Reassessment of morphological characteristics in freshwater eels (genus *Anguilla*, Anguillidae) shows congruence with molecular phylogeny estimates

YEONG-SHIN LIN, CHYNG-SHYAN TZENG & JENN-KANG HWANG

Accepted: 2 February 2005
doi:10.1111/j.1463-6409.2005.00192.x

Lin, Y.-S., Tzeng, C.-S. & Hwang, J.-K. (2005). Reassessment of morphological characteristics in freshwater eels (genus *Anguilla*, Anguillidae) shows congruence with molecular phylogeny estimates. — *Zoologica Scripta*, *34*, 225–234.

The morphology-based phylogeny of freshwater eels, proposed by V. Ege in 1939, has been accepted as the basis of eel classification since that time. However, this has been called into question by recent molecular studies. Most of the morphological characteristics recognized by Ege are morphometric. Since methods for the application of morphometric data to phylogeny construction have not been fully established, it is unclear whether the observed discrepancies between morphological and molecular data arise from intrinsic differences or from flawed analyses. Here, we have used two methods to assemble evolutionary trees from distance matrices constructed according to Ege's data, the neighbor-joining (NJ) method and the minimum network (MinNet) method; the latter is based on an evolutionary algorithm. After reanalysing Ege's morphological data, we found that both methods gave results consistent with those based on molecular data, although not with Ege's original classification. Therefore, we speculate that some morphological features Ege used to subdivide the eel groups may not be synapomorphic as he proposed, but symplesiomorphic or convergent . The method developed here may prove useful for constructing phylogeny for taxon groups where only continuous morphometric characteristics are recognized, such as the freshwater eels.

Yeong-Shin Lin & Jenn-Kang Hwang, Department of Biotechnology, National Chiao Tung University, Hsinchu, Taiwan 300, R.O.C. E-mail: leucine.bt90g@nctu.edu.tw or E-mail: jkhwang@faculty.nctu.edu.tw
Chyng-Shyan Tzeng, Department of Life Science, National Tsing Hua University, Hsinchu, Taiwan 300, R.O.C.

Introduction

The concept of numerical taxonomy was an offshoot of biostatistics, and mathematical analysis and computer science methodologies have frequently been used in systematics and evolutionary studies (e.g. Sneath & Sokal 1962; Goodfellow *et al.* 1992; Cohen & Farach 1997; Dietmann *et al.* 2001). However, definitive methods involving the use of geometric morphometric data to estimate phylogeny have yet to be fully established.

The practise of coding original shape data as discrete values based on some criteria and then treating them as input data for cladistic parsimony analyses (e.g. Fink & Zelditch 1995) has been called into question (Rohlf 1998; Adams & Rosenberg 1998). Even though the shape of each anatomical structure could be quantified separately (MacLeod 2002), the problem of the loss of shape information due to the coding procedure remains unresolved.

An alternative approach is to use phylogenetic methods that can utilize morphometric data in their original form rather than forcing them into integer codes (Rohlf 2002). Here we describe a new method based on this principle and use it to reanalyse the phylogeny of freshwater eels using their morphometric characteristics. We attempt to clarify the incongruence between molecular and morphological results (Lin *et al.* 2001a; Aoyama *et al.* 2001).

The eel is a kind of elopomorph fish identified by a unique leptocephalus larval stage. All freshwater eels are classified within the genus *Anguilla* Schrank (Anguillidae). To shed light on the chaotic systematics of this genus, Ege (1939) examined 25 265 specimens, including 12 793 adults and 12 472 elvers, and constructed a phylogenetic synopsis based on 12 morphological characteristics. Species with a short dorsal fin and those without variegated markings were thought to derive from a common ancestor.

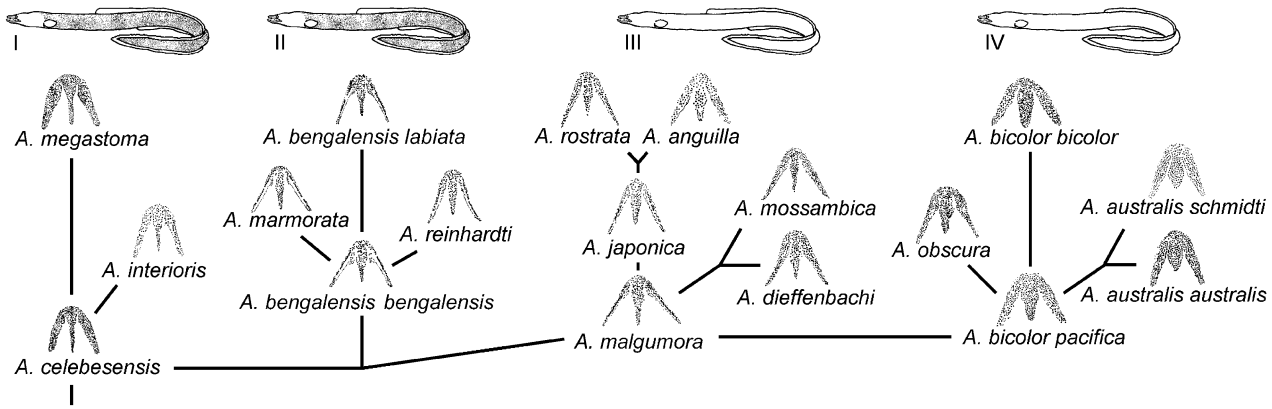

Fig. 1 Ege's (1939) phylogenetic synopsis of the genus *Anguilla*, modified according to Castle & Williamson (1974). This figure demonstrates the variegated markings and dorsal fin types for groups I–IV. Dentitions for each species are also shown. The relationship indicates the traditional view of freshwater eel phylogeny. The branch length is not proportional to divergence time.

Ege subdivided the genus into four groups: (I) variegated species with broad, undivided maxillary and mandibular bands of teeth; (II) variegated species with a toothless, longitudinal groove in the maxillary and mandibular bands of teeth; (III) species without variegated markings and with a long dorsal fin, and (IV) species without variegated markings and with a short dorsal fin.

The phylogenetic relationships thus constructed have been accepted for 60 years (Fig. 1) and even early molecular studies seemed to agree with Ege's phylogenetic synopsis, due to insufficient numbers of specimens and the necessity of analysing short sequences (Aoyama *et al*. 1996; Tsukamoto & Aoyama 1998; Bastrop *et al*. 2000). Lin (1998) proposed a phylogenetic tree, based on the analysis of the mitochondrial cytochrome *b* gene in 10 species, that displayed several significant differences from Ege's phylogeny. This was supported by the results of several other studies (e.g. Lin *et al*. 2001a; Aoyama *et al*. 2001).

