

On the dependence of speciation rates on species abundance and characteristic population size

ANASTASSIA M MAKARIEVA* and VICTOR G GORSHKOV

Theoretical Physics Division, Petersburg Nuclear Physics Institute, 188300, Gatchina, St.-Petersburg, Russia

Corresponding author (Fax, +7-812-7131963; Email, elba@infopro.spb.su)

The question of the potential importance for speciation of large/small population sizes remains open. We compare speciation rates in twelve major taxonomic groups that differ by twenty orders of magnitude in characteristic species abundance (global population number). It is observed that the twenty orders of magnitude's difference in species abundances scales to less than two orders of magnitude's difference in speciation rates. As far as species abundance largely determines the rate of generation of intraspecific endogenous genetic variation, the result obtained suggests that the latter rate is not a limiting factor for speciation. Furthermore, the observed approximate constancy of speciation rates in different taxa cannot be accounted for by assuming a neutral or nearly neutral molecular clock in subdivided populations. Neutral fixation is only relevant in sufficiently small populations with $4N_e\mu < 1$, which appears an unrealistic condition for many taxa of the smaller organisms. Further research is clearly needed to reveal the mechanisms that could equate the evolutionary pace in taxa with dramatically different population sizes.

[Makarieva A M and Gorshkov V G 2004 On the dependence of speciation rates on species abundance and characteristic population size; *J. Biosci.* **29** 119–128]

1. Introduction

Current views on the relationship of speciation rate on population size remain to be largely shaped by modelling studies. Shifting balance models envisage speciation events as shifts between high-fitness adaptive peaks (species) separated by low-fitness valleys. This implies that only if the population size is sufficiently small, the genotypic composition of the population is able to drift randomly through the valley of low fitness with a considerable probability (Wright 1931; Mayr 1954; Barton and Charlesworth 1984). Other modelling studies based on the assumption of existence of 'ridges' of highly fit genotypes (Dobzhansky 1937; Gavrillets *et al* 1998) that connect neighbouring adaptive peaks, suggest either independence of speciation rates from total population size (Orr and Orr 1996) or higher speciation rates in large populations with large geographic ranges (Gavrillets *et al* 1998).

On the other hand, the majority of studies in conservation biology are currently concentrated on the problem of maintaining sufficient intraspecific genetic variation, which is thought to be vital for species survival and further (adaptive) evolution (Cohn 1986; O'Brien 1994). Here population size together with generation time emerge as important factors governing the rate of generation of genetic variation in a species. It can be thought that species composed of a large number of individuals with short generation time should evolve more quickly, because they will sooner generate beneficial mutations needed for fitting into a new environment. The experimental studies aimed at revealing *in vitro* evolution in abundant, rapidly reproducing organisms (e.g. Papadopoulos *et al* 1999) are apparently theoretically grounded in such an approach.

So far, no large-scale analysis of the available empirical data on speciation rates versus characteristic population sizes of species has been ever performed, despite the

Keywords. Extinction; fixation; molecular clock; neutral theory; population size; species duration

potential importance of such an analysis for the speciation theory. To a large degree, this may reflect difficulties in obtaining estimates of characteristic population sizes of different species (Nei and Graur 1984). In this paper we compare rates of speciation between organisms that differ by up to twenty orders of magnitude in species abundances (i.e. global species population numbers). It proves that the twenty orders of magnitude's difference in species abundances scales to less than two orders of magnitude's difference in speciation rates. We further show that the observed pattern cannot be accounted for by the neutral molecular clock approach.

2. Methods

2.1 Estimates of speciation rates

According to the fossil record, morphologically distinct groups of organisms (morphospecies) succeed each other to dominate ecological space. We are interested in the mean time, t_s , between two successive speciation events in a given evolutionary lineage. If the number of species remains constant or changes only slowly, this time can be estimated by the mean time of species duration, t_d (Sepkoski 1998), i.e. the time elapsed between the origin and extinction of a species.

Life has existed on Earth for about $T \sim 4 \cdot 10^9$ years (Hayes 1996). Increase of species numbers within separate taxa as well as the global increase in biodiversity is sufficiently well described by exponential curves (Bush *et al* 1977; Benton 1995). One can thus write

$$n = e^{(S-E)T}, \quad (1)$$

where $n \sim 10^7$ is the current number of species in the biosphere (May and Nee 1995); $S \equiv 1/t_s$ is the mean speciation rate; $E \equiv 1/t_d$ is the mean extinction rate. Using these values of n and T and taking the average species duration to be about $t_d \sim 4 \cdot 10^6$ years (Raup 1991a), we obtain from eq. (1):

$$(S-E)/E = \frac{t_d}{T} \ln n \sim 0.016. \quad (2)$$

This agrees with the result obtained by Raup (1991b) who noted that more than 99% of species that have ever lived are now extinct. One can see from eq. (2) that, on average, the time between two successive speciation events $t_s \equiv 1/S$ can be estimated by species duration, $t_d \equiv 1/E$, to the accuracy of less than 2%.

This conclusion is very robust. Even if one assumes that life originated during the Phanerozoic and use in eq. (2) $T_{Ph} = 6 \cdot 10^8$ years instead of the whole time of life existence T , one still has $(S-E)/E \sim 0.11$, that is, about 10% difference between t_s and t_d .

Values of S and E remain close to each other when considered for particular taxonomic groups as well, even when the calculations are made for the initial period of the group's diversification. During this time the increase of the number of species within the group is most rapid and the difference between t_s and t_d values is the largest possible. Based on the data of Stanley (1979), Sepkoski (1998) plotted the per-species rate of diversification, $S-E$, versus the extinction rate, E , for twelve major groups of taxa. For nine of these groups, the ratio $(S-E)/E$ is less than unity, implying a less than two-fold difference between S and E . The largest ratio, $(S-E)/E \sim 4$, is observed for planktic foraminifers. It means that the speciation rate in these organisms is five times that of extinction rate, and that t_s is five times less than t_d . These figures show that even in the extreme cases of most rapid diversification, the orders of magnitude of t_s and t_d still coincide. This makes species duration, t_d , a reasonable estimate of t_s in all cases.

