.76, 0.78)	(0.74, 0.77)
$m = 0.003$	(0.46, 0.40)	(0.43, 0.38)	(0.78, 0.80)	(0.74, 0.76)
$m = 0.001$	([0.48, 0.42)	(0.43, 0.39)	(0.80, 0.80)	(0.71, 0.72)

Each interval shows values for 6 and 24 loci. Results from 12 loci were intermediate. Two genetic distances are used, F_{ST} estimated by Weir and Cockerham's (1984) and the standard distance (D_S) estimated by Nei (1978).

parametric genetic distances. The only notable difference between trees constructed from true distances and trees constructed from estimated genetic distances was the case of UPGMA trees constructed from populations having a hierarchical model of population fragmentation. In this case, UPGMA trees constructed from true genetic distances had an R^2 of 1.0. The R^2 for trees calculated from estimated genetic distances ranged from 0.62 to 0.85, depending on the number of loci sampled. R^2 values were higher when more loci were sampled.

Table 3 The amount error introduced by tree fitting algorithms relative to sampling error accumulated while estimating genetic distances from 12 loci

Model	*UPGMA*		*NJ*	
	F_{ST}	D_S	F_{ST}	D_S
Hierarchical fragmentation, eight populations				
$t = 50, 100, 200$	−0.53	−0.47	−0.12	−0.11
$t = 200, 400, 800$	−0.50	−0.47	−0.10	−0.12
$t = 800, 1600, 3200$	−0.44	−0.48	−0.09	−0.15
Linear stepping stones, eight populations				
$m = 0.01$	2.77	2.25	−0.19	−0.20
$m = 0.003$	2.80	2.26	−0.14	−0.20
$m = 0.001$	2.16	1.33	0.04	−0.20
Two-dimensional stepping stones, 16 populations				
$m = 0.01$	0.37	0.30	0.06	0.06
$m = 0.003$	0.82	0.58	0.17	0.17
$m = 0.001$	0.76	0.40	0.17	0.17

Negative values indicate that the tree fitting algorithm reduced the average squared error (that is, improved estimates; Equation (1)).

The error decomposition showed that the genetic distances within trees could be either more or less accurate than the matrix of genetic distances that the tree was constructed from (Table 3). For example, when UPGMA trees were constructed from populations in a linear stepping-stone structure, the relative error introduced by the tree construction algorithm was usually twice as large as the error caused by the estimation of the genetic distances. This is a substantial distortion of the tree relationships among populations. On the other hand, when populations had a hierarchical history of fragmentation, the estimates of genetic distances produced by UPGMA trees were more accurate than the raw estimates in the genetic distance matrix. In fact, UPGMA trees were able to remove approximately 50% of the squared error of the genetic distance matrix. NJ trees showed similar results. Genetic distances within NJ trees for the hierarchical model were more accurate than the raw matrix, but the degree of improvement was smaller than for the UPGMA trees. This is probably because the inherent flexibility of NJ algorithm allows it to fit branch lengths to sampling error within the genetic distance matrix, and do less smoothing. However, NJ trees were also capable of improving estimates of genetic distance in the linear stepping-stone model. In this case, NJ trees had up to 21% less squared error than the raw estimates of genetic distance. This compares very favorably to the UPGMA trees, which as mentioned above, usually more than doubled the amount of error in the matrix.

In general, results were similar for trees constructed from F_{ST} as for D_S. The only notable distinction is how each genetic distance handled the challenging case of highly differentiated populations. The linear stepping-stone model with very low migration rates ($m = 0.0001$) provides a good case study. D_S was ineffective for describing population structure in this model, because it is undefined when samples do not share any alleles, and this frequently happened when samples had only 6 or 12 loci. F_{ST} had more subtle problems. When the rate of gene flow in the linear stepping-stone model was 0.0001, F_{ST} between neighboring populations was approximately 0.35. Given this level of differentiation, the NJ algorithm would 'want' to separate each pair of adjacent populations by branches having a length of 0.35. If there were eight such populations lined up in a row, this would cause populations 1 and 8 to be separated in the tree by a genetic distance of $7 \times 0.35 = 2.45$. F_{ST}, however, cannot take a value greater than 1.0, so the fitted distance in the tree would be much greater than the genetic distance observed between populations 1 and 8. This reduces the R^2 value for the tree (Tables 1 and 2). The problem is actually more acute than this, because the maximum value of F_{ST} is equal to the homozygosity of the populations being compared (Kalinowski, 2002; Hedrick, 2005), which in this case was 0.51.

