UC Irvine
UC Irvine Previously Published Works

Title

Genetic change and rates of cladogenesis.

Permalink

https://escholarship.org/uc/item/07g899sr

Journal

Genetics, 81(4)

ISSN

0016-6731

Authors

Avise, JC

Ayala, FJ

Publication Date

1975-12-01

DOI

10.1093/genetics/81.4.757

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, availalbe at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

GENETIC CHANGE AND RATES OF CLADOGENESIS

JOHN C. AVISE* AND FRANCISCO J. AYALA

Department of Genetics, University of California, Davis, California 95616

ABSTRACT

Models are introduced which predict ratios of mean levels of genetic divergence in species-rich *versus* species-poor phylads under two competing assumptions: (1) genetic differentiation is a function of time, unrelated to the number of cladogenetic events and (2) genetic differentiation is proportional to the number of speciation events in the group. The models are simple, general, and biologically real, but not precise. They lead to qualitatively distinct predictions about levels of genetic divergence depending upon the relationship between rates of speciation and amount of genetic change. When genetic distance between species is a function of time, mean genetic distances in speciose and depauperate phylads of equal evolutionary age are very similar. On the contrary, when genetic distance is a function of the number of speciations in the history of a phylad, the ratio of mean genetic distances separating species in speciose *versus* depauperate phylads is greater than one, and increases rapidly as the frequency of speciations in one group relative to the other increases. The models may be tested with data from natural populations to assess (1) possible correlations between rates of anagenesis and cladogenesis and (2) the amount of genetic differentiation accompanying the speciation process.

The data collected in electrophoretic surveys and other kinds of studies can be used to test the predictions of the models. For this purpose genetic distances need to be measured in speciose and depauperate phylads of equal evolutionary age. The limited information presently available agrees better with the model predicting that genetic change is primarily a function of time, and is not correlated with rates of speciation. Further testing of the models is, however, required before firm conclusions can be drawn.

SOME evolutionary groups or phylads contain many more living species than others. A phylad may come to possess more species than another phylad by one or a combination of factors entailing (1) a higher rate of speciation, (2) a lower rate of extinction, or (3) a longer period of time since the origin of the phylad. In some cases, paleontological or other evidence may permit elimination of one or more such alternatives. For example, the evolutionary origin of the highly species-rich (speciose) North American minnows, containing about 250 species, presently traces to the mid-Miocene. By contrast, the species-poor (depauperate) North American sunfish genus Lepomis contains only 11 species, and yet also appears to have originated in the Miocene, or at the Miocene-Pliocene boundary, and there is no evidence suggesting that Lepomis was ever much more speciose than it is today.

* Present address: Department of Zoology, University of Georgia, Athens, Georgia 30602.

Perhaps the paramount question in evolutionary genetics concerns the mechanisms of speciation, and in particular the genetic changes involved. Are rates of anagenesis (phyletic evolution) and rates of cladogenesis (splitting) highly correlated? We might expect this to be the case precisely because populations evolving anagenetically at a faster rate would be likely to speciate more often. Alternatively, rates of anagenesis and of cladogenesis could be positively correlated because the process of speciation *per se*, i.e. the development of reproductive isolation, involved major genetic changes. In this paper, we propose theoretical models allowing tests of this possible correlation between anagenesis and cladogenesis. One set of models assumes that genetic differentiation between species is a function of time, unrelated to the number of speciation events in a phylad. Another set of models assumes that genetic differentiation between species is correlated with the number of cladogenetic events in the group. These models generate distinct and testable predictions of the mean levels of genetic divergence expected between species in speciose *versus* depauperate phylads of equal evolutionary age.

The models are developed to permit meaningful comparisons with the types of data currently being generated by electrophoretic and similar procedures that estimate genetic divergence between species. Possible observations that would lead to falsification of one or the other set of models are suggested.

THE MODELS

An ideal biological model would simultaneously maximize simplicity, generality, reality, and precision. Models dealing with genetic divergence in different groups of organisms are not likely to satisfy all the criteria of an ideal model. Genetic divergence may be affected by a number of factors including time since divergence, number of speciation events, selective regimes, population sizes, mutation rates, and other parameters (AYALA 1974; AYALA and GILPIN 1974), and these factors may differ widely in different lineages. Nonetheless, alternative models designed to allow testing of particular hypothesis for particular sets of data are needed. The models advanced here are of this type. They allow predictions of the mean amount of genetic divergence between species within a group, under two competing assumptions: (1) genetic distance is a function of mean time since divergence, independent of the number of speciations, and (2) genetic distance is a function of the number of speciations in the group. These assumptions lead to distinct predictions about mean levels of genetic divergence in speciose *versus* depauperate phylads.

Barring differences in extinction rates, a phylad may come to possess more species than another phylad of equal evolutionary age by one or both of two processes: (a) speciation events occur more frequently, and (b) more species arise per speciation episode. Examples of alternatives (a) and (b) are pictured in Figures 1 and 2, respectively. These methods of generating speciose and depauperate phylads, and the assumptions involved in the various models advanced below, are summarized in Table 1.

TABLE 1

Factors determining amount of genetic change according to four phylogenetic models

---

Model 1: Amount of genetic change is proportional to time.

    1A: $m_r < m_p$, $l_r = l_p$.

    1B: $m_r = m_p$, $l_r > l_p$.

Model 2: Amount of genetic change is proportional to the number of speciation events.

    2A: $m_r < m_p$, $l_r = l_p$; genetic change is proportional to number of cladogenetic events separating species.

    2B: $m_r = m_p$, $l_r > l_p$; genetic change is proportional to number of species generated at each cladogenetic event.

    The following assumptions apply to all four models: *

    1) cladogenetic events occur at regular time intervals;

    2) no species becomes extinct; and

    3) there are no parallel or convergent genetic changes in different lineages.