It has been suggested that the increase in biodiversity can sometimes be better fitted by logistic rather than by exponential curves (Courtyllot and Gaudemer 1996). The proposed logistic models account for the periods of relative stasis in the number of species. Such periods are interrupted by mass extinctions, which, in their turn, are followed by a rapid regain in the species number. When the number of species remains constant, species duration, t_d , is by definition equal to the reciprocal of speciation rate, i.e. to t_s . On the other hand, the direct estimates of the $(S-E)/E$ ratio made for the periods of active growth of biodiversity, see above, show that t_d and t_s remain close to each other during these periods as well. This shows that, on an average, one species during its lifespan gives rise to no more than a few species, even when the species number grows most rapidly. This being the case, the conclusion that t_d and t_s are close to each other, is independent of the model of biodiversity growth applied (exponential, logistic or other).

Species durations in marine diatoms, planktic and benthic foraminifers, higher plants and vertebrates [marked with (L) in table 1] were estimated by Stanley (1985) and Bush *et al* (1977) using the Lyellian approach. At any given time point in the past, a Lyellian curve depicts the percentage of species within the considered fossil biota that survives to the present. Mean species duration is estimated from the absolute value of the slope of the curve at zero. The number of species analysed in such studies never falls significantly below one hundred (Stanley 1979).

Species durations in dinoflagellates and diatoms [marked with (V) in table 1] were calculated by us from the slopes of survivorship curves (plots of the proportion of the original sample that survives for various intervals) constructed by Van Valen (1973). For most taxa survivorship curves are approximately log linear (Van Valen

1973; Stanley 1985). Mean species duration can be estimated from the constant linear slope of the curve. Number of species analysed ranged from several hundreds in dinoflagellates to several thousands in diatoms.

The upper estimate of species duration in higher plants (> 20 Myr) represents the mean age of the extant North American species (Stanley 1985). This estimate is supported by a study of Mediterranean pines by Klaus (1989), who observed that many of the extant species have fossils dating to Miocene. Mean species durations in bryophytes and beetles are estimated, respectively, from the observations that nearly all fossil bryophyte species < 20 Myr old and all known Pleistocene beetle species are extant (Stanley 1985). Species duration in *Drosophila* was estimated from below by the age of the extant species of *Drosophila melanogaster* subgroup (Li *et al* 1999).

The absence of reliable fossil records hinder the analysis of speciation patterns in prokaryotes. However, the limited evidence available, although it should be interpreted with caution, suggests speciation rates not dramatically different from those of eukaryotes. Bacteriologists claim, based on the studies of morphological, ecological

and genetic characteristics of bacteria, that close but different bacterial species display less than 97% of sequence identity (Stackebrandt and Goebel 1994). To obtain an upper estimate of the bacterial speciation rate, one can assume that close bacterial species differ by about 1% of their genome. Rate of large-scale molecular evolution in bacteria is of the order of $(4-8) \cdot 10^{-9}$ substitutions per synonymous site per year (Ochman *et al* 1999). From this one obtains an estimate of $t_s > 1-2$ million years for bacteria. That is, not oftener than every one million year there appears a new type of bacteria separated from its ancestor by the same amount of genetic difference that is characteristic of the extant bacterial species.

2.2 Estimates of rates of generation of genetic variation per species

2.2a Rate of generation of intraspecific genetic variation: Appearance of new genetic variants in a population is an indispensable prerequisite for speciation and could be a limiting factor for the latter. We estimate the rate R of appearance of new genotypes in a species, con-

Table 1. Species durations in taxa with known global species abundances^{a,b}.

Taxon	t_d species duration (Myr)	N Global species abundance, individuals	g generation time (year)	G genome size (bp)	n_g mutation rate ^d , (bp) ⁻¹ (generation) ⁻¹	R rate of appearance of new genotypes per species, geno- types (species) ⁻¹ yr ⁻¹
1 Bacteria	> 1-2	10^{24}	$10^{-3}-10^{-2}$ [5,6]	$G_{Vg}: \sim 10^{-3}$ [7]		$10^{23}-10^{24}$
2 Marine diatoms	25 (L) [1] 8-15 (V)	10^{18}	$10^{-3}-10^{-2}$ [6]	10^7-10^{10} [8]	10^{-10}	$10^{17}-10^{21}$
3 Dinoflagellates	16 (V)	10^{17}	10^{-2} [6]	10^7-10^{10} [8]	10^{-10}	$10^{16}-10^{19}$
4 Planktic foraminifers	> 20 (L) [1]	10^{17}	10^{-2} [6]	10^7-10^{10} [8]	10^{-10}	$10^{16}-10^{19}$
5 Benthic foraminifers	> 20 (L) [1]	$> 10^{15}$	10^{-2} [6]	10^7-10^{11} [8]	10^{-10}	$10^{14}-10^{17}$
6 Insects (beetles, <i>Drosophila</i>)	> 2	$10^{11}-10^{13}$ [3]	10^{-1} [3]	10^8 [9]	10^{-9}	$10^{11}-10^{14}$
7 Higher plants ^c	> 8 (L) [1] > 20	$10^{11}-10^{12}$	10 [3]	10^{10} [10]	10^{-9}	$10^{10}-10^{11}$
8 Bryophytes	> 20	10^{16}	1	10^8 [11]	10^{-9}	10^{15}
9 Lizards	26 (L) [2]	10^5-10^9 [3]	1 [3]	10^9 [12]	10^{-9}	10^5-10^9
10 Turtles, crocodiles	5 (L) [2]	10^3 [3,4]	10 [3]	10^9 [12]	10^{-9}	10^4
11 Rodents	1 (L) [2]	10^5-10^8 [3]	1 [3]	10^9 [13]	10^{-9}	10^5-10^8
12 Carnivores	1-2 (L) [2]	10^4 [3]	1-10 [3]	10^9 [13]	10^{-9}	10^3-10^4

^aNumbers in square brackets indicate sources of data that are not explained in the text. [1], Stanley 1985; [2], Bush *et al* 1977; [3], Nei and Graur 1984; [4], Bourn *et al* 1999; [5], Ochman *et al* 1999; [6], Stanley 1979; [7], Drake 1991; [8], Li 1997; Cornelissen *et al* 1984; [9], Laird 1973; Juan and Petitpierre 1991; Petitpierre *et al* 1993; [10], Wakamiya *et al* 1993; [11], Voglmayr 2000; [12], Mirsky and Ris 1951; Olmo 1976; Capriglione *et al* 1987; [13], Orlov and Bulatova 1983.