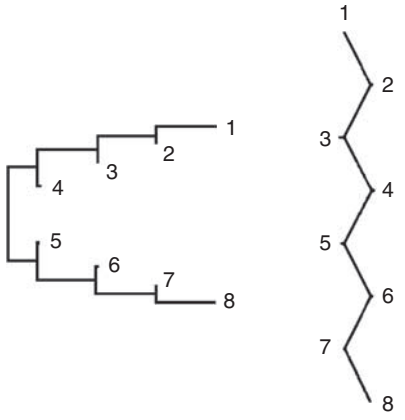

Figure 4 A rectangular (left) and radial (right) NJ tree constructed from the true genetic distances (Nei, 1978) of the linear stepping-stone model ($m = 0.01$). Both trees have the same topology and same branch lengths. NJ, neighbor joining.

An incidental lesson that arose from these results is that the style in which a tree is displayed can affect the ease in which it is interpreted (Figure 4). Unrooted trees (such as UPGMA and NJ) can be displayed in either a radial or a rectangular format (Figure 4). Both types of trees show the same topology with the same branch lengths, but a radial tree can more clearly illustrate isolation by distance.

Software

A Window-based computer program, *TreeFit*, is available from the author for calculating R^2 for UPGMA and NJ trees. *TreeFit* also provides the user with a list of observed and fitted genetic distance for all pairs of populations. This allows the user to construct a scatter plot of these values and identify where the largest discrepancies are. *TreeFit* reads matrixes of genetic distances provided by the user, so any genetic distance can be used to construct trees. The program was checked for accuracy by comparing results to trees described by Felsenstein (2004) and Swofford *et al*. (1996). *TreeFit* is available for download at the author's website www.montana.edu/kalinowski.

Discussion

The results show that evolutionary trees can be a useful tool for describing genetic relationships among populations, but that they must be used with care because populations connected by gene flow may have a genetic

structure that cannot be represented with a tree. Particular caution must be exercised using UPGMA trees. UPGMA trees assume that all the populations clustered in two nodes are equally different from each other, and this assumption can severely distort genetic relationships between populations. Branch lengths in NJ trees are more flexible, and NJ trees can faithfully depict genetic structure for some populations that have an isolation-by-distance population structure. This was seen with the NJ trees constructed for the linear stepping-stone model. However, the flexibility of the NJ algorithm has its limits, and it fails for the two-dimensional stepping-stone model (as does the UPGMA algorithm).

The results suggest that R^2 should be useful for deciding whether to use a tree to summarize population structure. Establishing a threshold for deciding how high R^2 should be to use a tree requires making an arbitrary decision, and I suggest that if R^2 is much less than 0.90, evolutionary trees should probably not be used to describe population structure. More stringent criteria are probably not appropriate because tree construction algorithms can compensate for sampling error and improve estimates of genetic distances within a tree. This is desirable, but reduces the R^2 value of the tree. If the NJ algorithm is used to make trees, this decrease in R^2 is usually less than 0.10. If a less stringent criterion was used for deciding whether to use a tree, this would accept trees constructed from two-dimensional stepping-stones models of evolution, and such trees do not provide a useful picture of population structure. If the R^2 value for a tree is low, other statistical method should be used to describe population structure (for example, multidimensional scaling).

The results from this investigation clearly show that R^2 values should not be used as a test for whether populations have had a hierarchical evolutionary history—especially if the NJ algorithm is used to construct the tree. NJ trees can have a high R^2 even when populations display an isolation-by-distance population structure (for example, linear stepping stones). R^2 should be interpreted, therefore, not as a test for how populations evolved, but as a measure of how well an evolutionary tree summarizes a genetic distance matrix. R^2 measures how well an evolutionary tree describes the genetic structure of a set of populations, but not the evolutionary process that created the structure.

