Phylogenetic inheritance of genetic variability produced by neutral models of evolution

J. José[1], W.J. Puma-Villanueva[2], F.J. Von Zuben[2] and J.A.F. Diniz-Filho[3]

[1]Instituto de Biologia, Universidade Estadual de Campinas,
Campinas, SP, Brasil
[2]Departamento de Engenharia da Computação e Automação Industrial,
Laboratório de Bioinformática e Computação Bio-inspirada,
Faculdade de Elétrica e Engenharia da Computação,
Universidade Estadual de Campinas, Campinas, SP, Brasil
[3]Departamento de Biologia Geral, Laboratório de Ecologia Evolutiva,
Universidade Federal de Goiás, Goiânia, GO, Brasil

Corresponding author: J. José
E-mail: juliana.jose@gmail.com

ABSTRACT. The amount of genetic variability in species and populations has been mainly related to microevolutionary forces operating in natural populations and the influence of phylogenetic processes for the distribution of genetic variability has been neglected. To investigate how the current genetic variability distribution depends on the genetic variability of ancestral species, we simulated the evolution of heterozygosity on a pre-determined phylogeny under three neutral models of evolution: genetic drift, drift *vs* mutation and drift *vs* migration. The distribution of genetic variability resulting from the simulations was used to estimate the phylogenetic signal by the phylogenetic comparative method of autocorrelation. Phylogenetic signal in genetic variability was observed for each of the three models, and its intensity was generally higher and persisted longer when forces of drift, mutation and migration were reduced. The prediction of a phylogenetic signal in genetic variability has consequences for: population genetics, which must consider biological processes acting at the species level influencing the amount and distribution of genetic variability; the macroevolutionary

theory, by giving a theoretical basis for species selection by suggesting a heritability of genetic variability between species, and the meta-analyses of genetic variability, which must deal with the non-independence of species. The patterns observed in phylogenetic signal produced by different models of evolution can be used further to compare with data obtained from molecular markers. This is the first study that analyzes the theoretical expectations for the existence of a phylogenetic signal in a population genetic trait.

Key words: Population genetics; Heterozygosity; Neutral evolution; Phylogenetic comparative method; Phylogenetic autocorrelation; Macroevolution

INTRODUCTION

The amount of genetic variability in natural populations has been an important focus in the study of population genetics, especially its relationship with the microevolutionary forces acting in these populations (Altukhov and Salmenkova, 2002; Lewontin, 2002). The study of genetic variability increased in importance in the 1960s (Lewontin, 1985), with the debate between the relative importance of selection or neutral models of evolution, a debate that persists to this day (Skibinski and Ward, 1998). Genetic variability has also been considered to be an important trait at the species level, resting on the assumption that the reservoir of genetic variability acts as a hedge against extinction (Dobzhansky, 1939) and is thus an important pre-requisite for evolutionary change (Hedrick, 1986; Lewontin, 2002). Genetic variability has been studied at the species level as a good trait to show species selection (Lloyd and Gould, 1993; Lieberman and Vrba, 2005) and as a marker to study extinction risk (Frankham, 1996; Spielman et al., 2004).

Population geneticists usually consider the amount of genetic variability present in natural populations as a result of a balance among microevolutionary forces (Hedrick et al., 1976; Nevo, 1978; Hedrick, 1986). Genetic variability estimates by molecular markers are usually explained with a genetic drift *vs* mutation, or genetic drift *vs* gene flow balance. This causal relationship assumes that the effective population size and mutation or migration rate determine the amount of genetic variability (Crow and Kimura, 1970; Frankham, 1996, 1998; Garner et al., 2005). For allozymes and single nucleotide polymorphisms, natural selection has also been considered to determine the amount of genetic variability (Nevo, 1978; Hedrick, 1986; Kreitman and Akashi, 1995; Panova and Johannesson, 2004; Nielsen 2005).