^bThe following taxa with unknown population sizes display similar species durations (in Myr): marine bivalves (11-14) and gastropods (10-14) (Stanley 1985); snakes and frogs (16), and salamanders (5) (Bush *et al* 1977).

^cGenome size estimated for *Pinus* genus to match the estimate of population size, which is also for *Pinus*.

^dMutation rate n_g can be written as $n_g = nk$, where n is mutation rate per base pair per cell division, k is the average number of divisions in the germ line. For unicellular organisms $k = 1$ and $n = n_g$. For all the eukaryotes we took $n \sim 10^{-10}$ (bp)⁻¹ (div)⁻¹ based on the data of Drake (1991) for *Saccharomyces cerevisiae* and *Neurospora crassa*. For most multicellular organisms $k \geq 10$ (Orr 1995), which gives $n_g \sim 10^{-9}$ (bp)⁻¹ (generation)⁻¹.

sisting of N individuals with genome size G and generation time g , as

$$R = \mathbf{n}_g GN/g, \quad \text{if } \mathbf{n}_g G < 1; \text{ and} \quad (3a)$$

$$R = N/g, \quad \text{if } \mathbf{n}_g G > 1, \quad (3b)$$

where \mathbf{n}_g ($\text{bp}^{-1} \text{ generation}^{-1}$) is the mutation rate. If there is less than one mutation per genome per generation, the proportion of new genotypes within each generation is equal to $\mathbf{n}_g G$, see eq. (3a). Otherwise, on average every genotype will be a new one (assuming an infinite allele model), so that the dependence on $\mathbf{n}_g G$ vanishes, see eq. (3b).

The above definition of R is not the only possible one. One could, for example, consider mutations in the coding region only. However, since a large part of the genome of higher organisms is noncoding, this would mean that for most organisms the product $\mathbf{n}_g G$ (where G is understood as the size of the coding region) would be less than unity. In such a case in all organisms R will be determined by eq. (3a). Alternatively, one can apply a different logic. Due to recombination and random assortment of the parental genetic material, any newly born individual in a sexually reproducing species represents a new genotype absent in the previous generation, even if no new mutations occurred in its genome. Then, for the vast majority of species, R should be estimated by eq. (3b), which is merely the number of individuals born per unit time. However, the exact way of defining R has no impact on our major result, that is, on the observation of a many orders of magnitude's contrast between the scatter of R values and the relevant scatter of speciation rates (see §3). In fact, the definition that we chose, see eq. (3a) and eq. (3b), appears to be most conservative with respect to this result. It diminishes the value of R in organisms with largest N (bacteria) due to the condition $\mathbf{n}_g G < 1$, thus reducing the overall scatter of R values.

2.2b Estimates of species abundance N: Within ecological communities heterotrophs of the smallest size (bacteria) claim the largest energy fluxes of the order of total net primary production (Whittaker and Likens 1975), which is of the order of $P = 2 \text{ W m}^{-2}$. Metabolic power of one bacterial cell per unit area of its projection on the ground surface can be calculated as $j = p/l^2$, where $l \sim 10^{-6} \text{ m}$ is the characteristic linear size of bacteria (Curtis and Barnes 1989), p is the metabolic power per individual. Given growth efficiency in bacteria of about 0.5, the value of p can be estimated as $p = mK/g$, where $m = \mathbf{r}l^3 \sim 10^{-18} \text{ t}$ is body mass, $\rho \sim 1 \text{ t m}^{-3}$ is the density of living matter, $K \sim 4.2 \cdot 10^9 \text{ J t}^{-1}$ is the mean energy content of living matter, g is the average generation time.

Taking $g \sim 0.003 \text{ year} \sim 10^5 \text{ s}$ (Ochman *et al* 1999; Stanley 1979, p. 210) we obtain $j = mK/gl^2 = lK/g \sim 0.04 \text{ W m}^{-2}$ and $d = P/j = 50$. That is, there are about 50 bacterial cells along a vertical axis of the ecosystem (d is analogous to

leaf area index in plants). Assuming that bacteria form a continuous cover both on land and in the sea (Sieburth 1976) of thickness $L = dl = 5 \cdot 10^{-5} \text{ m}$, the total live mass of all bacteria (spread across the Earth's surface of $S = 5 \cdot 10^{14} \text{ m}^2$) is estimated as $M = \mathbf{r}LS = 3 \cdot 10^9 \text{ t}$. The total number of bacterial cells on Earth is estimated as $M/m \sim 3 \cdot 10^{27}$. Very similar estimates of bacterial abundances were reported by Schmidt *et al* (1998) and Whitman *et al* (1998). As there are about $3 \cdot 10^3$ bacterial species (Curtis and Barnes 1989), the number of bacteria of one and the same species estimates as $N_b \sim 10^{24}$ (Gorshkov 1995).

Total biomass of aquatic organisms tends to be uniformly distributed over logarithmically equally spaced size classes of organisms (Sheldon *et al* 1972; Witek and Krajewska-Soltys 1989; Gaedke 1992; Cohen *et al* 2003). The mean cell size of autotrophic unicellular eukaryotes is about $50 \cdot 10^{-6} \text{ m}$ (Curtis and Barnes 1989). Using the order of magnitude of the obtained value of total bacterial biomass, $M \sim 10^9 \text{ t}$, and estimating the mass of eukaryotic cell to be $m_E \sim 10^{-13} \text{ t}$, we obtain that the global number of marine unicellular autotrophic eukaryotes is of the order of $M/m_E \sim 10^{22}$. There are about 10^4 known species of marine unicellular autotrophic eukaryotes (Curtis and Barnes 1989). The average species abundance of a marine diatom is thus about $N_{\text{dia}} \sim 10^{18}$ individuals.

The majority of foraminifers and dinoflagellates are larger than marine diatoms, $l > 100 \cdot 10^{-6} \text{ m}$ (Curtis and Barnes 1989). On average, they should therefore have species abundances at least several times smaller than those of marine diatoms, $N_{\text{f,din}} \leq 5 \cdot 10^{17}$.

The few available direct estimates of species abundance in marine organisms confirm the applied approach. Global abundance of unicellular *Emiliania huxleyi* with cell size $l \sim 4 \cdot 10^{-6} \text{ m}$ is of the order of $7 \cdot 10^{22}$ (Emiliani 1993). Being about an order of magnitude smaller in linear size than the average diatom and about four times larger than the average bacteria, *E. huxleyi* should have population size less than two orders of magnitude smaller than bacteria and about three orders of magnitude larger than diatoms. This agrees satisfactorily with the above estimates of N_b and N_{dia} .