When using molecular data to test alternative trees in both maximum parsimony (MP) and maximum likelihood (ML) analyses, the topology from Ege's (1939) phylogeny is always rejected ($P < 0.0001$) (Lin *et al*. 2001a). Lin *et al*. (2001b) reviewed recent molecular analyses and found the following major differences (Fig. 2): (1) none of the four groups proposed by Ege is monophyletic; (2) *A. australis* (Richardson, 1841) is not clustered with other group IV species, but rather with the Atlantic eels; (3) *A. interioris* (Whitley, 1938) is not clustered with other group I species, and (4) *A. reinhardti* (Steindachner, 1867) (group II) and *A. japonica* (Temminck & Schlegel, 1846) (group III) are clustered together, and are neighbours to the group I species *A. celebesensis* (Kaup, 1856) and *A. megastoma* (Kaup, 1856).

The eight morphometric and four numerical characteristics measured by Ege (1939) are continuous. It is surprising

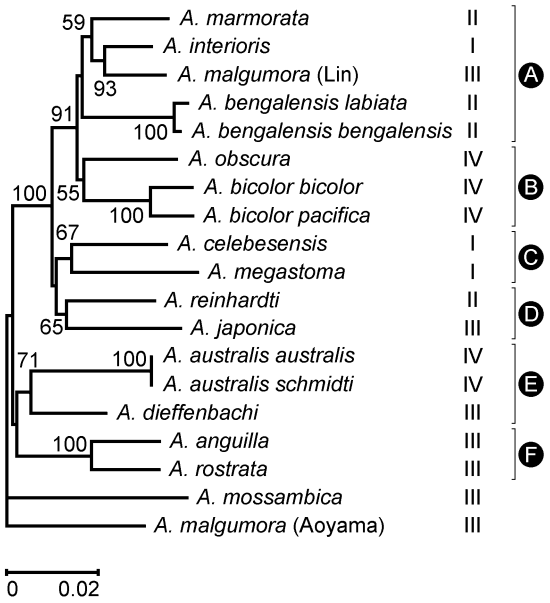

Fig. 2 The neighbor-joining tree (Saitou & Nei 1987) based on the Tamura-Nei distances (Tamura & Nei 1993) from the complete cytochrome *b* and 12S rRNA genes combined with molecular data from Lin *et al*. (2001a) and Aoyama *et al*. (2001), as proposed by Lin *et al*. (2001b). The *A. malgumora* specimens analysed in these two studies may be different species, and are denoted (Lin) and (Aoyama), respectively. The cytochrome *b* gene of *A. bicolor bicolor* in Aoyama *et al*. (2001) is incorrect and was eliminated, as indicated in Lin *et al*. (2002). The numbers at the nodes are bootstrap values (Felsenstein 1985) from 5000 replicates (only those larger than 50 are shown). The scale bar indicates the branch length. The group numbers (I~IV) proposed by Ege (1939) are also shown as well as additional clustering symbols (A~F). The clustering of groups A and B is strongly supported with a bootstrap value of 91, whereas their subsequent divergence is also confirmed by morphological analyses.

that there has been no attempt to construct the phylogeny of *Anguilla* using numerical taxonomy based on these morphological characteristics. One reason for this may be that, among the sepcies, the ranges of most characteristics partially or completely overlap, as indicated by Watanabe *et al*. (2004). In this study, we show that if proper methodologies are adopted, it is possible to reconstruct a morphological tree that is consistent with molecular phylogeny (Lin *et al*. 2001b).

We used two approaches involving the distance matrix and minimum network (MinNet) methods, which we developed as part of this study. Using a distance matrix to construct the dendrogram, we preferred the neighbor-joining (NJ) method (Saitou & Nei 1987), originally designed for molecular data, over traditional methods such as UPGMA (Sokal & Michener 1958), which assume that the evolutionary rate is constant.

Most phenetic methods only represent current similarity relationships among species and, unlike cladistic methods, they do not include historical evolutionary processes. To include this dynamic concept, we developed the MinNet method based on evolutionary computation (EC) to obtain the shortest evolutionary path in the feature space. EC is a generally adaptable concept that solves problems using evolution as an algorithmic tool (Foster 2001) and is especially well-suited for global minimization (e.g. Lewis 1998; Yang & Kao 2000; Quesneville & Anxolabehere 2001; Lemmon & Milinkovitch 2002; Ronen *et al*. 2002). The corresponding coordinates in the resulting evolutionary path represent a possible morphological transformation history for freshwater eels.

Materials and methods

Dataset

We used the 12 morphological characteristics measured by Ege (1939) and listed in Table 1. The first eight are so-called induced characteristics that are designed to eliminate the effects of body growth and make the characteristics stable. However, since the proportions of some characteristics still vary significantly with sexual maturation, we used only characteristic measures for eels of identical body length, t, which implies similar age.

In order to include all possible information, we used linear regression to obtain interpolated values (Fig. 3). The simplest model was adopted since only the frequency tables were available from Ege (1939). We performed one linear regression for each characteristic with the exception of *e/g* (see Table 1), where male and female *A. japonica* were treated separately due to significant differences in regression slope between the sexes. We considered all characteristics for $t = 200$ mm. Eels of this length have passed the elver stage but have not matured sexually. Henceforth, all characteristics referred to are those for $t = 200$ mm.

Table 1 The 12 morphological characteristics.

Morphological characteristics	Notation*
Variation of preanal length, in percentage of total length	*a/t*
Variation of preanal length without head, in percentage of total length	$(a - h)/t$
Variation of distance between verticals through anus and origin of dorsal fin, in percentage of total length	$(a - d)/t$
Variation of predorsal length without head, in percentage of total length	$(d - h)/t$
Variation in length of head, in percentage of total length	*h/t*
Variation in length of gape, in percentage of length of head	*g/h*
Variation of distance from perpendicular through eye-centre on margin of upper jaw to angle of gape, in percentage of length of gape	*e/g*
Variation in length of the intermaxillary-vomerine band of teeth, in proportion to that of the maxillary band	*v/mx*
Number of prehaemal vertebrae	*pv*
Total number of vertebrae	*tv*
Number of branchiostegal rays	*br*
Number of pectoral rays	*pr*

*The notations are: the preanal length (*a*), the total length (*t*), the length of head (*h*), the predorsal length (*d*), the length of gap (*g*), the distance from perpendicular through eye-centre on margin of upper jaw to angle of gape (*e*), the length of the intermaxillary-vomerine band of teeth (*v*), the length of the maxillary band (*mx*), number of prehaemal vertebrae (*pv*), total number of vertebrae (*tv*), the number of branchiostegal rays (*br*), and the number of pectoral rays (*pr*).