Some papers and reviews have tried to establish relationships between species genetic variability and species life-history characteristics (Nevo, 1978; Mukherjee et al., 1987; Ward et al., 1992; Frankham, 1996, 1998; Spielman et al., 2004; Garner et al., 2005). Nevo (1978) found different degrees of allozyme genetic variability in species of different life zones, biotas, habitats, specialization levels, and also taxonomic groups and concluded that genetic variability can be seen as an adaptation to environment, in space and time. Ward et al. (1992) not only found relationships between allozyme genetic variability and some life-history traits but also relationships with the biochemical properties of each enzyme. More recent works (Frankham, 1996, 1998; Spielman et al., 2004; Garner et al., 2005) showed relationships between species genetic diversity and population size, inbreeding rate and extinction risk.

In spite of the generality of the proposed associations between genetic variability and species bionomic characteristics, those propositions ignored the influence of species evolutionary history and the characteristics shared at the species level as a result of ancestral-decent relationships. Species are connected, at a higher or lower degree, by a common ancestor because of the genealogical process of ancestry and decent between individuals, lineages and species (Felsenstein, 1985).

In ignoring these connections, a serious theoretical problem emerges. The evolutionary history of a species has an important role in constraining its characteristics and in the directions and rates of evolution of its traits (Cheverud et al., 1985). Since species share common ancestors, many species characteristics are shared as a consequence of common ancestral characteristics resulting in a phylogenetic dependence on the characteristics of contemporaneous species (Cheverud et al., 1985; Felsenstein, 1985).

Besides the theoretical problem, there is also a methodological one, if we assume independence when analyzing a trait in a group of species. To overcome these errors, phylogenetic comparative methods have been proposed (Cheverud et al., 1985; Felsenstein, 1985; Gittleman and Kot, 1990; Diniz-Filho et al., 1998; Blomberg et al., 2003) either to estimate the phylogenetic signal retained in a trait or to isolate the phylogenetic signal and study the correlation of two traits among species. The term phylogenetic signal has been widely used to represent the phylogenetic dependency or effect and was recently defined by Blomberg and Garland Jr. (2002) as the tendency of related species to resemble each other more than they resemble species drawn at random from a tree.

Some clues for a phylogenetic signal in species genetic variability can be found in previous studies. One clue is the correlation between intraspecific polymorphism and interespecific evolutionary rate, a prediction of the neutral theory of molecular evolution (Kimura, 1983). This correlation can be considered as evidence of the phylogenetic signal because it relates the genetic variability to the evolutionary time of species isolation, which represents the degree of phylogenetic relatedness.

A series of studies by Skibinski and collaborators explored these correlations (Ward and Skibinski, 1985; Mukherjee et al., 1987; Skibinski and Ward, 1982, 1998). They used heterozygosity as a genetic variability estimate and genetic distance (Nei, 1972) as an evolutionary rate estimate. For vertebrates, they found a correlation of 70% (Skibinski and Ward, 1982), and high correlations were also found for invertebrates and plants (Ward and Skibinski, 1985). These correlations were not sensitive to variations in reproductive mode, population sizes, gene-flow intensity, mutation rate or time of divergence (Mukherjee et al., 1987).

Simulations showed that some selective models can also produce these correlations, but neutral mutation is still the most plausible explanation (Skibinski and Ward, 1998). Skibinski, Ward and their co-authors do not invoke phylogenetic effects to explain their data, although their results provide good evidence for the existence of a phylogenetic signal in genetic variability.

Another clue can be found in the review of Nevo (1978). This author divided species into taxonomic groups for analysis and found differences in the amount of genetic variability between groups. In the review by Ward et al. (1992), the care in considering the correlation between genetic variability and genetic distance in comparing taxonomic groups produced results that differ from Nevo (1978). If we assume that there is some correspondence between taxonomy and phylogeny for the species studied, this difference suggests that part of the asso-

ciations found by Nevo (1978) can be due to a phylogenetic dependence on genetic variability and on the life-history characteristics that he used.