In the Black Sea a dinoflagellate *Noctiluca scintillans* has a population density of about 10^4 ind m^{-3} (Konsuloff 1976). In the Black Sea (total area of the order of $4 \cdot 10^9 \text{ m}^2$) living organisms are concentrated within the 100 uppermost meters of the water column. Using these data we obtain that the species abundance of *N. scintillans* is about $N \sim 4 \cdot 10^{17}$, which agrees with our theoretical estimates of $N_{\text{f,din}}$.

In coral reef ecosystems normal population density of the widespread large benthic foraminifers of *Amphistegina* spp. ($l \sim 10^{-4} \text{ m}$) is more than 50 individuals per 100 cm^2 bottom of rubble (Hallock 1984). Given that

coral reefs occupy about less than 1% of total surface area on a planet and assuming that *Amphistegina* spp. are abundant in the majority of coral reef ecosystems (Hallock 1984), we estimate the global population number for an *Amphistegina* sp. to be of the order of 10^{15} individuals. This value can be regarded as the lower estimate of average population sizes in benthic foraminifers due to the relatively large body size of *Amphistegina* organisms.

Species abundances in plants, $N_{pl} \sim 10^{11} - 10^{12}$, were estimated using pine *Pinus sylvestris*, which has a geographic habitat range $H \sim 10^6$ km² (Sokolov and Svyazeva 1965) and population density of about $D \sim 10^5 - 10^6$ ind km⁻² (Sannikov 1992). Species abundance in bryophytes, $N_{br} \sim 10^{16}$, was estimated from the data on population density of approximately $\sim 10^4$ ind m⁻² for *Sphagnum fuscum*, a species common to the bogs of the temperate zone (Grabovik 1998), assuming that bogs constitute about 20% of the forest-covered area, $7.6 \cdot 10^6$ km², of the temperate zone (Lanly and Allan 1991). Generation time in bryophytes was taken to be of the order of one year.

Using the available estimates of G , g , n_g and N ; values of R were estimated for 12 taxa of organisms (table 1).

3. Results: constancy of speciation rates

Data of table 1 plotted in figure 1 make it clear that in the first approximation speciation rates are independent of the rate of appearance of new genotypes in a species, R , which is largely determined by species abundance N . A more than 20 orders of magnitude's range of R values scales to less than two orders of magnitude range of species durations. [Linear regression of $\log t_d$ on $\log R$ (middle points of uncertainty intervals, table 1) gives the following results: $r = 0.30$, $p > 0.3$ with 10 df.) Even if one excludes bacteria, the range of R values remains practically unaffected – 18 orders of magnitude.

The character of this major result justifies the procedure of data selection that we applied when constructing table 1 and figure 1. Our primary goal was to find taxa which would show the largest possible scatter in R values and to test the corresponding scatter in speciation rates. By choosing the most abundant organisms (bacteria) and the least abundant (carnivorous mammals) we outlined the maximum range of R values. Now the task was to enhance the statistics for t_d values and to show that their scatter is narrow as compared to the scatter in R . Within such an approach, the exact value of R for a given taxon is not very significant, as one knows that it is in any case confined somewhere between $R_{bacteria}$ and $R_{mammals}$. In addition to the 12 taxa shown in table 1, there are five more taxa with unavailable R estimates, but with t_d values falling within the same narrow range (these are bivalves, gastropods, snakes, frogs, and salamanders shown under comment *b* in table 1).

The observed differences in R values between different taxa are so huge that the obtained results are hardly sensitive to particular estimates of any involved parameters. For example, were the inverse proportionality of t_d to R to hold, one would observe marine diatoms evolving at a rate of about 10^8 acts of speciation per second per evolutionary lineage (if one takes species duration in carnivores as a reliable reference point). As this rate by far exceeds the reproduction rate of these organisms, such a result would mean that marine diatoms (and other abundant organisms) would be in a state of continuous morphological changes. This would render recognition of morphological uniformity in any species impossible, which is evidently not the case.

Before proceeding to the discussion, it is necessary to comment on the validity of the species abundance N for the present analysis. As already noted, an indispensable prerequisite for evolution is the appearance of new genetic variants. In the limit, if the genetic composition of populations were not changing with time, no evolution would be possible. For a species with genome size G one can envisage 4^G possible genotypes. The overwhelming majority of those genotypes are biologically meaningless. A few genotypes or a few groups of related genotypes correspond to different species. Thus, evolution can be viewed

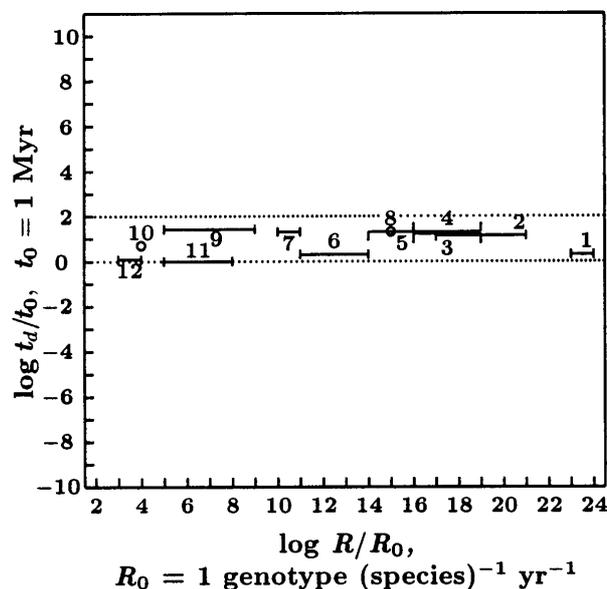


Figure 1. Species durations t_d versus rates R of generation of genetic variation in different taxa. Numbers correspond to taxonomic groups in table 1. Dotted line shows the observed scatter of speciation durations (approx. two orders of magnitude). The figure segment where the speciation rates are shown, is chosen to be narrow to help visualize the difference in the scatter of speciation durations vs species abundances.

as a continuous trial of newly appearing genotypes for their ability to give rise to a new species. A species with a large global abundance N produces more genetic variants to be tested by natural selection than a species with a small N , independent of the degree of subdivision of the global population. Namely the global species abundance determines the rate of generation of genetic variability within a species.