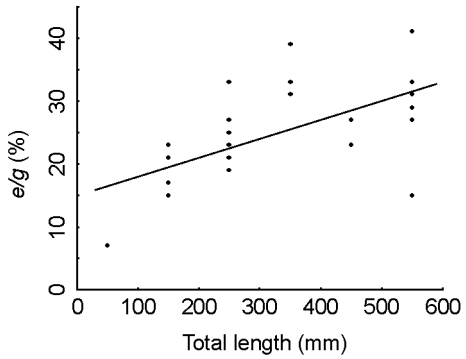

Fig. 3 The relationship between the seventh morphological characteristic, *e/g*, and total body length for *A. celebesensis*. The regression line is also shown. This example illustrates that characteristics may change slightly with maturation stage.

The mean $\mu_{s,c}$ and standard deviation $\sigma_{s,c}$ of the measurements for each species s and characteristic c were calculated (data can be requested from the authors). For each characteristic, the average and standard deviation of $\mu_{s,c}$ over species, denoted by $<\mu_c>$ and $<\sigma_c>$, respectively, were also derived. To weight each characteristic equally, we standardized both $\mu_{s,c}$ and $\sigma_{s,c}$ as follows:

$$\mu^*_{s,c} = \frac{\mu_{s,c} - <\mu_c>}{\sigma_c}$$

$$\sigma^*_{s,c} = \frac{\sigma_{s,c} - <\mu_c>}{\sigma_c}$$

Table 2 Mean $\mu^*_{s,c}$ and standard deviation $\sigma^*_{s,c}$ of the measurements for each species, *s*, and each characteristic, *c*, after standardization.

Species	Morphological characteristics												
	a/t	*(a − h)/t*	*(a − d)/t*	*(d − h)/t*	*h/t*	*g/h*	*e/g*	*v/mx*	*pv*	*tv*	*br*	*pr*	
Anguilla celebesensis (Kaup, 1856)	−0.80 ± 0.04	−0.75 ± 0.03	0.05 ± 0.15	−0.23 ± 0.09	−0.37 ± 0.08	1.41 ± 0.10	0.52 ± 0.27	1.79 ± 0.11	−1.45 ± 0.02	−1.36 ± 0.01	−0.84 ± 0.06	−1.32 ± 0.05	
A. interioris (Whitley, 1938)	−0.17 ± 0.04	0.07 ± 0.04	0.89 ± 0.11	−0.83 ± 0.06	−0.04 ± 0.12	—	—	—	−0.89 ± 0.02	−0.86 ± 0.01	−0.19 ± 0.05	1.36 ± 0.07	
A. megastoma (Kaup, 1856)	−2.20 ± 0.02	−1.55 ± 0.03	0.49 ± 0.10	−0.85 ± 0.05	−1.69 ± 0.05	1.76 ± 0.06	1.33 ± 0.18	0.84 ± 0.07	−0.40 ± 0.02	1.02 ± 0.01	0.03 ± 0.06	−0.80 ± 0.05	
A. bengalensis bengalensis (Gray, 1831)	−1.22 ± 0.03	−1.27 ± 0.05	0.53 ± 0.11	−0.79 ± 0.07	−0.38 ± 0.06	0.25 ± 0.07	0.69 ± 0.20	−0.63 ± 0.07	−0.90 ± 0.02	0.07 ± 0.01	−1.24 ± 0.06	1.32 ± 0.06	
A. bengalensis labiata (Peters, 1852)	−0.12 ± 0.02	−1.23 ± 0.04	0.62 ± 0.12	−0.89 ± 0.06	1.79 ± 0.06	0.96 ± 0.07	1.26 ± 0.16	0.13 ± 0.08	−0.77 ± 0.02	0.61 ± 0.01	−1.69 ± 0.06	0.14 ± 0.05	
A. marmorata (Quoy & Gaimard, 1824)	0.42 ± 0.03	−0.11 ± 0.03	1.51 ± 0.11	−1.43 ± 0.06	0.86 ± 0.06	0.59 ± 0.05	0.21 ± 0.22	−0.62 ± 0.07	−0.66 ± 0.02	−0.81 ± 0.01	−1.31 ± 0.06	1.16 ± 0.05	
A. reinhardti (Steindachner, 1867)	1.35 ± 0.03	0.90 ± 0.03	0.41 ± 0.11	−0.16 ± 0.05	0.96 ± 0.06	−0.63 ± 0.06	−0.29 ± 0.21	−0.26 ± 0.06	0.04 ± 0.01	−0.27 ± 0.01	0.21 ± 0.05	1.02 ± 0.05	
A. malgumora (Kaup, 1856)	0.54 ± 0.03	−0.20 ± 0.03	0.58 ± 0.19	−0.60 ± 0.10	1.19 ± 0.12	1.20 ± 0.04	1.41 ± 0.25	—	−0.97 ± 0.02	−0.84 ± 0.01	−1.19 ± 0.06	—	
A. japonica (Temminck & Schlegel, 1846)	−1.27 ± 0.03	−0.56 ± 0.03	0.09 ± 0.12	−0.21 ± 0.05	−1.43 ± 0.05	−0.97 ± 0.06	−0.73 ± 0.18	0.96 ± 0.08	0.59 ± 0.02	1.74 ± 0.01	0.86 ± 0.06	−0.37 ± 0.06	
A. rostrata (Le Sueur, 1817)	1.59 ± 0.03	1.82 ± 0.03	0.00 ± 0.12	0.45 ± 0.06	0.05 ± 0.06	−1.05 ± 0.05	−1.56 ± 0.20	−0.80 ± 0.08	0.20 ± 0.02	−0.41 ± 0.01	0.27 ± 0.07	−1.31 ± 0.06	
A. anguilla (Linnaeus, 1758)	0.83 ± 0.03	1.66 ± 0.03	0.54 ± 0.15	−0.11 ± 0.08	−0.90 ± 0.06	−0.95 ± 0.07	−1.78 ± 0.30	0.17 ± 0.09	1.42 ± 0.02	1.47 ± 0.01	−0.14 ± 0.06	−0.19 ± 0.06	
A. dieffenbachi (Gray, 1842)	0.76 ± 0.03	0.62 ± 0.03	0.45 ± 0.10	−0.25 ± 0.05	0.40 ± 0.06	0.49 ± 0.05	−0.17 ± 0.16	0.62 ± 0.06	0.92 ± 0.02	0.95 ± 0.01	−0.07 ± 0.05	1.07 ± 0.05	
A. mossambica (Peters, 1852)	−0.18 ± 0.04	−0.45 ± 0.05	1.09 ± 0.11	−1.12 ± 0.07	0.36 ± 0.05	0.62 ± 0.09	−0.13 ± 0.26	−0.84 ± 0.07	−0.99 ± 0.02	−1.49 ± 0.01	−0.54 ± 0.06	1.18 ± 0.06	
A. bicolor pacifica (Schmidt, 1928)	−0.54 ± 0.03	−0.93 ± 0.03	−1.80 ± 0.14	1.46 ± 0.07	0.43 ± 0.05	−1.06 ± 0.06	−0.67 ± 0.29	0.66 ± 0.08	0.33 ± 0.02	−0.43 ± 0.01	1.06 ± 0.06	−1.19 ± 0.06	
A. bicolor bicolor (McClelland, 1844)	0.01 ± 0.03	−0.10 ± 0.04	−1.59 ± 0.12	1.46 ± 0.08	0.21 ± 0.05	−0.45 ± 0.08	0.94 ± 0.23	1.44 ± 0.07	0.45 ± 0.02	0.17 ± 0.01	0.78 ± 0.06	−0.28 ± 0.06	
A. obscura (Günther, 1872)	0.49 ± 0.03	0.17 ± 0.03	−1.06 ± 0.12	1.03 ± 0.07	0.58 ± 0.06	0.18 ± 0.07	0.76 ± 0.21	−0.51 ± 0.06	−0.50 ± 0.02	−1.21 ± 0.01	0.77 ± 0.05	0.05 ± 0.05	
A. australis australis (Richardson, 1841)	−0.63 ± 0.02	0.56 ± 0.03	−1.51 ± 0.12	1.55 ± 0.07	−1.97 ± 0.05	−1.30 ± 0.04	−1.15 ± 0.18	−1.94 ± 0.09	1.91 ± 0.02	0.95 ± 0.01	1.80 ± 0.06	−1.02 ± 0.05	
A. australis schmidti (Phillipps, 1925)	1.14 ± 0.03	1.35 ± 0.04	−1.28 ± 0.12	1.54 ± 0.07	−0.05 ± 0.05	−1.04 ± 0.06	−0.65 ± 0.18	−1.00 ± 0.07	1.67 ± 0.02	0.71 ± 0.01	1.42 ± 0.06	−0.82 ± 0.05	