Considering the theoretical expectation of phylogenetic dependence between species and the clues we have from the literature, we infer that there is a phylogenetic signal in species genetic variability. To understand how this phylogenetic signal can emerge from the evolution of genetic variability, we used simulation procedures to evolve the genetic variability over a predefined topology under neutral evolutionary models. Our results are discussed in the context of phylogenetic comparative theory, macroevolution theory and population genetics theory and data, and a mechanism of phylogenetic inheritance is proposed. The patterns we observed from the simulated models can be further compared to available data on species genetic variability.

The study of a population genetic trait in a phylogenetic comparative way is a novelty, and there is just one report of phylogenetic comparative analysis for a population genetic trait, namely population structure (Duminil et al., 2007). This is the first study that analyzes the theoretical expectations for the existence of a phylogenetic signal in a population genetic trait, and also the first study that examines species genetic variability as a phylogenetic character.

MATERIAL AND METHODS

Evolutionary models

We simulated the evolution of genetic variability under three stochastic models of evolution. The first model is evolution by genetic drift (Wright, 1931), given by

$$H_t = \left[1 - \left(\frac{1}{2N_e}\right)\right] \times H_{t-1} \qquad \text{(Equation 1)}$$

where, H_t is the heterozygosity at generation t, H_{t-1} is the heterozygosity at generation $t-1$ and N_e is the effective population size.

The second model is evolution by drift and mutation (Kimura and Crow, 1964), given by

$$H_t = 1 - \left[\frac{1}{2N_e} + \left(1 - \frac{1}{2N_e}\right)(1 - H_{t-1})\right](1 - \mu)^2 \qquad \text{(Equation 2)}$$

where, μ is the mutation rate. The term $(1-\mu)^2$ then is the probability that both alleles do not mutate.

The third model is evolution by drift and migration (Crow and Kimura, 1970), given by

$$H_t = 1 - \left[\frac{1}{2N_e} + \left(1 - \frac{1}{2N_e}\right)(1 - H_{t-1})\right](1 - M)^2 \qquad \text{(Equation 3)}$$

where, M is the migration rate. This model assumes the island model of migration, and the term $(1-M)^2$ gives the probability that neither allele is a migrant. Since we simulated heterozygosity changes throughout species, in the three models we considered H_t as the heterozygosity after t generations at the species n, H_0 as the initial heterozygosity at the lineage n-1 and N_e as

the effective population size averaged for the lineage.

Parameter values for simulations

The parameter values for the three models of evolution are presented in Table 1. The three models had 4 parameters varying in simulations; for each parameter we chose three mean values resulting in 81 simulations for each model. In simulations of genetic drift, we also varied the parameter $VarN_e$, which represents variance of the distribution of N_e, from which the values of effective population size for each lineage were randomly sampled throughout the simulations. Since the variation of this parameter did not influence the results, in simulations of the other two models we fixed the variance of N_e at 0.5.

Table 1. Parameter values for the three models used in simulations.

Genetic drift			
H_0	N_e	$VarN_e$	t
0.1	100	0.01	100
0.3	1000	0.03	500
0.5	10,000	0.05	1000
Genetic drift *vs* mutation			
H_0	N_e	μ	t
0.1	100	0.0001	100
0.3	1000	0.00001	1000
0.5	10,000	0.000001	3000
Genetic drift *vs* migration			
H_0	N_e	M	t
0.1	50	0.01	50
0.3	100	0.001	100
0.5	1000	0.0001	500

H_0 is the inicial heterozygosity, N_e is the effective size of the lineages simulated, $VarN_e$ is the variance of a normal distribution allowed for the effective size, μ is the mutation rate, M is the migration rate, and t is time measured in number of generations.