Species abundance is a product of characteristic population density D and the area of species range H , $N = DH$. On a larger scale, species range H does not considerably depend on body size (although such a dependence may exist within certain taxonomic groups, see, e.g. Gaston 1996). This becomes clear from the fact that species of both unicells (Finlay and Clarke 1999) as well as of largest animals (Gaston 1996) may range over territories of continental and global scale. Hence, a major determinant of N is population density D , which is heavily dependent on body size of the organism – there are many more bacteria per square meter than elephants.

Isolation of local subpopulations (e.g. geographic isolation on a island) is one of the widely discussed mechanisms of speciation (Mayr 1963). However, population isolation appears to be largely irrelevant to the present analysis. Due to the fact that N is proportional to D , while D is determined by body size and approximately independent of H , the ratio between local population numbers of differently-sized species on an isolated area is roughly proportional to the ratio between their global abundances N . [The same logic was applied in the few existing attempts to quantify the impact of population size on the rate of molecular evolution, when global species abundance was used as a proportional estimate of the effective population size (see, e.g. Nei and Graur 1984).] Consideration of isolated instead of panmictic populations would therefore yield the same magnitude of relative changes as those presented in figure 1 and table 1, providing no clues for their possible causes. In a somewhat different context, we return to the problem of population subdivision in the next section.

4. Discussion

4.1 The neutral theory approach to speciation

The question arises what are the factors that ensure similar evolutionary pace (in terms of speciation events) in the taxa studied? One possibility to be explored is the neutral theory approach, i.e. assuming that speciation occurs in the course of random genetic drift along high-fitness ridges due to fixation of (nearly) neutral mutations (Orr and Orr 1996). Within such an approach, speciation (\equiv reproductive isolation) occurs as soon as the genetic distance between diverging populations becomes suffi-

ciently high. If close species differ on average by the same $d \sim 1\%$ in all taxa, and mutations are fixed at a roughly constant rate $r \sim 0.1\text{--}1\%$ per million year, speciation will proceed at an approximately constant rate in all organisms, namely each $d/r \sim 1\text{--}10$ million years, which agrees with the data of table 1.

If the mutation rate per generation is constant, then, according to the classical neutral theory, fixation of neutral mutations occurs at a constant rate per generation in all organisms, independent of their population size. In such a case the four orders of magnitude's difference in the generation time of the considered taxa (table 1) will make the shortest-lived organisms speciate four orders of magnitude more rapidly than the longest-lived ones. Such a trend is apparently not observed. Fixation rate of slightly deleterious (nearly neutral) mutations may be constant over real time rather than over time calibrated in generations, if there is an inverse correlation between generation time and effective population size (Ohta 1987; Chao and Carr 1993). This might yield a time-constant speciation rate.

However, there are serious limits to the applicability of the (nearly) neutral theory to the explanation of molecular evolution (Gorshkov 1995). According to Kimura and Ohta (1969), the average time to fixation of a neutral allele in a population equals $4N_e$ generations, where N_e is the effective population size. The probability distribution of fixation time is very significantly skewed to the right, so that fixation in less than $0.2 N_e$ generations is practically improbable (see figure 1 in Kimura 1970).

Let us for simplicity consider a population of N_e haploid unicellular asexual organisms. (In cases of diploidy and sexual reproduction one can envisage particular genotypes instead of haploid individuals and obtain essentially the same results.) The number of mutations in a particular site per generation is equal to nN_e , where $n \geq 10^{-10}$ is mutation rate per base pair per cell division. Consider a single point mutation, e.g. $G > T$, which occurred in site a of an allele that was destined to fix. In the absence of new mutations this allele will be fixed in about $4N_e$ generations. That is, in $4N_e$ generations all N_e individuals in the population will have T instead of G in site a . However, during the same $4N_e$ generations there will on average appear about $vN_e \cdot 4N_e$ individuals that will contain some other base pair (A, C or G) in that very site a , because new mutations appear in the population at a rate of nN_e . For fixation to occur, one has to demand that the number of such individuals should be considerably less than the total number of individuals, $vN_e \cdot 4N_e < N_e$, which gives

$$4N_e n < 1. \quad (4)$$

Only in those populations where this condition is satisfied, one can speak about neutral fixation and, consequently, evolution by neutral molecular clock. With a

conservative estimate of $(\mathbf{n} \sim 10^{-10} \text{ (bp)}^{-1} \text{ (generation)}^{-1})$, condition (4) implies that N_e may not be considerably larger than 10^9 , $N_e \leq 10^9$. The average time to fixation of deleterious mutations is longer than that of neutral mutations, in very much the same manner as the average time to fixation of advantageous mutations is shorter than that (Kimura and Ohta 1969). It means that for a slightly deleterious mutation to reach fixation, the effective population size should be even smaller than demanded by (4). When $4N_e\mathbf{n} > 1$, there will be an approximately equal frequency of all the four base pairs in all neutral sites in the population. In other words, in very large populations no neutral or nearly neutral fixation is possible at all, due to appearance of new mutations in the course of fixation (Kimura and Ohta 1973; Gorshkov 1995).

This result was first obtained by Kimura and Ohta (1973) but was never broadly discussed within the biological community. One possible reason for such a situation is that when Kimura and Ohta formulated restriction (4) for the applicability of the neutral theory, it was implicitly assumed that for most organisms studied genetically at that time (mammals, insects and other higher organisms) the effective population sizes never go beyond 10^9 , thus satisfying eq. (4).

There is another interpretation of condition (4). One can envisage $4N_e\mathbf{n}$ as the average number of mutations that occurred in a nucleotide site during the time of fixation of an allele to which it belongs. The probability that a genotype randomly picked out from the population remains unaffected by mutation in this site is given by $e^{-4N_e\mathbf{n}}$. The probability that there are N_e genotypes unaffected by mutation in this particular site in the population (which corresponds to fixation) is $p_{\text{fix}} = [e^{-4N_e\mathbf{n}}]^{N_e}$. At $\mathbf{n} \sim 10^{-10}$ and $4N_e\mathbf{n} \sim 1$ (which means $N_e \sim 10^9$), p_{fix} becomes virtually zero. Hence, in large populations with $N_e > 10^9$ no neutral fixation can take place.