The time units between cladogenetic events are m_r and m_p, and the numbers of species generated at each cladogenetic event are l_r and l_p, for species-rich and species-poor phylads respectively.

* Relation of these assumptions does not significantly affect the conclusions derived from the models.
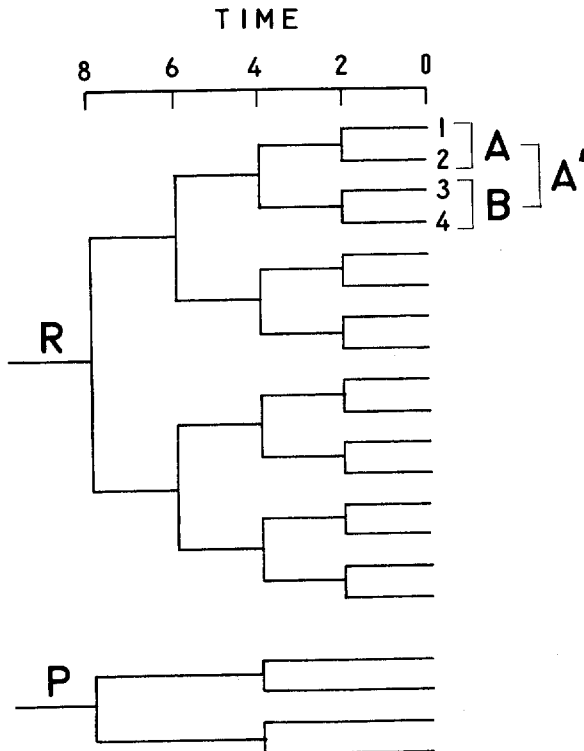

FIGURE 1.—Two phylads, one species-rich (R) and the other species-poor (P), of equal evolutionary age. Cladogenetic events occur every two time units in R ($m_r = 2$), and every four time units in $P(m_p = 4)$. In either one of the phylads, two species arise at each cladogenetic event ($l_r = l_p = 2$). The ratio of mean genetic distances between the two phylads ($\bar{d}_r/\bar{d}_p$) is 1.02 if genetic change is proportional to time (model 1A), but 2.37 is genetic change is proportional to the number of cladogenetic events between species (model 1B).
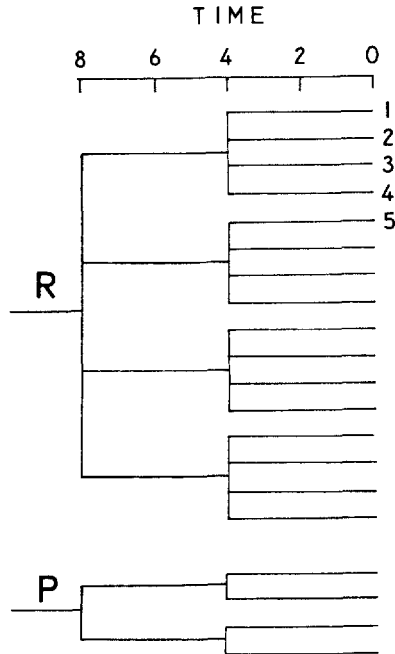
TIME

FIGURE 2.—A species-rich (R) and a species-poor (P) phylad of equal evolutionary age. Cladogenetic events occur every four time units in both phylads $(m_r = m_p = 4)$; four species arise per cladogenetic event in R $(l_r = 4)$, but only two in P $(l_p = 2)$. The ratio of mean genetic distances between the two phylads $(\bar{d}_r/\bar{d}_p)$ is 1.08 if genetic change is proportional to time (model 1B), but 3.90 if genetic change is proportional to the number of species arising per cladogenetic event (model 2B).

Definitions. Given a phylad X_i, let

$t_i \equiv$ number of time units since the origin of the phylad (time is counted in some absolute scale, such as years or multiples thereof); t_i starts one instant before the first speciation or cladogenetic event (clad) in the phylad;

$m_i \equiv$ number of time units between clads (clads are assumed to occur at regular intervals);

$k_i \equiv t_i/m_i$ (k_i is assumed, for simplicity, to be an integer);

$l_i \equiv$ number of species generated at each clad, $l_i > 1$ (l_i is assumed to be constant for a given phylad);

$s_i \equiv$ number of extant species in the phylad;

$d_{ab} \equiv$ genetic distance between two extant species, a and b;

$c_i \equiv$ number of all pairwise comparisons between all extant species in a phylad.

Thus (for simplicity, subscripts are dropped from here on whenever there is no ambiguity):

$$s = l^{t/m} = l^k , \qquad (1)$$

$$c = \frac{l^k(l^k-1)}{2} , \qquad (2)$$

$$\bar{d} = \frac{\Sigma d}{c} . \qquad (3)$$

Model 1.

Genetic differentiation is proportional to time. The genetic distance between two extant species, a and b, is assumed to be directly proportional to the time, t_{ab}, elapsed since the two species last shared a common ancestor, i.e., $d_{ab} = \alpha t_{ab}$. The average distance between all extant species is simply $\bar{d} = \alpha \bar{t}$, where $\bar{t}$ is the average of all t_{ab}'s. Since the constant α is assumed to be the same for all phylads, α can be ignored in the following calculations.