The choice of values was intended to cover all the range of variation in heterozygosity over time according to the distribution determined by the models. For the genetic drift model, the greater variation in heterozygosity occurs for $N_e < 10,000$ and for $t < 1000$. In this model, we also took care to maintain t between $t = N_e / 10$ and $t = N_e$ (Crow and Kimura, 1970); this is because, for $t > N_e$, heterozygosity is not maintained in the lineage. For the drift-mutation model, the maximum value for time was reduced because, for the defined mutation values, there is almost no variation in heterozygosity after 3000 generations. The mutation rate values were defined setting $4N\mu = 0.4$, used by Skibinski and Ward (1998). For the drift-migration model, the values of N_e and t had to be even more reduced because the equilibrium was attained too fast in the values we used in the other two models. This should not be a problem, because if we take the value of N_e and t very close to equilibrium, we will have results that can be extended to all higher values.

Computational basis for simulations

To run simulations on the models of evolution described above, we programmed a routine in the MatLab 6.5 software. The input was: i) a topology, ii) the value of H_0, iii) the mean and variance of the N_e and t distributions in which the values of these parameters will be randomly sampled, and iv) the number of times this simulation had to be replicated.

The algorithm is:

1. Enter a topology in Newick notation. The next steps will refer to the topology exemplified in Figure 1;
2. Define the initial value of H_0 and the mean and variance of N_e and t distribution;
3. Calculate the value of H_1 and H_2 using a random value of N_e at each iteration;
4. Take the value of H_1 as H_0 to calculate H_3 and H_5, with a new random value of N_e at each iteration;
5. Take the value of H_2 as H_0 to calculate H_4 and H_6, with a new random value of N_e at each iteration.

After step 5, the simulation proceeds repeating steps 4 and 5 using each new heterozygosity value already defined for a node as the ancestor heterozygosity to calculate the heterozygosity of its two descents. The simulation stops when the heterozygosity value is calculated for all internal and external nodes. The simulation was replicated the number of times requested. The output gives two files: i) the values of H to each internal and external node of the topology to each replicate, and ii) a mean value to each internal and external node across all repeats.
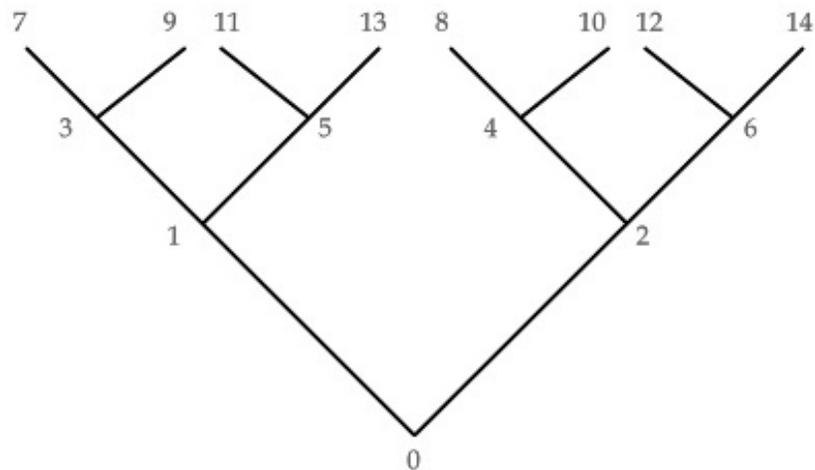

Figure 1. Symmetrical topology only with bifurcations and 8 with terminal nodes.

Topology

We defined a topology of 128 terminal nodes, with bifurcations only, and totally symmetrical. This topology has 7 hierarchical levels (Figure 2) that were analyzed, either in total,