Note that under fixation we understand a situation when the nucleotide in a particular site is consistently more abundant than the others, that is, its frequency is of the order of unity. The condition for strict fixation (when the nucleotide frequency exactly equals unity) is much stronger than given in eq. (4). In such a case one has to demand that no new mutations occur during the $4N_e$ generations to fixation, which gives $4N_e^2\mathbf{n} < 1$.

It is now of interest to check how condition (4) is fulfilled in organisms described in table 1. Due to difficulties in determining the real effective population sizes, the few studies concerned about checking the numerical predictions of the neutral theory with respect to effective population sizes derive the information about this parameter from the data on species abundance, see Nei and Graur (1984), by assuming a positive relationship between these two characteristics. If one follows this approach and takes the species abundances N as a rough estimate of

effective population size, N_e , it becomes clear from table 1 that speciation rates in either bacteria, or higher plants, or insects, or foraminifers, or dinoflagellates or diatoms, cannot be explained by the (nearly) neutral fixation, because for them condition considered in eq. (4) with $N_e = N$ does not hold.

4.2 Population subdivision

In reality, the effective population sizes are in most cases much smaller than the global species abundances, $N_e \ll N$. The question thus becomes whether the real values of N_e are sufficiently small to satisfy eq. (4) for the highly abundant organisms namely, bacteria and other unicells. This possibility is immediately rejected for the small organisms living within the well-mixed layer of the ocean. The average population density in these organisms is more than 10^6 individuals per 1 m^2 of the water column. For N_e to be less than 10^9 , one has to demand that the oceanic populations of bacteria and protists are divided into isolated subpopulations inhabiting areas of about $30 \times 30 \text{ m}^2$ and less. Taking into account the tremendous wind and wave mixing of the ocean, any prolonged existence of such isolated subpopulations is highly unlikely. Studies of small subunit ribosomal DNA in planktonic foraminifers confirm this statement and suggest a substantial gene flow even between bipolar populations of these organisms (Darling *et al* 2000), to say nothing about local adjacent areas of the size of 100 m^2 .

Furthermore, even if the sufficiently small ($N_e < 10^9$) isolated subpopulations as exemplified by foraminifers did exist somehow and evolved each independently governed by nearly neutral molecular clock, then in several million years (average species duration) a single foraminiferous species subdivided into $N/N_e > 10^{20}/10^9 > 10^{11}$ subpopulations would produce more than 10^{11} new species compared to the total of $\sim 10^4$ species known. (The situation with the other smaller organisms would be approximately the same. For example, the number of insect species would increase by a factor of 10^2 – 10^4 .) Although some studies of molecular phylogeny of planktic foraminifers did reveal cryptic species within morphologically uniform taxa, the number of cryptic species per morphospecies is usually small (de Vargas *et al* 1999). Besides, studies of other protists show that many species display a ubiquitous spread. Finlay and Clarke (1999) identified 78% of the world-wide recorded species of *Paraphysomonas* (this genus belongs to a sister class with respect to diatoms within the division *Chrysophyta*) in 0.1 cm^2 of sediment collected from a freshwater pond. This led the researchers to conclude that the global number of species from this genus cannot be vastly different from what is already known. In any case, the above esti-

mate of the number of species of small organisms, $n \sim 10^{11}$, appears as completely unrealistic.

On the other hand, if one assumes that the nearly neutral molecular clock explanation is invalid for the smaller (abundant) organisms, but remains valid for the others, one will face the problem of accounting for the observed coincidence of two independent factors. Indeed, if the constancy of speciation rates in the larger (less abundant) organisms is dictated by (nearly) neutral molecular clock, while in the abundant organisms it is dictated by a different factor, one has then to explain why there is no difference in speciation rates between the abundant and the less abundant organisms, as witnessed by table 1.

To retain the (nearly) neutral molecular clock, one needs to assume that from time to time individuals from a single subpopulation acquire the ability to expand nearly instantaneously (on an evolutionary timescale) into the whole ecological space occupied by the species. One of the ways of how this may happen is through a bottleneck, i.e. reduction of the species to a single local subpopulation with N_e satisfying eq. (4) and further expansion of this population to the whole ecological space of a species. (In the abundant organisms; e.g. foraminifers, this would correspond to reduction of the global species population number by a factor of 10^{-11} or less.) However, it remains unclear why bottlenecks should happen at the same rate in different evolutionary lineages. Mass extinctions, which are most probably caused by environmental factors and during which species become extinct in a correlated fashion are short-lived events (Raup and Sepkoski 1982) and on average account for no more than several percent of the total number of extinctions during the Phanerozoic.

Furthermore, it remains unclear why bottlenecks should happen at a rate approximately coinciding with the rate of molecular evolution. Indeed, for the rate of molecular evolution to remain constant, bottlenecks should occur exactly at those moments when the genetic distance between isolated subpopulations reaches the level characteristic of two species. If bottlenecks happen more often, molecular evolution will speed up and its rate will be determined by the rate of bottleneck appearance rather than by the rate of neutral genetic drift. On the contrary, if bottlenecks happen only rarely than during the time period between two successive bottlenecks, the number of species in a given evolutionary lineage will fluctuate from one to N/N_e , which is a pattern hardly corresponding to the available paleoevidence.

Another factor that could account for the instantaneous spread of one subpopulation to the whole ecological range of the species, is the appearance of some non-neutral genetic change that would impart to its carriers an advantage over other conspecifics. This would mean uncoupling between molecular evolution and the process of speci-

ation, the latter to be accompanied by rapid non-neutral morphological changes (Nei 1984). If so, it is the rate of appearance of such advantageous genetic changes, that limits the rate of speciation. However, it remains unclear why such advantageous genetic changes leading to speciation should occur tens of orders of magnitude less frequent in diatoms and foraminifers than, for example, in mammals, to compensate for the population size differences and yield an approximately equal speciation rates in all taxa.

5. Conclusions

In this paper the following problem was addressed: In a given evolutionary lineage, how often do there appear morphologically distinct forms capable of claiming noticeable ecological space (compared to the other species in the considered taxon)? Rate of thus defined morphological evolution in different taxa was calibrated in terms of species durations, t_d , and plotted against the rate of appearance of new genetic variation within a given species, R . The value of R appeared to be largely determined by global population numbers of the considered species. It was shown that the more than twenty orders of magnitude's range of R values scaled to less than two orders of magnitude's range of species durations. Virtually the same statement was put forward by Gorshkov (1995), who compared the rates of appearance of interspecific genetic differences between close bacterial, insect and mammalian species, i.e. species differing from each other by orders of magnitude in their population numbers (see table 1).