We want to find out, for all pairwise comparisons between extant species in a phylad, the average number of time units since each pair of species had the last common ancestor. Consider a cohort of species (such as A, in Figure 1) with their last common ancestor m time units ago. In such a cohort, there are l species, and thus $l(l-1)/2$ comparisons between species; since there are l^{k-1} such cohorts (A, B, C, . . .), the total distance, in time units, for all pairwise comparisons between species separated m time units before present, is $ml(l-1)l^{k-1}/2 = ml^k(l-1)/2$. Members of different cohorts (A, B, C, . . .) split $2m$ time units ago; there are $l^3(l-1)/2$ comparisons between species that diverged $2m$ units ago in the A' group, and there are l^{k-2} such groups. Therefore the total distance for all pairwise comparisons between species separated $2m$ time units before present is $2ml^3(l-1)l^{k-2}/2 = 2ml^{k+1}(l-1)/2$. In general the total distance, in time units, for all pairwise comparisons between species with the last common ancestor im time units ago is $iml^{k+i-1}(l-1)/2$. The total distance for all possible pairwise comparisons between the species in a phylad is, according to model 1:

$$\Sigma d(1) = ml^k \frac{(l-1)}{2} \sum_{i=1}^{i=k} il^{i-1} \ . \tag{4}$$

To obtain the average distance for all pairwise comparisons, we simply divide (4) by the number of all pairwise comparisons, given by (2):

$$\bar{d}(1) = \frac{ml^k \dfrac{l-1}{2} \sum_{i=1}^{i=k} il^{i-1}}{\dfrac{l^k(l^k-1)}{2}} = \frac{m(l-1) \sum_{i=1}^{i=k} il^{i-1}}{l^k-1} \ . \tag{5}$$

The solution of (5) can be shown to be (see APPENDIX):

$$\bar{d}(1) = m(k - \frac{1}{l-1} + \frac{k}{l^k-1}) \ ; \tag{6}$$

or, since $t = mk$:

$$\bar{d}(1) = t(1 - \frac{1}{k(l-1)} + \frac{1}{l^k-1}). \tag{7}$$

Since l and k are integers generally larger than one, the total value of the terms in parentheses will not be very different from one, and therefore $\bar{d}(1)$ will not be very different from t, the number of time units since the origin of the phylad.

Let R and P be two phylads started at the same time, t times units ago. Let R be a species-rich phylad in which l_r species are formed every m_r time units, and P be a species-poor phylad in which l_p species are formed every m_p time units

(either $m_r < m_p$, and/or $l_r > l_p$). According to model 1, the average distances, in time units, for the two phylads are:

$$\bar{d}(1)_r = t(1 - \frac{1}{k_r(l_r-1)} + \frac{1}{l_r^{k_r}-1}) \text{ , and} \qquad (8)$$

$$\bar{d}(1)_p = t(1 - \frac{1}{k_p(l_p-1)} + \frac{1}{l_p^{k_p}-1}) \qquad (9)$$

Since the value of $\bar{d}(1)$ is not very different from t for both (8) and (9), the ratio $\bar{d}(1)_r/\bar{d}(1)_p$ will be approximately one. In conclusion, then, when the rate of genetic change is proportional to time, the average genetic distance between species should be approximately the same for phylads of equal evolutionary age, independently of the number of species in the phylad.

Numerical values of $\bar{d}(1)$ for phylads with assorted values of t, l and k are shown in Table 2. It can be seen in the table that for any given t, the values $\bar{d}(1)$ are fairly similar for any values of k and l.

Model 2.

Genetic differentiation is proportional to the number of speciation events. We shall consider two alternative forms of this model. First (Model 2A), we shall consider a model in which the rate of genetic differentiation is proportional to the number of cladogenetic events in the phylad; the same numbers of species

TABLE 2

Mean genetic distances, measured in time units, between species in phylads of various ages, when genetic change is proportional to time (models 1A and 1B)

$k = t/m$	t	l				
		2	3	4	5	6
2	4	3.33	3.50	3.60	3.67	3.71
	6	5.00	5.25	5.40	5.50	5.57
	8	6.67	7.00	7.20	7.33	7.43
	10	8.33	8.75	9.00	9.17	9.29
	12	10.00	10.50	10.80	11.00	11.14
3	6	4.86	5.23	5.43	5.55	5.63
	12	9.71	10.46	10.86	11.10	11.26
4	4	3.27	3.55	3.68	3.76	3.80
	8	6.53	7.10	7.36	7.51	7.61
	12	9.80	10.65	11.05	11.27	11.41
5	10	8.32	9.04	9.34	9.50	9.60
6	6	5.09	5.51	5.67	5.75	5.80
	12	10.19	11.02	11.34	11.50	11.60
12	12	11.00	11.60	11.67	11.75	11.80

t = age of phylad; m = time units between cladogenetic events; l = number of species generated per cladogenetic event. The number of extant species per phylad is simply $l^{t/m}$.

The maximum ratio of genetic distances between a species-rich and a species-poor phylad of equal evolutionary age is for any combination of values in the table 1.16; the minimum ratio is 1.004.

are assumed to arise per clad in speciose and in depauperate phylads $(l_r = l_p)$, although clads occur more often in the former than in the latter $(m_r < m_p)$. Secondly (Model 2B), we shall assume that the rate of genetic differentiation is proportional to the number of species arising per clad $(l_r > l_p)$, but clads occur with equal frequency in speciose and in depauperate phylads $(m_r = m_p)$. One of these two versions of model 2, or a combination of both, could be the case if it were true that phylads in which genetic change occurs at a faster rate (for whatever reason) would speciate more often. Models 2A and/or 2B could also apply if it were the case that speciation events (for whatever reason they might come about) would entail a "genetic revolution," i.e., a substantial amount of genetic change.

Model 2A: Let genetic distance between two species be proportional to the number of clads separating two species, i.e., $d(2A)_{ab} = \beta \cdot$ clads between a and b. Thus in Figure 1, $d(2A)$ between species 1 and 2 is equal to one, between species 1 and 3 is equal to three, and so on. Since β is assumed to be the same for all phylads, is may be dropped from the calculations, so that genetic distances will be expressed simply in terms of the number of clads between any two species.