The standardized results are listed in Table 2. We performed principal component analysis (PCA) to determine the contribution of each characteristic in the feature space, using Gene Cluster (Eisen 1998).

Distance matrix method

Due to some characteristic deficiencies of the species *A. interioris* and *A. malgumora* (Kaup, 1856), we constructed a 12-dimensional feature space containing 16 operational taxonomic units (OTU) according to $\mu^*_{s,c}$ listed in Table 2.

$$OTU_s^{12} = (\mu^*_{s,a/t}, \mu^*_{s,(a-h)/t}, \ldots, \mu^*_{s,pr})$$

We applied the same procedure to feature spaces with fewer dimensions containing species such as *A. interioris* or *A. malgumora*. We computed the Euclidean distance, *D*, between OTUs and constructed the distance matrix, which was then used to construct the phylogenetic tree using NJ.

$$D_{s1,s2} = \sqrt{\sum_c (\mu^*_{s1,c} - \mu^*_{s2,c})^2}$$

Minimum network method

The distance matrix method described above is based on the static traditional phenetic concept, although each coordinate in the feature space is dynamic and represents a morphological state. Hence, during evolutionary history, the morphological transformation can be expressed as a trajectory in the feature space. Therefore, we can construct a network with all the OTUs at external nodes to represent the phylogenetic tree. The problem is to find a network with the shortest tree length to be as close to the actual evolutionary history as possible.

We attempted to find the shortest pathway, *G*, using the EC method. Figure 4 illustrates the process of our evolution strategy, which relies on mutation as the search operator (Foster 2001). First, we randomly picked $n - 2$ internal nodes, *I*, in the feature space. From these nodes, we constructed a network based on the minimum spanning principle, $G = (OTU, I, E)$, where *E* is a set of $2n - 3$ edges such that the sum of *E* is minimized, and all nodes are connected without a cycle while each *OTU* is only connected once. We then generated a mutated node *M* for each internal node.

$$M_{i,c} = N(I_{i,c}, d)$$

The random variable, *N*, is based on a Gaussian distribution whose standard deviation value, *d*, declines with increasing numbers of generations. We then examined each *G* by replacing I_i with M_i, following a reconnection of the network, to check whether the sum of *E* is reduced. We kept the best mutation for the next generation and repeated the whole evolution process until *E* was convergent.

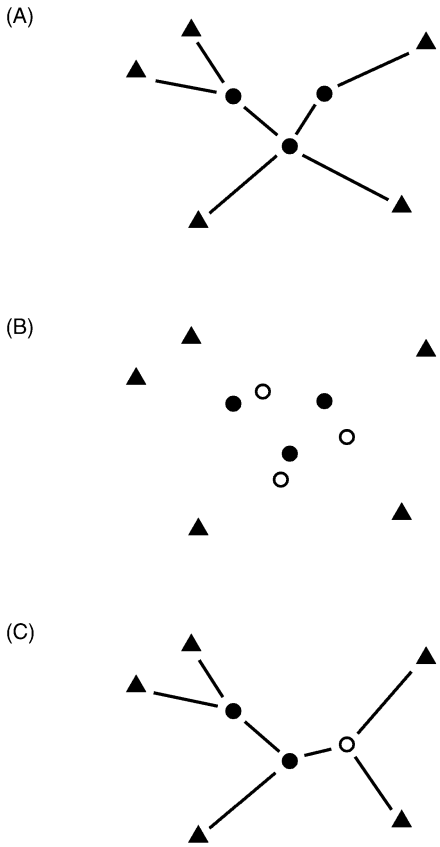

Fig. 4 A–C. Evolution strategy. —A. Triangles represent the OTUs in a feature space. An ideal network contains n OTUs, the external nodes, and $n-2$ internal nodes, symbolized by solid circles. The first step of our evolution strategy was to produce $n-2$ internal nodes randomly in the feature space. Sequentially, these nodes are connected to construct a network based on the minimum spanning principle, with all external nodes connected once. —B. A mutated node, represented by an open circle, is generated for each internal node. —C. These mutations are examined following a reconnection of the network to reveal whether the total tree length is reduced. The best mutation is retained for the next generation so that the mutation and selection process can be repeated.

Random sampling supporting value

We also propose a supporting value to examine the stability of a cluster in a newly constructed phylogenetic tree. This value indicates the probability that the clustering is supported when using the measurements of one randomly selected specimen from each Anguilla population, instead of $\mu^*_{s,c}$ to construct the phylogeny. Each characteristic of these OTUs is randomly assigned as $N(\mu^*_{s,c}, \sigma^*_{s,c})$ (Table 2). Unstable morphological characteristics might also contribute intraspecific diversity and sequentially decrease the supporting value. This random sampling (RS) process was repeated for 5000 replicates. The RS value is presented as a percentage.

Results

Figure 5 presents the evolutionary trees with 16 OTUs and all 12 characteristics constructed using NJ and MinNet. Surprisingly, the group proposed by Ege (1939), containing species with a short dorsal fin (group IV) or those without variegated markings (group III + IV), ceases to be monophyletic. On the other hand, although the clustering of south Pacific and Atlantic eels (group E + F in Fig. 2) is not strongly supported by the molecular phylogeny, the Atlantic eels and *A. australis* are clustered together here and are neighbour to *A. dieffenbachi*, confirming the Central American Isthmus hypothesis of Lin *et al*. (2001a).