It is well-known that there is a significant variability in the observed rates of molecular clock and numerous models involving weak selection, speciation and other factors have been proposed to explain this variability (Gillespie 1991). It might be interesting to explore whether the approximately one order of magnitude's difference in speciation rates between different taxa (table 1) is reflected in different rates of molecular evolution. That is, whether rapidly speciating taxa diverge more rapidly at the molecular level than more slowly speciating taxa. Such a research would shed further light on the interrelated dynamics of the processes of molecular divergence and speciation.

On the basis of the results obtained so far, it is possible to draw the following two conclusions. First, the rate of generation of endogenous genetic variation in a species does not appear to be the limiting factor for the speciation process. Second, the approximately constant speciation rate observed in different taxa is unlikely to be explained by the neutral molecular clock, i.e. evolutionary fixation of (nearly) neutral substitutions. This result adds to the existing inconsistencies between the molecu-

lar clock approach and the paleontological record (Conway Morris 1998).

In the light of these findings, we conclude that further research is needed to reveal those mechanisms that could equate speciation rates in different taxa, including exploring the possibility of a major role played in macroevolution by the acquisition of exogenic genetic material through horizontal gene transfer (Gorshkov 1995; Marienfeld *et al* 1997; Garcia-Vallvé *et al* 2000; Gorshkov *et al* 2000).

Acknowledgements

Work of AMM was supported by the European Science Foundation Scientific Programme on Theoretical Biology of Adaptation and the Russian Foundation for Basic Research, grant No. 1124-2003-2 for scientific schools. We thank three anonymous referees for their critical comments.

References

- Barton N and Charlesworth B 1984 Genetic revolutions, founder effects and speciation; *Annu. Rev. Ecol. Syst.* **15** 133–164
- Benton M J 1995 Diversification and extinction in the history of life; *Science* **268** 52–58
- Bourn D, Gibson C, Augeri D, Wilson C J, Church J and Hay S I 1999 The rise and fall of the Aldabran giant tortoise population; *Proc. R. Soc. London B.* **266** 1091–1100
- Bush G L, Case S M, Wilson A C and Patton J L 1977 Rapid speciation and chromosomal evolution in mammals; *Proc. Natl. Acad. Sci. USA* **74** 3942–3946
- Capriglione T, Olmo E, Odierna G, Improra B and Morescalchi A 1987 Cytofluorometric DNA base determination in vertebrate species with different genome sizes; *Bas. Appl. Histochem.* **31** 119–126
- Chao L and Carr D E 1993 The molecular clock and the relationship between population size and generation time; *Evolution* **47** 688–690
- Cohen J E, Jonsson T and Carpenter S R 2003 Ecological community description using the food web, species abundance, and body size; *Proc. Natl. Acad. Sci. USA* **100** 1781–1786
- Cohn J P 1986 Surprising cheetah genetics; *Bioscience* **36** 358–361
- Conway Morris S 1998 The evolution of diversity in ancient ecosystems: A review; *Philos. Trans. R. Soc. London B* **353** 327–345
- Cornelissen A W C A, Overdulve J P and van der Ploeg M 1984 Determination of nuclear DNA of five Eucoccidian parasites, *Isospora (Toxoplasma) gondii*, *Sarcocystis cruzi*, *Eimeria tenella*, *E. acervulina* and *Plasmodium berghei*, with special reference to gamontogenesis and meiosis in *I. (T.) gondii*; *Parasitology* **88** 531–553
- Courtillot V and Gaudemer Y 1996 Effects of mass extinctions on biodiversity; *Nature (London)* **381** 146–148
- Curtis H and Barnes N S 1989 *Biology* 5th edition (New York: Worth Publishers)
- Darling K F, Wade C M, Stewart I A, Kroon D, Dingle R and Brown A J 2000 Molecular evidence for genetic mixing of Arctic and Antarctic subpolar populations of planktonic foraminifers; *Nature (London)* **405** 43–47
- de Vargis C, Norris R, Zaninetti L, Gibb S W and Pawlowski J 1999 Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces; *Proc. Natl. Acad. Sci. USA* **96** 2864–2868
- Dobzhansky Th G 1937 *Genetics and the origin of species* (New York: Columbia University Press)
- Drake J W 1991 A constant rate of spontaneous mutation in DNA-based microbes; *Proc. Natl. Acad. Sci. USA* **88** 7160–7164
- Emiliani C 1993 Viral extinctions in deep-sea species; *Nature (London)* **366** 217–218
- Finlay B J and Clarke K J 1999 Apparent global ubiquity of species in the protist genus *Paraphysomonas*; *Protist* **150** 419–430
- Gaedke U 1992 The size distribution of plankton biomass in a large lake and its seasonal variability; *Limnol. Oceanogr.* **37** 1202–1220
- Garcia-Vallvé S, Romeu A and Palau J 2000 Horizontal gene transfer of glycosyl hydrolases of the rumen fungi; *Mol. Biol. Evol.* **17** 352–361
- Gaston K J 1996 Species-range-size distributions: Patterns, mechanisms and implications; *Trends Ecol. Evol.* **11** 197–201
- Gavrilets S, Li H and Vose M D 1998 Rapid parapatric speciation on holey adaptive landscapes; *Proc. R. Soc. London B* **265** 1483–1489
- Gillespie J H 1991 *The causes of molecular evolution* (Oxford: Oxford University Press)
- Gorshkov V G 1995 *Physical and biological bases of life stability. Man, Biota, Environment* (Berlin: Springer)
- Gorshkov V G, Gorshkov V V and Makarieva A M 2000 *Biotic regulation of the environment: Key issue of global change* (London: Springer)
- Grabovik S I 1998 Ecological peculiarities of reproduction in *Sphagnum* mosses; *Bot. J.* **88** 92–97 (in Russian)
- Hallock P 1984 Distribution of selected species of living algal symbiont-bearing foraminifera on two Pacific coral reefs; *J. Foram. Res.* **14** 250–261
- Hayes J M 1996 The earliest memories of life on Earth; *Nature (London)* **384** 21–22
- Juan C and Petitpierre E 1991 Evolution of genome size in darkling beetles (Tenebrionidae, Coleoptera); *Genome* **34** 169–173
- Kimura M 1970 The length of time required for a selectively neutral mutant to reach fixation through random frequency drift in a finite population; *Genet. Res.* **15** 131–133
- Kimura M and Ohta T 1969 The average number of generations until fixation of a mutant gene in a finite population; *Genetics* **61** 763–771
- Kimura M and Ohta T 1973 The age of a neutral mutant persisting in a finite population; *Genetics* **75** 199–212
- Klaus W 1989 Mediterranean pines and their history; *Plant Syst. Evol.* **162** 133–163
- Konsuloff A 1976 *Zooplankton in the Black Sea along the Bulgarian Coast*, Ph.D. thesis, Institute of Fisheries, Varna
- Laird C D 1973 DNA of *Drosophila* chromosomes; *Annu. Rev. Genet.* **7** 177–204
- Lanly J P and Allan T 1991 Overview of status and trends of world's forests; in *Proceedings of the Technical Workshop to Explore Options for Global Forestry Management* (eds) D Howlett and C Sargent (London: International Institute for Environment and Development) pp 17–39
- Li W-H 1997 *Molecular evolution* (Sunderland: Sinauer Associates)
- Li Y J, Satta Y and Takahata N 1999 Paleo-demography of the *Drosophila melanogaster* subgroup: application of the maximum likelihood method; *Genes Genet. Syst.* **74** 117–127