Consider again Figure 1. The species in cohort A are separated by one clad; there are $l(l-1)/2$ comparisons per cohort, and l^{k-1} such cohorts; the total genetic distance for all pairwise comparisons between species separated m time units before the present is $l(l-1)l^{k-1}/2 = l^k(l-1)/2$. Species whose last common ancestor was $2m$ units before present are separated by three clads; there are $l^3(l-1)/2$ comparisons within each such group of species, and l^{k-2} groups; the total genetic distance between pairs of species separated by three clads is $3l^3(l-1)l^{k-2}/2 = 3l^{k+1}(l-1)/2$. In general the total distance, in clads, between pairs of species with the last common ancestor im time units before present is $(2i-1)l^{k+i-1}(l-1)/2$. The total distance for all pairwise comparisons between species in a phylad is, according to model 2A:

$$\Sigma d(2A) = l^k \frac{l-1}{2} \sum_{i=1}^{i=k} (2i-1)l^{i-1} . \tag{10}$$

Dividing by the number of pairwise comparisons, given by (2), we obtain:

$$\bar{d}(2A) = \frac{l^k \dfrac{l-1}{2} \sum\limits_{i=1}^{i=k} (2i-1)l^{i-1}}{\dfrac{l^k(l^k-1)}{2}} = \frac{(l-1) \sum\limits_{i=1}^{i=k} (2i-1)l^{i-1}}{l^k-1} . \tag{11}$$

The solution of (11) is shown in the APPENDIX to be:

$$\bar{d}(2A) = 2k - \frac{l+1}{l-1} + \frac{2k}{l^k-1} . \tag{12}$$

Therefore, since $t = mk$:

$$\bar{d}(2A) = \frac{2t}{m} - \frac{l+1}{l-1} + \frac{2t/m}{l^{t/m}-1} . \tag{13}$$

If t is large relative to m,

$$\bar{d}(2A) \simeq \frac{2t}{m} - \frac{l+1}{l-1} \ . \tag{13a}$$

Let R and P be two phylads of equal evolutionary age, t, such that $m_r < m_p$ and $l_r = l_p$; that is, the number of species arising per clad is the same in both phylads, but cladogenetic events occur more frequently in the speciose phylad, R, than in the depauperate phylad, P. According to model 2A, the average distances, in clads, are for the two phylads:

$$\bar{d}(2A)_r = \frac{2t}{m_r} - \frac{l+1}{l-1} + \frac{2t/m_r}{l^{t/m_r}-1} \ , \text{ and} \tag{14}$$

$$\bar{d}(2A)_p = \frac{2t}{m_p} - \frac{l+1}{l-1} + \frac{2t/m_p}{l^{t/m_p}-1} \ . \tag{15}$$

If t is large relative to m_p, the ratio $\bar{d}(2A)_r/\bar{d}(2A)_p$ will be larger than but not be very different from m_p/m_r, or k_p/k_r. The ratio between the mean genetic distances in two phylads is, approximately, the reciprocal of the ratio of the number of time units between clads in the two phylads.

Numerical values of $\bar{d}(2A)$ for phylads with various k and l values are given in Table 3. It can be seen in that table that for any l, the ratio between any two $\bar{d}(2A)$'s is larger than the ratio between the corresponding k's (although it approaches the latter ratio as k becomes larger), and may be as large as 9.02 for the values in Table 3. In contrast, according to model 1 the mean distance ratio of speciose to depauperate phylads is always close to one.

Model 2B: Let genetic distance between two species be a function of the number of species arising per clad, so that $d(3A)_{ab} = \gamma l^{k'}$, where $k' = t'/m$, and t' is the number of time units since the two species last shared a common ancestor. For example, in Figure 2 the genetic distance between species 1 and 2 is $\gamma l^1 = 4\gamma$, between species 1 and 5 is $\gamma l^2 = 16\gamma$, and so on. (The proportionality

TABLE 3

Mean genetic distances between species when genetic change is proportional to the number of cladogenetic events separating species (Model 2A)

l	$k = t/m$							
	2	3	4	5	6	8	10	12
2	2.33	3.86	5.53	7.32	9.19	13.06	17.02	21.01
3	2.50	4.22	6.10	8.04	10.01	14.00	18.00	22.00
4	2.60	4.43	6.36	8.34	10.34	14.33	18.33	22.33
5	2.67	4.55	6.51	8.50	10.50	14.50	18.50	22.50
6	2.71	4.63	6.61	8.60	10.60	14.60	18.60	22.60

Symbols, and numbers of extant species in a phylad, as in Table 2.
The maximum ratio of genetic distances between a species-rich and a species-poor phylad with equal number of species formed per clad (l) is 9.02 (for $l = 2$; $k = 12$ and 2); the minimum ratio is 1.23 (for $l = 6$, $k = 6$ and 5). Compared phylads are assumed to be of equal evolutionary age ($t_r = t_p$).

constant γ may be dropped, as before.) Given a species-rich phylad, R, and a species-poor phylad, P, it is assumed that $l_r > l_p$, but $m_r = m_p$.