Nevertheless, even though groups C and D in Fig. 2 are not stably clustered, they remain arranged around the central area in our newly generated trees. These four species were subdivided into three different groups in Ege's system. Comparing our trees with those of Lin *et al*. (2001b), only *A. mossambica* displays a significantly different position (i.e. clustering with *A. marmorata*). However, this clustering does not support Ege's tree either. The results of PCA (R mode) indicate that only the first seven components contribute more than 5% of the variation (23.85%, 18.58%, 15.61%, 11.12%, 9.17%, 6.57%, 6.25%, 3.70%, 3.36%, 1.45%, 0.28%, and 0.04%), which implies that the feature space can, as an approximation, be simplified as a seven-dimensional space.

To include *A. malgumora* in our analysis, two characteristics, *v/mx* and *pr*, had to be eliminated when constructing the phylogenetic relationships (Fig. 6A). Although *A. malgumora* is clustered with group A, as indicated by Lin *et al*. (2001a), this NJ tree represents a monophyletic group of the short dorsal fin eels — groupIV in Ege's (1939) phylogeny. After eliminating another characteristic $(a-d)/t$, as suggested by Lin (1998), *A. australis* rejoins the Atlantic eels, the RS supporting values are improved, and the tree topology (Fig. 6B) is nearly identical to that shown in Fig. 5A.

The variations contributed by principal components before eliminating $(a-d)/t$ are 29.00%, 20.57%, 18.36%, 11.78%, 7.92%, 4.89%, 4.53%, 2.50%, 0.38%, and 0.07%. After eliminating $(a-d)/t$, the variations are 29.05%, 20.49%, 17.16%, 12.36%, 8.49%, 5.13%, 4.30%, 2.64%, and 0.39%. Comparison of these two data sets shows that $(a-d)/t$ does not provide much additional information.

The original last component, which has a contribution ratio of 0.07% before elimination of $(a-d)/t$, projects mostly at two features $(a-d)/t$ and $(d-b)/t$ (Table 1), with vector values of –0.67 and –0.72, respectively. These results imply that this component is strongly associated with dorsal fin length. After eliminating the three characteristics, *v/mx*, *pr* and $(a-d)/t$, an approximate six-dimensional feature space is yielded, which is insufficient to provide enough resolution for a MinNet tree (data not shown).

The nine available characteristics of *A. interioris* do not appear to be sufficient to resolve the phylogeny (Table 2).

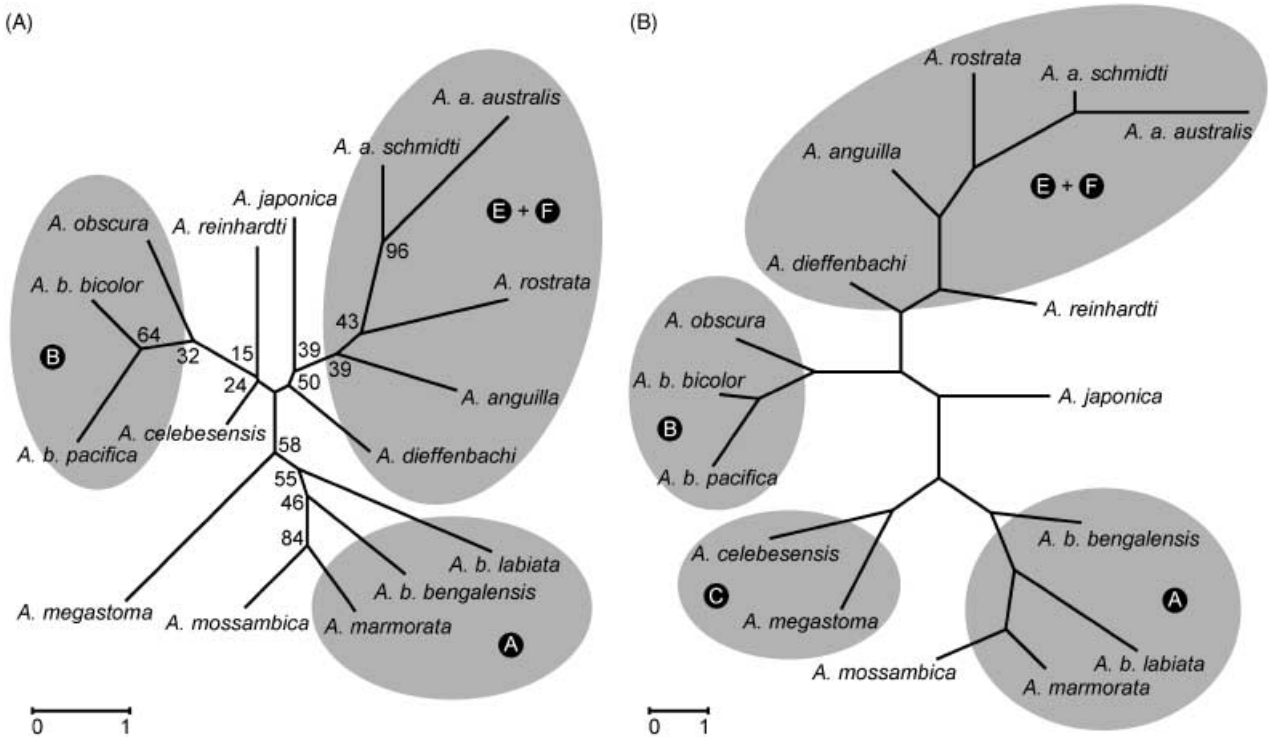

Fig. 5 A, B. Inferred phylogenetic trees using all characteristics and excluding *A. interioris* and *A. malgumora*. Scale bars represent the branch length. The clustering symbols indicated in Fig. 2 are also shown. —A. NJ tree with RS values at the nodes from 5000 replicates. —B. MinNet tree, with population size 10 000, and through 1187 generations.