or isolated to construct correlograms. The branch lengths were determined by the value of *t* randomly sampled in each simulation.
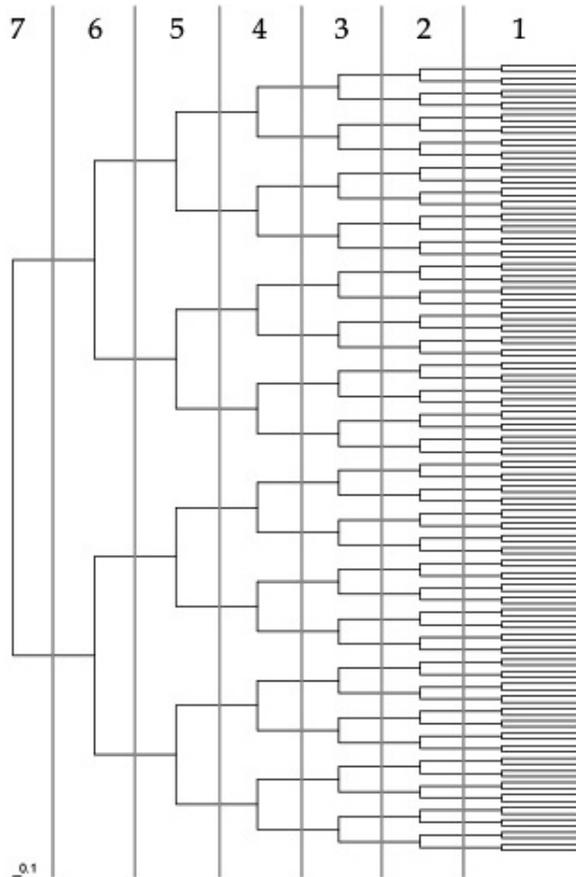

Figure 2. Comb-like topology of 128 terminal nodes used in simulations. Vertical lines define the classes of phylogenetic distance used for autocorrelation analysis.

Comparative phylogenetic analysis

We analyzed 1000 replications of each simulation, i.e., for each combination of parameter values for each of the three models. We used Moran's I, estimated by the phylogenetic comparative analysis of phylogenetic autocorrelation (Gittleman and Kot, 1990; Diniz-Filho, 2001). Moran's I is a good estimate of the phylogenetic signal retained (Gittleman and Kot, 1990; Martins et al., 2002), and it allowed us to estimate the phylogenetic signal to each simulation, and also for each of the 7 hierarchical levels of the topology, generating a phylogenetic correlogram. These analyses were executed using the Autophy 5.1 program (Diniz-Filho, 2001).

The mean value for Moran's I obtained for each combination of parameters was used in a principal components analysis to summarize correlogram profiles along a single dimension. The

first axis corresponded to 77.6% of the total variance, and its scores described a continuous change in correlogram profile, from an abrupt decay curve at higher scores to a gradual decay curve at lower values (Figure 3). Correlograms with an abrupt decay lose phylogenetic signal faster, whereas those with a gradual decay retain more phylogenetic signal (Figure 3). Differences between average correlogram profiles produced by alternative models and by different parameter combinations were tested by an analysis of variance on PC1 scores with the Bonferroni *post hoc* test.
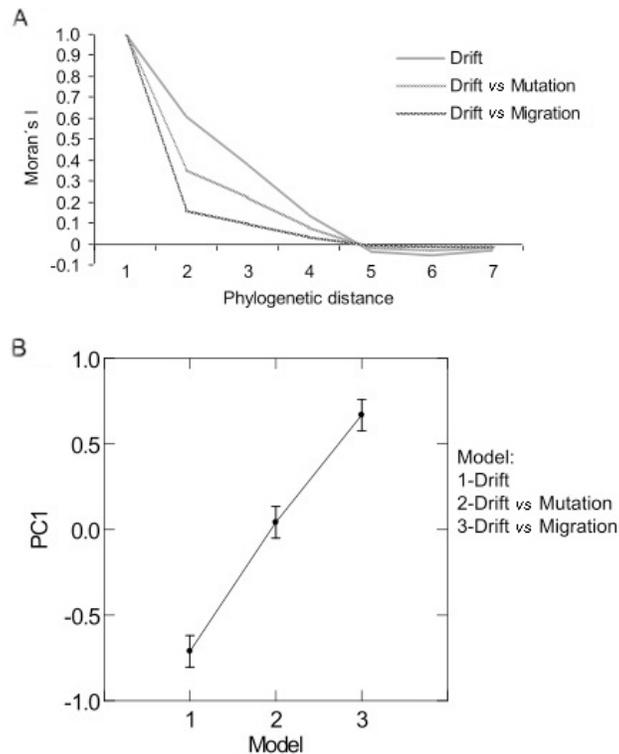

Figure 3. A. Average correlograms for the three simulated models of evolution. **B.** Scores obtained for factor 1, represented by the PC1 axis, of principal components analysis for each model studied.