Proceeding as before, we obtain:

$$\bar{d}(2B) = \frac{\Sigma d(2B)}{c} = \frac{l^{k-1}\dfrac{l-1}{2}\overset{i=k}{\underset{i=1}{\Sigma}}l^{2i}}{\dfrac{l^k(l^k-1)}{2}} = \frac{(l-1)\overset{i=k}{\underset{i=1}{\Sigma}}l^{2i}}{l(l^k-1)} \;. \tag{16}$$

The solution of (16) is shown in the APPENDIX to be:

$$\bar{d}(2B) = \frac{l(l^k+1)}{l+1} + \frac{l-1}{l(l^k-1)} \;. \tag{17}$$

If k is large, the second term can be ignored, and $\bar{d}(2B)$ will be approximately $(l^{k+1}+l)/(l+1)$. Given two phylads, P and R, such that $l_p > l_r$ and $k_p = k_r$, the ratio of their mean genetic distances will be, approximately:

$$\frac{\bar{d}_r(2B)}{\bar{d}_p(2B)} \simeq \frac{(l_r^{k+1}+l_r)/(l_r+1)}{(l_p^{k+1}+l_p)/(l_p+1)} \;. \tag{18}$$

Numerical examples of $\bar{d}(2B)$ for various k and l values are given in Table 4. The ratio of mean genetic distances between a speciose and a depauperate phylad will be ratios between values in a given row of Table 4.

The results of the various models are summarized in Table 5. When genetic distance is a function of time, the ratio of mean genetic distances between species in speciose *versus* depauperate phylads of equal evolutionary age approximates 1, no matter whether the greater number of species in the speciose phylad is due to more frequent speciation events per lineage (model 1A), or to more species being generated at each cladogenetic episode (model 1B). On the other hand, when genetic distance is a function of the number of speciation events in the group, the ratio $\bar{d}_r/\bar{d}_p$ is greater than one and increases as the frequency of spe-

TABLE 4

Mean genetic distances between species when genetic change is proportional to the number of species generated per clad (Model 2B)

$k = t/m$	l		
	2	3	4
2	3.50	7.58	13.65
3	6.07	21.03	52.01
4	11.37	61.51	204.80
5	22.02	182.25	819.20
6	43.34	546.75	3276.80

Symbols, and number of extant species in a phylad, as in Table 2.

The maximum ratio of genetic distances between a species-rich and a species-poor phylad with equal k is 75.6 ($k = 6$, $l = 4$ and 2); the minimum ratio is 1.68 ($k = 2$, $l = 4$ and 3). Compared phylads are assumed to be of equal evolutionary age ($t_r = t_p$).

TABLE 5

Summary of the results predicted according to the various models

Differences between species-rich and species-poor phylads	Genetic distance proportional to	
	(1) time	(2) number of speciations between species (model 2A), or number of species generated per clad (model 2B)
$m_r < m_p;\ l_r = l_p$	Model 1A $\bar{d}_r/\bar{d}_p \simeq 1$	Model 2A $\bar{d}_r/\bar{d}_p \gg 1$
$m_r = m_p;\ l_r > l_p$	Model 1B $\bar{d}_r/\bar{d}_p \simeq 1$	Model 2B $\bar{d}_r/\bar{d}_p \gg 1$

The body of the table gives the ratios of mean genetic distances $(\bar{d}_r/\bar{d}_p)$ between a species-rich and a species-poor phylad. m_r and m_p are the time units between cladogenetic events, and l_r and l_p are the numbers of species generated per clad, in species-rich and species-poor phylads respectively.

ciations (or as the number of species generated per cladogenetic episode) becomes greater in one group relative to another.

RESULTS

The models predict that if the rate of genetic change in evolution is primarily a function of time, and is not correlated with the rate of cladogenesis, the average degree of genetic differentiation between living species would be approximately the same in speciose as in depauperate phylads of similar evolutionary age. On the contrary, if rates of anagenesis and of cladogenesis are strongly correlated, species should on the average be more genetically different in species-rich than in species-poor phylads. Clearly these alternatives could be tested by studying genetic differentiation in groups of organisms in which speciose and depauperate phylads of comparable age can be identified. At present, few such data are available.

The Hawaiian Drosophila could provide excellent materials to test the models suggested above. Most species are endemic to a single island, and upper limits to their evolutionary ages may be reasonably inferred from known geological histories of the islands. Three monophyletic assemblages of species provide examples of speciose, intermediate, and depauperate phylads. The crassifemur group contains only six known species, the white-tipped scutellum group contains about 60 species, and the group of picture-winged flies consists of approximately 120 species. Thirty-one genetic loci have been examined in four picture-winged species and in three white-tip species all from a single island (Table 6; AYALA 1975). Among the four picture-winged species $\bar{d}_r = 0.821 \pm 0.129$, and among the three white-tipped species $\bar{d}_p = 0.715 \pm 0.195$ (d calculated using NEI's [1972] method). The ratio of $\bar{d}_r/\bar{d}_p = 1.15$ would seem to favor model 1, but the small number of species studied as well as many uncertainties make it unwarranted to draw any conclusions from this limited information.

TABLE 6

Mean genetic distances between seven species of Drosophila from Hawaii, based on a sample of 31 gene loci studied by electrophoresis (Ayala 1975)

Species	2	3	4	5	6	7
Picture-winged species:						
1. *D. planitibia*	0.854	1.113	1.241	2.712	2.244	1.781
2. *D. adiostola*		0.737	0.560	1.738	1.753	1.565
3. *D. truncipenna*			0.421	2.101	2.090	1.990
4. *D. hamifera*				1.934	1.739	1.816
White-tipped species:						
5. *D. nigra*					0.331	0.850
6. *D. nigella*						0.964
7. *D. dolichotarsis*						—

Distances calculated according to the method of NEI (1972).

A set of data which is at present more complete concerns two groups of fish, the speciose minnows and the depauperate sunfish. Most of the 250 species of North American minnows belong to the subfamily Leuciscinae, which first appears in the fossil record in North America during the Miocene. A total of nine species of minnows native to California have been assayed at 24 genetic loci; $\bar{d}_r = 0.568 \pm 0.052$ (Table 7; AVISE and AYALA 1975). These nine species share geographic range but are probably not a monophyletic assemblage. Some of the species have their closest relatives among the eastern United States fauna. Furthermore, one of the species examined, *Notemigonus crysoleucus*, is a probable outlier among the North American minnows and likely belongs to a distinct subfamily, the Abramidinae. The maximum genetic distance observed between any pair of minnows examined involves *Notemigonus crysoleucus* and *Pogonichthys macrolepidotus*, where $d = 1.12$. On the average, the level of divergence among minnows native to California may approximate the mean genetic distance between all North American minnows (AVISE and AYALA 1975).