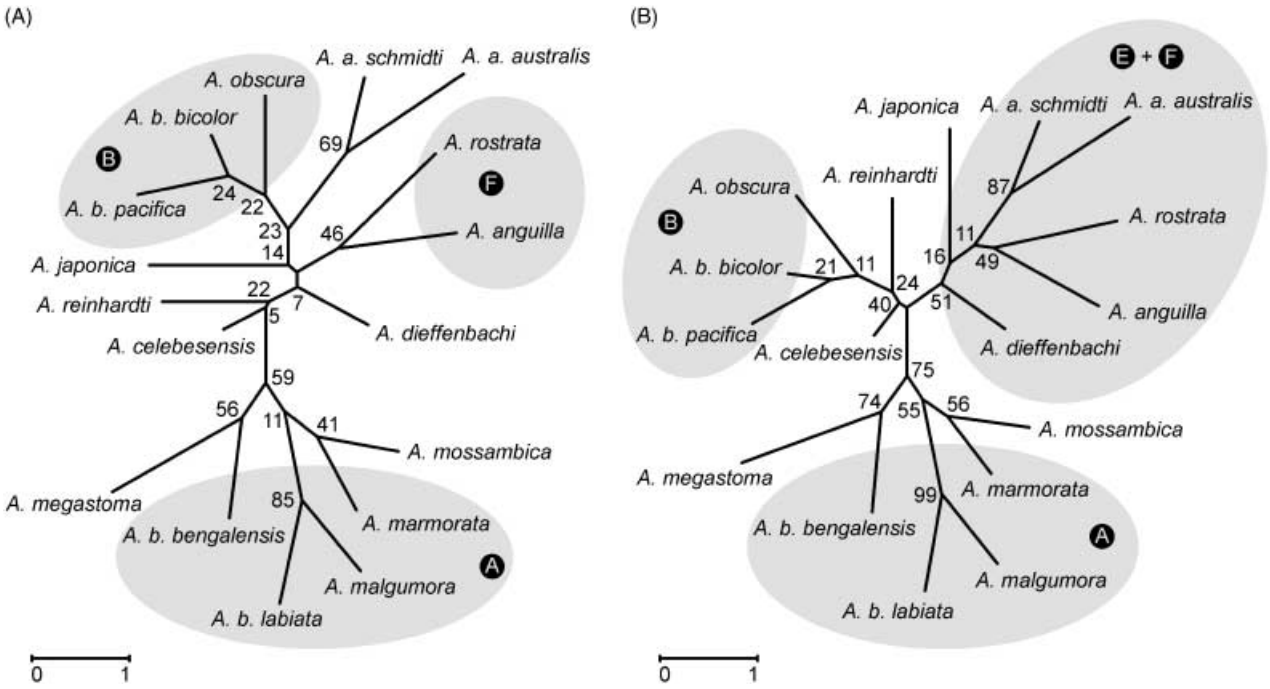

Fig. 6 A, B. The position of *A. malgumora* in the inferred NJ trees. Scale bars represent the branch length, and the numbers at the nodes represent RS values from 5000 replicates. The clustering symbols indicated in Fig. 2 are also displayed. —A. Eliminating two characteristics, *v/mx* and *pr*. —B. Eliminating three characteristics, *v/mx*, *pr*, and $(a - d)/t$.

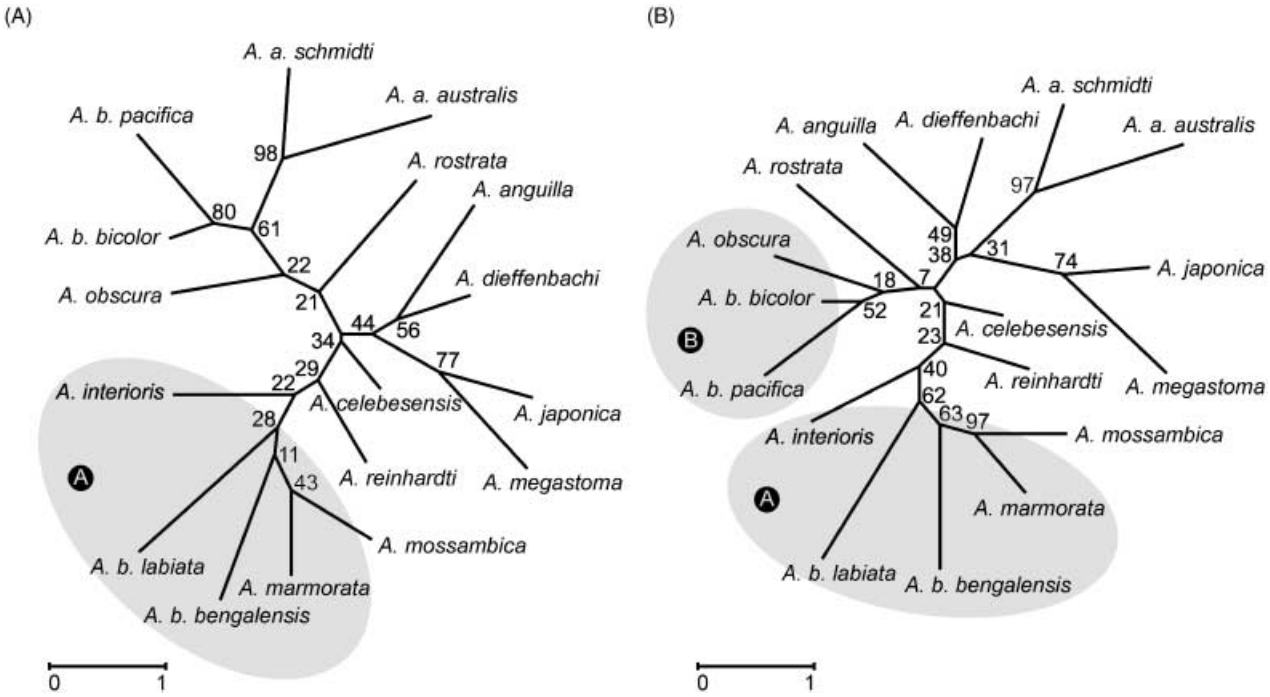

Fig. 7 A, B. The position of *A. interioris* in the inferred NJ trees. Scale bars represent the branch length, and the numbers at the nodes represent RS values from 5000 replicates. The clustering symbols indicated in Fig. 2 are also displayed. —A. Eliminating three characteristics, *g/h*, *e/g*, and *v/mx*. —B. Eliminating four characteristics, *g/h*, *e/g*, *v/mx*, and $(a - d)/t$.

Therefore, the feature space that is generated has only about five dimensions after simplification. The variations contributed by principal components before eliminating $(a - d)/t$ are 28.52%, 21.90%, 19.00%, 11.73%, 10.42%, 4.44%, 3.15%, 0.78%, and 0.06%. After eliminating $(a - d)/t$ they are 28.13%, 22.18%, 18.70%, 11.59%, 10.93%, 4.22%, 3.43%, and 0.82%. However, the clustering of *A. interioris* with group A, described in the molecular phylogenetic results (Fig. 2), could still be recognized in the NJ tree we generated (Fig. 7).

Discussion

In this study we attempted to reconstruct the phylogenetic relationships of the genus *Anguilla* using a morphological data set identical to that used in Ege's (1939) study. However, all four groups proposed by Ege, which were recently confirmed by extensive taxonomic re-examination (Watanabe *et al.* 2004), are nonmonophyletic in our newly generated trees.