RESULTS

All three simulated models showed some amount of phylogenetic signal for genetic variability. This signal increases in strength from the ancestral to the derivate nodes as expected (Diniz-Filho, 2001). For the first class of distance the phylogenetic signal measured with Moran's I was always equal to 1 (Figure 3A).

Genetic drift

Just the variation in two parameters, N_e and t, influenced the phylogenetic signal in genetic drift simulations (F = 70.991, P = 0.00, respectively; Figure 4). The variation

in the initial heterozygosity and in the variance of N_e had no effect for the estimates of phylogenetic signal (F = 0.00 and P = 1 for both). The phylogenetic signal is retained more with an increase in the values of N_e and with a decrease in the values of t (Figure 4). The opposite influence of N_e and t is a consequence of their inverse influence on heterozygosity, as shown by Equation 1. For the phylogenetic signal, the influence of the highest N_e and the lowest t is similar, whatever the values are for the other parameters, and the correlogram profile is linear until it crosses zero.
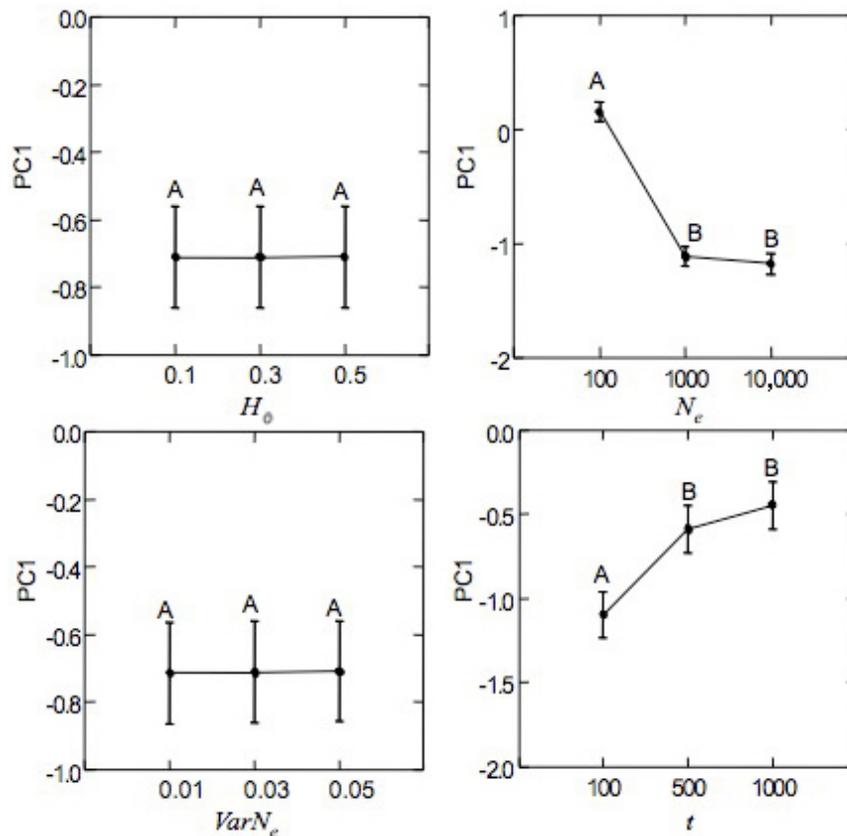

Figure 4. PC1 scores obtained for each parameter value used in simulations of genetic drift model. The different letters over the PC1 data on scores distribution represent significant differences at the 0.05 level (ANOVA). For abbreviations, see legend to Table 1.