Bootstrapping is frequently used to measure statistical confidence in the topology of evolutionary trees (for example, Felsenstein (2004), chapter 20, and references therein). High bootstrap support and high R^2 values are desirable if a tree is to be used to describe population structure, but they measure different quantities and the distinction is important. The goal of bootstrapping is to assess the statistical support for each interior branch in the tree. The concern is that the topology of the tree has been influenced by sampling error caused by sampling a limited number of loci. If the number of loci genotyped is increased, trees should approach the correct topology and the level of support is expected to increase. The goal of calculating R^2 is to determine whether a tree's topology and branch lengths accurately reflect the genetic distances in the genetic distance matrix. The concern is that imposing a bifurcating topology onto the populations distorts the actual relationships among

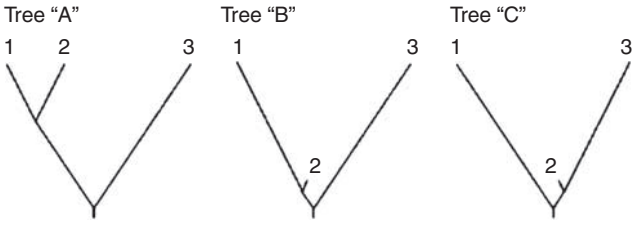

Figure 5 Branches from hypothetical trees that have different topologies and branch lengths.

populations. This value is not expected to increase if more loci are genotyped. A tree could have high bootstrap values, but a low R^2 value. This would happen if a large number of loci were genotyped, and there was consistent population structure across the loci, but the population structure was fundamentally incompatible with a tree topology.

The NJ trees constructed from populations in the linear stepping-stone model illustrate that evolutionary trees do not always cluster populations—at least not in a evolutionarily meaningful way—and that the branch lengths in a tree can be at least as important as a tree's topology. Consider the NJ tree for the linear stepping-stone model (Figure 3). This tree was created by forming a series of clusters, but because of the way branch lengths were assigned, the final tree does not cluster populations into recognizable groups. In fact, what it shows is that most populations are genetically intermediate between their neighbors (which is appropriate, because populations in this model are genetically intermediate between their neighbors). The importance of branch lengths in a tree is illustrated with a second example (Figure 5). In Figure 5, trees, 'A' and 'B' have the same topology, but have very different biological interpretations. Tree 'A' depicts populations 1 and 2 as a cluster, whereas tree 'B' depicts population 2 as having a genetic composition intermediate between populations 1 and 2 (although a little more similar to population 1). Tree 'C' has a topology different from 'A' and 'B', but would often be interpreted the same way as tree 'B'.

There are many reasons to describe the genetic structure of populations, ranging from indentifying management units to inferring the evolutionary processes that gave rise to current patterns of genetic diversity. For each of these purposes, it is helpful to have a simple and accurate summary of the genetic relationships among populations. This is not always an easy task, because genetic data are notoriously multivariate, and populations can have complex patterns of genetic structure. In my experience, it is useful to explore several different methods for describing population structure, including making trees, and then to select a method that conveys the pattern that emerges from these analyses most succinctly. There is some art to this process, and I hope that the R^2 values described in this study will be useful.

Acknowledgements

I thank two anonymous reviewers for thoughtful comments that substantially improved this manuscript. This work was supported by funding from the National Science Foundation (DEB-0717456).

References

Edwards AWF, Cavalli-Sforza LL (1964). Reconstruction of evolutionary trees. In: Heywood VH, McNeill J (eds). *Phenetic and Phylogenetic Classification*. Systamics Association: London. pp 67–76.

Felsenstein J (2004). *Inferring Phylogenies*. Sinauer Associates Inc.: Sunderland, MA.

Gutiérrez-Espeleta GA, Kalinowski ST, WM Boyce WM, Hedrick PW (2000). Genetic variation and population structure in desert bighorn sheep: implications for conservation. *Conserv Genet* **1**: 3–15.

Hedrick PW (2005). A standardized genetic differentiation measure. *Evolution* **59**: 1633–1638.

Hudson RR (1991). Gene genealogies and the coalescent process. *Oxf Surv Evol Biol* **7**: 1–44.

Kalinowski ST (2002). Evolutionary and statistical properties of genetic distances. *Mol Ecol* **11**: 1263–1273.

Nei M (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583–590.

Page RDM (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**: 357–358.

Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.

Sokal RR, Rohlf FJ (1962). The comparison of dendrograms by objective methods. *Taxon* **11**: 33–40.

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1996). Phylogenetic Inference. In: Hillis DM, Moritz C, Mable BK (eds). *Molecular Systematics* 2nd edn, Sinauer Associates Inc.: Sunderland, MA. pp 407–514.

Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.