Specifically, *A. interioris* (group I), *A. reinhardti* (group II), *A. malgumor*a (group III), *A. mossambica* (group III), and *A. australis* (group IV) always cluster with species from other groups, as indicated in Figs 5–7. This result implies that the characteristics Ege used to subdivide these four groups (i.e. without variegated markings, with a short dorsal fin, or with a toothless longitudinal groove in the maxillary and mandibular bands of teeth), may not be synapomorphic, but rather symplesiomorphic or convergent, as suggested by Lin *et al.* (2001a).

Contrary to Ege's (1939) classification, the molecular tree proposed by Lin *et al.* (2001b) (Fig. 2) and our newly constructed morphological tree have substantial congruence. Ignoring group C + D and *A. mossambica*, the remaining three groups — A, B and E + F — cluster consistently. Assuming that the 12 applied morphological characteristics are selectively neutral, the phylogenetic relationships thus constructed should be more reliable than Ege's subjective tree.

Lin *et al.* (2001a) proposed that variegated markings resulted from convergent evolution and suggested that the short dorsal fin is from an inappropriate impression. The reason is that the positions of the anus and the origin of the dorsal fin developed independently and should not be combined as one homologous characteristic. Our argument therefore concentrates on dentition types, the remaining critical characteristic adopted by Ege.

Ege (1939) noted that *A. australis* differs from all other species in a single feature, the posterior position of the boundary for the beginning of the constriction in the intermaxillary-vomerine band (Fig. 1 and Ege 1939, fig. 40). This lies so far back that the length of the anterior part varies from a little more than 1× to more than 3× the length of the back part. However,

the definition of the constriction is obscure. The tip of a broad intermaxillary-vomerine band always displays a constriction. Therefore, this feature may be a consequence of a short band.

Compared with *A. bicolor* and *A. obscura*, *A. australis* has a significantly shorter intermaxillary-vomerine band, little more than half the length of its maxillary band. On the other hand, *A. rostrata* has a similarly short intermaxillary-vomerine band, although it is not as broad. Considering that only two group III species, *A. malgumora* and *A. japonica*, with a toothless longitudinal groove in the maxillary and mandibular bands of teeth, do not belong to group E or F, we can conclude that groups B, E and F all comprise species with similar dentition types.

The remainder contain species without the toothless longitudinal groove (group I) and those with the groove, i.e. *A. japonica*, *A. malgumora* and group II. Lin *et al.* (2001a) showed that *A. reinhardti* and *A. japonica* share similar dental characteristics, as do *A. bengalensis* and *A. malgumora*. These results are consistent with the clustering of groups A and D. However, group A also includes one group I species, *A. interioris*, which has the most sparse dentition among the group, with some small teeth in the inner boundary of the maxillary and mandibular bands. It seems to have an immature or degenerated groove. *A. malgumora*, clustered with *A. interioris* in Fig. 2, displays a much less obvious groove. These observations imply that the dentition of group A species may have derived from a common ancient type.

The clustering of C + D, where *A. celebesensis* and *A. megastoma* (group C) represent extremely different dentition types compared with *A. reinhardti* and *A. japonica* (group D), is of particular interest. This clustering is unstable in the trees we generated (Figs 5–7), and it is not strongly supported by molecular phylogeny (Fig. 2). Interestingly, groups C + D and A + B form a clade with bootstrap support of 100 in Fig. 2, where their dentition types exhibit a mosaic composition.

The species with a toothless longitudinal groove (groups A and D) are not clustered. Other species without that groove (groups B and C) are also separated. A clade including groups A and B instead has bootstrap support of 91. However, it should be noted that group B has a broad intermaxillary-vomerine band, while group C has an extremely narrow one, and the shapes of their maxillary and mandibular bands of teeth are also different.

Combined with the difference in dentition characteristics between groups A and D, each of these four groups actually has a unique dentition type. Molecular evidence suggests that groups C and D and the clade A + B diverged almost at the same time, followed by the divergence of A and B. Therefore, we suggest that these four dentition types may have been generated from a radiation event.

One consequence of this radiation event is that the phylogenetic relationships among these four groups cannot easily be verified, especially given that the morphological characteristics used to construct the phylogenetic trees in this study are insufficient. Another coincidental phenomenon is the inconsistent localization of groups C and D in the phylogeny. Low RS supporting values also imply that the variations of these characteristics are significant.

The following examples illustrate the deficiency of this morphological data set. Two geographically separated subspecies of *A. australis* were found to be genetically identical (Jellyman 1987; Lin *et al.* 2001b). However, the morphological characteristics of these two populations are statistically different, and consequently they are subdivided in our trees. Molecular data (Lin *et al.* 2001b) also suggest that the two subspecies of *A. bengalensis* are genetically identical, although this result needs to be confirmed by a comprehensive analysis. Nevertheless, these two subspecies have divergent morphological measurements, and each clusters with different species in Fig. 6.

These two cases imply that the morphological characteristics used in this study are not entirely neutral. These features would be altered by environmental factors during their developmental processes, thus providing an explanation for the clustering of *A. mossambica*. Using our geometric tree-construction method, sufficient dimensions of the available feature space are necessary to reveal the relationships between numerous OTUs. Insufficient dimensions would cause some of the separate branches to combine in the tree-searching procedures. *Anguilla mossambica* inhabits the same area as one population of *A. marmorata*. Thus, similar morphological characteristics, derived from similar environmental pressures, may explain the clustering of these two species.

Despite the deficiencies mentioned above, our morphological phylogenies are still generally consistent with previous molecular-based data. The two phenetic tree-construction methods, NJ and MinNet, derived from different logical concepts and applied in this study, show good agreement. One particular advantage of cladistic methods is the representation of characteristic transformation with time, which is absent in traditional phenetic methods. Edwards & Cavalli-Sforza (1964) represented chosen characters by axes in a multidimensional space with the organisms as points, and proposed a tree-construction method, maximum likelihood, based on the probability estimation.

Our MinNet method also utilizes this space concept to display characteristic transformation over time. Compared with the NJ tree, the MinNet representation is more reasonable in that it clusters *A. celebesensis* and *A. megastoma* (Fig. 5). However, the difference is not significant, and the trajectory of the evolutionary path in the MinNet tree does not necessarily reflect the true history.

This description is easily illustrated by the following example. The extinction of one species would generate a totally

different trajectory path in our MinNet tree compared with the original path, although their topologies may be similar. The implication is that the space concept only helps us find a candidate of the possible evolutionary paths. A reliable tree topology can be generated from either NJ or MinNet.

It should also be noted that some characteristics show significant intraspecific variation, and their distributions apparently overlap among species. Such characteristics may have been unstable and therefore easily transformed during evolution. Our RS supporting value is an indicator of the reliability of the clustering based on random sampling processes. A high supporting value implies that the clustering is strongly recommended by the selected morphological characteristics. Comprehensive taxon sampling and the number of neutral characteristics play a crucial role in obtaining a reliable phylogenetic tree.