- Marienfeld J R, Unseld M, Brandt P and Brennicke A 1997 Viral nucleic acid transfer between fungi and plants; *Trends Genet.* **13** 260–261
- May R M and Nee S 1995 The species alias problem; *Nature (London)* **378** 447–448
- Mayr E 1954 Change of genetic environment and evolution; in *Evolution as a process* (eds) J Huxley, A C Hardy and E B Ford (New York: Macmillan) pp 157–180
- Mayr E 1963 *Animal species and evolution* (Cambridge: Harvard University Press)
- Mirsky A E and Ris H 1951 The desoxyribonucleic acid content of animal cells and its evolutionary significance; *J. Gen. Physiol.* **34** 451–462
- Nei M 1984 Genetic polymorphism and neomutationism; *Lect. Notes Biomath.* **53** 214–241
- Nei M and Graur D 1984 Extent of protein polymorphism and the neutral mutation theory; *Evol. Biol.* **17** 73–118
- O'Brien S J 1994 A role for molecular genetics in biological conservation; *Proc. Natl. Acad. Sci. USA* **91** 5748–5755
- Ochman H, Elwyn S and Moran N A 1999 Calibrating bacterial evolution; *Proc. Natl. Acad. Sci. USA* **96** 12638–12643
- Ohta T 1987 Very slightly deleterious mutations and the molecular clock; *J. Mol. Evol.* **26** 1–6
- Olmo E 1976 Genome size in some reptiles; *J. Exp. Zool.* **195** 305–310
- Orlov V N and Bulatova N Sh 1983 *Comparative cytogenetics and carosystematics of mammals* (Moscow: Nauka) (in Russian)
- Orr H A 1995 Somatic Mutation Favors the Evolution of Diploidy; *Genetics* **139** 1441–1447
- Orr H A and Orr L H 1996 Waiting for speciation: the effect of population subdivision on the waiting time to speciation; *Evolution* **50** 1742–1749
- Papadopoulos D, Schneider D, Meier-Eiss J, Arber W, Lenski R E and Blot M 1999 Genomic evolution during a 10,000-generation experiment with bacteria; *Proc. Natl. Acad. Sci. USA* **96** 3807–3812
- Petitpierre E, Segarra C and Juan C 1993 Genome size and chromosomal evolution in leaf beetles (Coleoptera, Chrysomelidae); *Hereditas* **119** 1–6
- Raup D M 1991a A kill curve for Phanerozoic marine species; *Paleobiology* **17** 37–48
- Raup D M 1991b *Extinction: bad genes or bad luck?* (New York: Norton)
- Raup D M and Sepkoski J J Jr 1982 Mass extinctions in the marine fossil records; *Science* **215** 1501–1503
- Sannikov S N 1992 *Ecology and geography of natural regeneration of Skots pine* (Moscow: Nauka) (in Russian)
- Schmidt J L, Deming J W, Jumars P A and Keil R G 1998 Constancy of bacterial abundance in surficial marine sediments; *Limnol. Oceanogr.* **43** 976–982
- Sepkoski J J Jr 1998 Rates of speciation in the fossil record; *Philos. Trans. R. Soc. London B* **353** 315–326
- Sheldon R W, Prakash A and Sutcliffe W H Jr 1972 The size distribution of particles in the ocean; *Limnol. Oceanogr.* **17** 327–340
- Sieburth J M 1976 Bacterial substrates and productivity in marine ecosystems; *Annu. Rev. Ecol. Syst.* **7** 259–285
- Sokolov S Ja and Svyazeva O A 1965 *Geography of trees in the USSR* (Moscow: Nauka) (in Russian)
- Stackenbrandt E and Goebel B 1994 Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology; *Int. J. Syst. Bacteriol.* **44** 846–849
- Stanley S M 1979 *Macroevolution: pattern and process* (San Francisco: Freeman)
- Stanley S M 1985 Rates of evolution; *Paleobiology* **11** 13–26
- Van Valen L 1973 A new evolutionary law; *Evolutionary Theory* **1** 1–30
- Voglmayr H 2000 Nuclear DNA amounts in Mosses (Musci); *Ann. Bot.* **85** 531–546
- Wakamiya I, Newton R J, Jonston J S and Price H J 1993 Genome size and environmental factors in the genus *Pinus*; *Am. J. Bot.* **80** 1235–1241
- Whitman W B, Coleman D C and Wiebe W J 1998 Prokaryotes: The unseen majority; *Proc. Natl. Acad. Sci. USA* **95** 6578–6583
- Whittaker R H and Likens G E 1975 The biosphere and man; in *Primary productivity of the biosphere* (eds) H Lieth and R Whittaker (Berlin: Springer) pp 305–328
- Witek Z and Krajewska-Soltys A 1989 Some examples of the epipelagic plankton size structure in high latitude oceans; *J. Plankton Res.* **11** 1143–1155
- Wright S 1931 Evolution in Mendelian populations; *Genetics* **16** 97–159

MS received 31 July 2003; accepted 23 December 2003

Corresponding editor: VIDYANAND NANJUNDIAH