TABLE 7

Mean genetic distances between nine species of California minnows, based on a sample of 24 gene loci studied by electrophoresis (Avise and Ayala 1975)

Species	2	3	4	5	6	7	8	9
1. *Hesperoleucus symmetricus*	0.055	0.095	0.194	0.518	0.705	0.432	0.251	0.901
2. *Lavinia exilicauda*		0.147	0.216	0.616	0.746	0.519	0.354	0.919
3. *Mylopharodon conocephalus*			0.131	0.546	0.600	0.453	0.174	0.790
4. *Ptychocheilus grandis*				0.541	0.600	0.526	0.333	0.989
5. *Orthodon microlepidotus*					1.079	0.776	0.518	1.094
6. *Pogonichthys macrolepidotus*						0.519	0.679	1.118
7. *Richardsonius egregius*							0.443	0.976
8. *Gila bicolor*								0.884
9. *Notemigonus crysoleucus*								—

Distances calculated according to the method of NEI (1972).

Ten of the eleven extant species of sunfish, Lepomis, were assayed for genetic differentiation at 14 loci (AVISE and SMITH 1974). Using these data, AVISE and AYALA (1975) calculated genetic distances according to NEI's (1972) formula (Table 8). For all possible pairwise comparisons of Lepomis, $d_p = 0.627 \pm 0.029$. The maximum observed genetic distance, that between *Lepomis humilis* and *L. marginatus*, is $d = 1.02$. The genus Lepomis appears first in the fossil record near the Miocene-Pliocene border, and apparently has never contained many species.

Both Lepomis and the Leuciscinae are of similar evolutionary age; one phylad contains many more species than the other. Yet the ratio $\bar{d}_r/\bar{d}_p = 0.91$. These results are in fair agreement with the predictions of model 1 but conflict with model 2. For these fishes, time appears to be a more important determinant of levels of genetic differentiation than does the number of speciations in a phylad. Another prediction can be made from the models. If genetic distance is a function of time (model 1), the ratio of maximum distances observed between any two species in speciose *versus* depauperate phylads should approximate 1, if the phylads are of comparable evolutionary age. On the other hand, if genetic distance is correlated with the number of speciations (model 2), the maximum distance between species should be much greater in the speciose phylad. As pointed out above, the maximum d's observed are 1.12 and 1.02 for the minnows and Lepomis, respectively; their ratio equals 1.10. Again, these results are in good agreement with model 1.

The results just presented are far from satisfactory. First, there is the question of the actual evolutionary ages of the two groups. In particular, it seems likely that the speciose minnows may be older than Lepomis. This bias would, nevertheless, make the data even more inconsistent with model 2; if the minnows are older than presently suggested they should be more genetically different than Lepomis for this reason as well, but they are not. Perhaps a more serious difficulty concerns the true number of cladogenetic events involved in the two phylads. At present, the minnows are highly speciose, Lepomis fairly depauperate.

TABLE 8

Mean genetic distances between ten species of Lepomis, based on a sample of 14 gene loci

Species	2	3	4	5	6	7	8	9	10
1. *L. humilis*	0.574	0.848	0.543	0.991	0.875	0.553	0.794	1.019	0.757
2. *L. microlophus*		0.661	0.980	0.512	0.706	0.416	0.601	0.502	0.948
3. *L. auritus*			0.726	0.523	0.664	0.266	0.333	0.542	0.710
4. *L. gulosus*				0.723	0.682	0.660	0.714	0.679	0.378
5. *L. megalotis*					0.579	0.416	0.414	0.160	0.711
6. *L. cyanellus*						0.518	0.810	0.530	0.784
7. *L. punctatus*							0.329	0.426	0.673
8. *L. gibbosus*								0.569	0.685
9. *L. marginatus*									0.670
10. *L. macrochirus*									—

Distances have been calculated according to NEI's (1972) method from data by AVISE and SMITH (1974).

It seems likely that a greater number of cladogenetic events have occurred in the minnows than in Lepomis. Yet, it is conceivable that the rate of speciation may have been at least as large in Lepomis as in the minnows, and that many more extinctions have taken place among Lepomis. This possibility cannot be excluded altogether. One more difficulty is that we might have sampled a disproportionately large number of closely related minnow species, and thus underestimated the average genetic distance in the group. However, if we exclude the comparisons among the four most closely related minnow species, the mean genetic distance in the group becomes $\bar{d} = 0.65$, which is not significantly larger than the mean genetic distance among Lepomis species ($\bar{d} = 0.63$).

KOHNE, CHISCON and MOYER (1972) have studied the rates of DNA change in various primates by hybridizing nonrepetitive DNA sequences. Their "*very* tentative conclusion" is that rate of nucleotide sequence divergence was much higher during the period when the fossil record indicates that extensive speciation occurred in the primates, relative to other periods. That is, apparently the rates of cladogenesis and anagenesis are positively correlated in the evolution of the primates. Thus, the "*very* tentative conclusion" based on the primate DNA data is inconsistent with the no less tentative conclusion based on the minnows *versus* Lepomis comparison or in the Hawaiian Drosophila.