The methods proposed in this study are extremely useful, especially for some taxon groups, such as the freshwater eels, for which only continuous morphometric characteristics are recognized. Cladistic methods are already efficient and suitable for the remaining groups. Combining molecular and morphological data in this way can significantly improve our understanding of evolutionary history.

Acknowledgements

The work was supported in part by research grants from National Science Council (NSC 93-3112-B-009-001 to J.K.H., and NSC 90-2311-B-007-030 to C.S.T.), and from University System of Taiwan — Veterans General Hospital to J.K.H. We thank Dr Shen-Horn Yen and Dr Jinn-Moon Yang for valuable comments. We are also grateful to the Structural Bioinformatics Core at National Chiao Tung University for their hardware support.

References

Adams, D. C. & Rosenberg, M. S. (1998). Partial warps, phylogeny, and ontogeny: a comment on Fink and Zelditch (1995). *Systematic Biology*, *47*, 168–173.

Aoyama, J., Kobayashi, T. & Tsukamoto, K. (1996). Phylogeny of eels suggested by mitochondrial DNA sequences. *Nippon Suisan Gakkaishi*, *62*, 370–375.

Aoyama, J., Nishida, M. & Tsukamoto, K. (2001). Molecular phylogeny and evolution of the freshwater eel, genus *Anguilla*. *Molecular Phylogenetics and Evolution*, *20*, 450–459.

Bastrop, R., Strehlow, B., Jss, K. & Sturmbauer, C. (2000). A new molecular phylogenetic hypothesis for the evolution of freshwater eels. *Molecular Phylogenetics and Evolution*, *14*, 250–258.

Castle, P. H. J. & Williamson, G. R. (1974). On the validity of the fresh-water eel species, *Anguilla ancestralis* Ege, from Celebes. *Copeia*, *1974* (*2*), 569–570.

Cohen, J. & Farach, M. (1997). Numerical taxonomy on data: experimental results. *Journal of Computational Biology*, *4*, 547–558.

Dietmann, S., Park, J., Notredame, C., Heger, A., Lappe, M. & Holm, L. (2001). A fully automatic evolutionary classification of protein folds: Dali Domain Dictionary, Version 3. *Nucleic Acids Research*, *29*, 55–57.

Edwards, A. W. F. & Cavalli-Sforza, L. L. (1964). Reconstruction of evolutionary trees. In V. H. Heywood & J. McNeill (Eds) *Phenetic and Phylogenetic Classification, vol. 6* (pp. 67–76). Oxford: Systematics Association.

Ege, V. (1939). A revision of the genus *Anguilla* Shaw: a systematic, phylogenetic and geographical study. *Dana Report*, *16*, 1–256.

Eisen, M. (1998). *Gene Cluster* [Computer Software]. Stanford University.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, *39*, 783–791.

Fink, W. L. & Zelditch, M. L. (1995). Phylogenetic analysis of ontogenetic shape transformations: a reassessment of the piranha genus *Pygocentrus* (Teleostei). *Systematic Biology*, *44*, 343–360.

Foster, J. A. (2001). Evolutionary computation. *Nature Reviews Genetics*, *2*, 428–436.

Goodfellow, M., Ferguson, E. V. & Sanglier, J. J. (1992). Numerical classification and identification of *Streptomyces* species — a review. *Gene*, *115*, 225–233.

Jellyman, D. J. (1987). Review of the marine life history of Australasian temperate species of *Anguilla*. *American Fisheries Society Symposium*, *1*, 276–285.

Lemmon, A. R. & Milinkovitch, M. C. (2002). The metapopulation genetic algorithm: An efficient solution for the problem of large phylogeny estimation. *Proceedings of the National Academy of Sciences, USA*, *99*, 10516–10521.

Lewis, P. O. (1998). A genetic algorithm for maximum-likelihood phylogeny inference using nucleotide sequence data. *Molecular Biology and Evolution*, *15*, 277–283.

Lin, Y.-S. (1998). *Phylogenetic Study on Relationships of the Genus Anguilla*. [Masters Thesis]. Hsinchu: National Tsing Hua University.

Lin, Y.-S., Poh, Y.-P., Lin, S.-M. & Tzeng, C.-S. (2002). Molecular techniques to identify freshwater eels: RFLP analyses of PCR-amplified DNA fragments and allele-specific PCR from mitochondrial DNA. *Zoological Studies*, *41*, 421–430.

Lin, Y.-S., Poh, Y.-P. & Tzeng, C.-S. (2001a). A phylogeny of freshwater eels inferred from mitochondrial genes. *Molecular Phylogenetics and Evolution*, *20*, 252–261.

Lin, Y.-S., Poh, Y.-P. & Tzeng, C.-S. (2001b). Which is the dispersal route of the ancient Atlantic eels: Reanalysis of the phylogeny of freshwater eels. *Journal of Taiwan Fisheries Research*, *9*, 161–173.

MacLeod, N. (2002). Phylogenetic signals in morphometric data. In N. MacLeod & P. L. Forey (Eds) *Morphology, Shape and Phylogeny* (pp. 100–138). London: Taylor & Francis.

Quesneville, H. & Anxolabehere, D. (2001). Genetic algorithm-based model of evolutionary dynamics of class II transposable elements. *Journal of Theoretical Biology*, *213*, 21–30.

Rohlf, F. J. (1998). On applications of geometric morphometrics to studies of ontogeny and phylogeny. *Systematic Biology*, *47*, 147–158.

Rohlf, F. J. (2002). Geometric morphometrics and phylogeny. In N. MacLeod & P. L. Forey (Eds) *Morphology, Shape and Phylogeny* (pp. 175–193). London: Taylor & Francis.

Ronen, M., Shabtai, Y. & Guterman, H. (2002). Optimization of feeding profile for a fed-batch bioreactor by an evolutionary algorithm. *Journal of Biotechnology*, *97*, 253–263.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*, 406–425.

Sneath, P. H. A. & Sokal, R. R. (1962). Numerical taxonomy. *Nature*, *193*, 855–860.

Sokal, R. R. & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, *28*, 1409–1438.

Tamura, K. & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, *10*, 512–526.

Tsukamoto, K. & Aoyama, J. (1998). Evolution of freshwater eels of the genus *Anguilla*: a probable scenario. *Environmental Biology of Fishes*, *52*, 139–148.

Watanabe, S., Aoyama, J. & Tsukamoto, K. (2004). Reexamination of Ege's (1939) use of taxonomic characters of the genus *Anguilla*. *Bulletin of Marine Science*, *74*, 337–351.

Yang, J.-M. & Kao, C.-Y. (2000). Flexible ligand docking using a robust evolutionary algorithm. *Journal of Computational Chemistry*, *21*, 988–998.