Clearly, no general conclusions can be drawn concerning the correlation or lack thereof between rates of anagenesis and cladogenesis until the models advanced here are tested in a variety of groups of organisms. We know of no other sets of data giving genetic distances in speciose and depauperate phylads of similar evolutionary age. However, such information would not be very difficult to obtain. We have pointed out above that the Hawaiian Drosophila provide excellent materials to test the models advanced here. Among the fishes, a promising group for comparison with the depauperate Lepomis is the speciose minnow genus Notropis, which dates from the mid-Pliocene (MILLER 1965). Phylads of about equal evolutionary age but different rates of cladogenesis undoubtedly exist in many other groups of organisms as well.

DISCUSSION

Two processes of differentiation occur in evolution. Anagenesis or phyletic evolution refers to changes occurring through time in a single line of descent. Cladogenesis or splitting involves the formation of independent lines of evolutionary decent. It is of general importance to know whether or not these two processes are positively correlated. If anagenesis and cladogenesis were known to be correlated, evolutionary knowledge would advance by a large leap, since inferences could be made from known rates of cladogenesis to relative rates of anagenesis and *vice versa*. Of course, if anagenetic and cladogenetic evolution were correlated, we would want to find out the direction of the cause-effect relationship. At least two alternatives are possible in this respect. One is that whenever anagenetic change is occurring at a faster rate (for whatever reasons), there would be a greater probability for speciation events (cladogenesis) to occur. In

such case, rates of cladogenesis would be the effect of rates of anagenesis. The second alternative is that speciation *per se* may entail a substantial amount of genetic change. Any process that increases rates of speciation would also increase rates of anagenesis. Changes in rates of anagenesis would, then, be the effect of changes in rates of cladogenesis.

We have advanced simple and general models to ascertain whether or not rates of anagenesis and cladogenesis are correlated. The models advanced are fairly robust. They lead to qualitatively distinct predictions about ratios of mean genetic distances in speciose *versus* depauperate phylads. The models obtain for a wide variety of cases, and are realistic because they are relatively insensitive to relaxations of the underlying assumptions (Table 1). We may relax the assumption that no parallel or convergent genetic changes occur in different lineages, as long as the proportion of the genome that is free to diverge is large and is roughly equal in the two phylads. The models are not strongly dependent either upon the assumption that cladogenetic events occur at exactly regular time intervals, particularly when phylads are old and contain many species.

In obtaining data to test the models, it may be impossible to sample speciose phylads exhaustively. For example, we accepted a sample of nine species of minnows as representative, to a first approximation, of mean genetic divergence among the Leuciscinae. An ideal sample of a speciose phylad is large and random with respect to level of genetic divergence. The problems created by sampling bias are similar to those due to nonrandom extinctions. If extinctions are random events and occur at equal rates in speciose and depauperate phylads, they will not seriously alter the predictions of the models. However, a phylad may appear depauperate because of a high rate of extinction, despite a large number of cladogenetic events in its history.

Although the models are simple, general and robust, they are not precise. Genetic divergence is likely to be affected both by the time elapsed and by the number of speciations, as well as by other factors. But we are interested in main effects. The ratio of mean genetic divergence in speciose *versus* depauperate phylads approximates 1, when genetic change is primarily a function of time. If genetic distance is a function of the number of speciations (or if lineages evolving faster speciate more often), members of a speciose phylad should be much more distinct, on the average, than members of a depauperate phylad.

Whether rates of cladogenesis and anagenesis are correlated with each other or not, the accumulated amount of anagenesis is expected to be correlated with time. Needless to say, there are many factors that may contribute to anagenetic change, including population sizes, mutation rates, natural selection and others. But the longer the time involved in the evolution of a lineage, the more closely the correlation between anagenesis and time is likely to be. This is simply because the longer the time span involved, the greater the probability that the effects of other contributing factors would cancel out, or at least be reduced, relative to the effect of time.

The arguments just advanced suggest that tests of the models presented in this paper are not likely to provide evidence with respect to the current controversy

on the role of natural selection in molecular evolution. In particular, whenever large spans of time are involved, a close correlation between amount of genetic change and length of time is compatible with the neutralist as well as with the selectionist view of molecular evolution. According to the neutrality theory, rates of molecular evolution should be correlated with time, since they depend on the rates of mutation to neutral alleles (KIMURA 1968; KING and JUKES 1969; SARICH and WILSON 1967). If molecular evolution is subject to natural selection, rates of genetic change would be affected by a number of processes, including mutation rates and variable selective pressures. For the sake of argument, assume that the rate of genetic change is a function of the rate of environmental change because it results solely from the adaptation of organisms to their environments. The longer the time involved, the greater the amount of genetic change that would have accumulated. Rates of change may have fluctuated from protein to protein and from time to time. The fluctuations are likely to cancel out when averages are made over many proteins and over long times. The fluctuations would not be expected to cancel out if consistent trends with opposite effects persist in different lineages. Cladogenetic rates are such trends, since rates of speciation are known to vary greatly from phylad to phylad. The models presented here provide a way to test whether rates of clado-genesis are correlated with rates of anagenesis.

LITERATURE CITED

AVISE, J. C. and F. J. AYALA, 1975 Genetic differentiation in speciose versus depauperate phylads: evidence from the California minnows. Evolution (In press.)

AVISE, J. C. and M. H. SMITH, 1974 Biochemical genetics of sunfish. II. Genic similarity between hybridizing species. Am. Naturalist **108**: 458–472.

AYALA, F. J., 1974 Biological evolution: natural selection or random walk? Am. Scientist **62**: 692–701. ——, 1975 Genetic differentiation during the speciation process. Evol. Biol. **8**: (In press.)

AYALA, F. J. and M. E. GILPIN, 1974 Gene frequency comparisons between taxa: support for the natural selection of protein polymorphisms. Proc. Natl. Acad. Sci. U.S. **71**: 4847–4849.

KIMURA, M., 1968 Evolutionary rate at the molecular level. Nature **217**: 624–626.

KING, J. L., and T. H. JUKES, 1969 Non-Darwinian evolution. Science **164**: 788–798.

KOHNE, D. E., J. A. CHISCON and B. H. MOYER, 1972 Evolution of primate DNA sequences. J. Human Evolution 1:627–644.

LEVINS, R., 1966 The strategy of model building in population biology. Am. Scientist **54**: 421–431.

MILLER, R. R., 1965 Quaternary freshwater fishes of North America. Pp. 569–581. In: *The Quaternary of the United States*. Princeton Univ. Press, Princeton, New Jersey.

NEI, M., 1972 Genetic distance between populations. Am. Naturalist **106**: 283–292.

SARICH, V. M. and A. C. WILSON, 1967 Rates of albumin evolution in primates. Proc. Natl. Acad. Sci. U.S. **58**: 142–148.

APPENDIX

1. We want to find the solution of (5):

$$\bar{d}(1) = \frac{m(l-1) \sum\limits_{i=1}^{i=k} il^{i-1}}{l^k-1} \ .$$

(A1)

Let

$$\phi_k = \sum_{i=0}^{i=k} l^i \ ,$$

(A2)

then

$$l\phi_k = \sum_{j=1}^{j=k+1} l^j = \phi_{k+1} - 1 \ ,$$

(A3)

but

$$\phi_{k+1} = \sum_{i=0}^{i=k} l^i + l^{k+1} \ ,$$

(A4)

and, therefore,

$$l \sum_{i=0}^{i=k} l^i = \sum_{i=0}^{i=k} l^i + l^{k+1} - 1 \ ,$$

(A5)

or

$$\sum_{i=0}^{i=k} l^i = \frac{l^{k+1}-1}{l-1} \ .$$

(A6)

We now note that

$$\frac{\partial}{\partial l} \sum_{i=0}^{i=k} l^i = \frac{\partial}{\partial l} \left[\frac{l^{k+1}-1}{l-1} \right] = \frac{1}{l-1} \left[(k+1)l^k - \frac{l^{k+1}-1}{l-1} \right]$$

(A7)

but

$$\frac{\partial}{\partial l} \sum_{i=0}^{i=k} l^i = \sum_{i=1}^{i=k} il^{i-1} \ ,$$

(A8)

and, therefore,

$$\sum_{i=1}^{i=k} il^{i=1} = \frac{1}{l-1} \left[(k+1)l^k - \frac{l^{k+1}-1}{l-1} \right] \ .$$

(A9)

Replacing (A9) into (A1), we get

$$\bar{d}(1) = \frac{m(l-1) \dfrac{1}{l-1} \left[(k+1)l^k - \dfrac{l^{k+1}-1}{l-1} \right]}{l^k-1} =$$
$$= \frac{m[(k+1)(l^{k+1}-l^k) - l^{k+1}-1]}{(l^k-1)(l-1)} = \frac{m[kl^k(l-1) - l^k - 1]}{(l^k-1)(l-1)} =$$
$$= m\left(k - \frac{1}{l-1} + \frac{k}{l^k-1} \right) \ ,$$

(A10)

which is the solution given in (6).

2A. We now want to find the solution to (11):

$$\bar{d}(2\mathrm{A}) = \frac{(l-1) \sum\limits_{i=1}^{i=k} (2i-1) \, l^{i-1}}{l^{k-1}} \; . \tag{A11}$$

Let

$$\Psi_k = \sum_{i=1}^{i=k} (2i-1) \, l^{i-1}$$

then

$$\Psi_k = 2 \sum_{i=1}^{i=k} i l^{i-1} - \sum_{i=1}^{i=k} l^{i-1} = 2 \sum_{i=1}^{i=k} i l^{i-1} - \sum_{j=0}^{j=k-1} l^j \; . \tag{A12}$$

The latter two summations have been previously solved as (A9) and (A6). Substituting we obtain:

$$\Psi_k = \frac{2}{l-1} \left[(k+1) l^k - \frac{l^{k+1}-1}{l-1} \right] - \frac{l^k-1}{l-1} \; ; \tag{A13}$$

and, therefore, replacing the value of Ψ_k in (A11)

$$\bar{d}(2\mathrm{A}) = \frac{(l-1)\Psi_k}{l^{k-1}} = \frac{2 \left[(k+1) l^k - \frac{l^{k-1}-1}{l-1} \right] - l^k - 1}{l^{k-1}} =$$

$$= 2k + \frac{2k}{l^k-1} - \frac{l+1}{l-1} \; , \tag{A14}$$

which is the solution given in (12).

2B. Finally, we want to find the solution to (16):

$$\bar{d}(2\mathrm{B}) = \frac{(l-1) \sum\limits_{i=1}^{i=k} l^{2i}}{l(l^k-1)} \; . \tag{A15}$$

Let

$$\Omega_k = \sum_{i=1}^{i=k} l^{2i} = \sum_{i=1}^{i=k} \lambda^i \; , \tag{A16}$$

where $\lambda = l^2$. But the last expression is the sum of the terms of an arithmetic progression. Therefore,

$$\Omega_k = \frac{\lambda^{k+1}-1}{\lambda-1} = \frac{l^{2k+2}-1}{l^2-1} \; . \tag{A17}$$

Therefore

$$\bar{d}(2\mathrm{B}) = \frac{(l-1) l^{2k+2}-1}{l(l^k-1)(l^2-1)} = \frac{l^{2k+2}-1}{l(l^k-1)(l+1)} =$$

$$= \frac{l^{k+2}+l^2}{l(l+1)} + \frac{l^2-1}{l(l^k-1)(l+1)} = \frac{l(l^k+1)}{l+1} + \frac{l-1}{l(l^k-1)} \; , \tag{A18}$$

which is the solution given in (17).