Genetic drift *vs* mutation

When mutation was introduced into the genetic drift model (Equation 2), this new parameter influenced the observed phylogenetic signal (F = 3.324, P = 0.041) and changed the shapes of the correlogram profiles for N_e and t (F = 21.211, P = 0.00 and F = 23.445, P = 0.00), mainly for high mutation rate values (Figure 5). We also tested for an interaction between mutation and effective population size, but found that it was not significant (F = 1.354, P = 0.259; Figure 6).

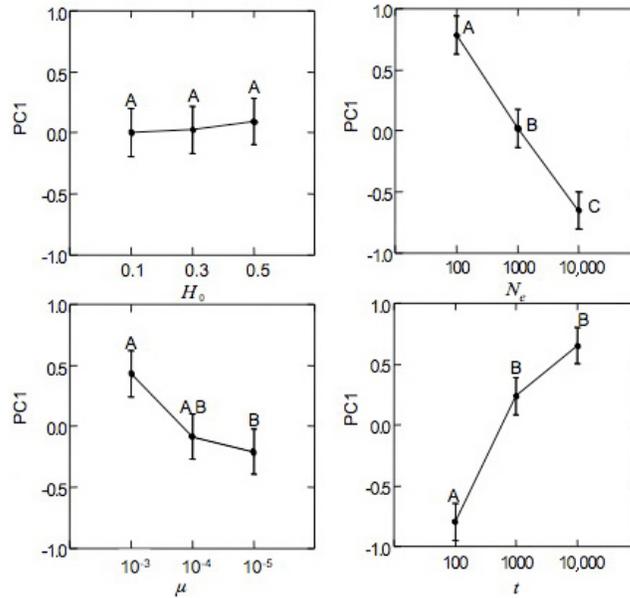**Figure 5.** PC1 scores obtained for each parameter value used in simulations of genetic drift *vs* mutation model. The different letters over the PC1 data on scores distribution represent significant differences at the 0.05 level (ANOVA). For abbreviations, see legend to Table 1.
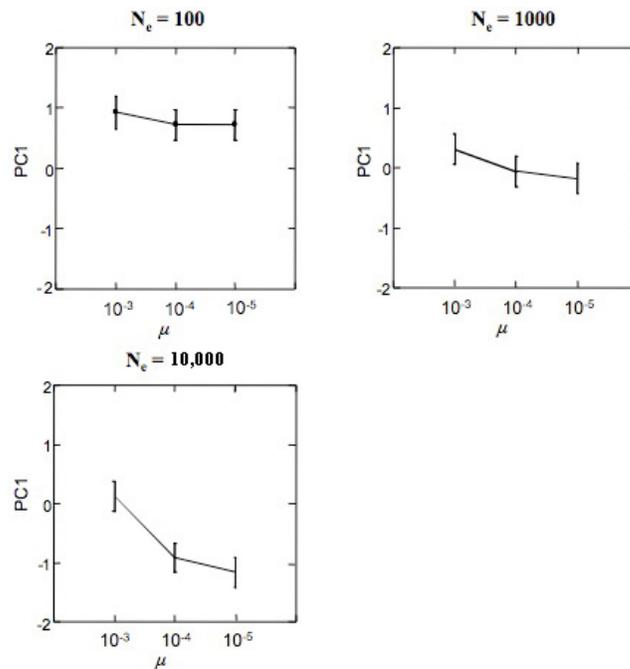

Figure 6. Interaction effect of effective size (N_e) with mutation rate (μ) in simulations of genetic drift *vs* mutation model.

Genetic drift *vs* migration

The role of migration counterbalancing genetic drift is similar to mutation because of the way the model was defined. However, migration rates are usually higher than mutation rates and the influence of migration rates in heterozygosity evolution is stronger. To study the effect of migration we had to increase the intensity of drift. All correlograms resulting from genetic drift *vs* migration simulations exhibited an abrupt decay of phylogenetic signal (F = 21.412, P = 0.00), at least for the higher migration rate (Figure 7). The effect of N_e and t was also high (F = 8.207, P = 0.001 and F = 6.693, P = 0.002). The interaction between migration and effective population size was significant (F = 7.514, P = 0.00) (Figure 8). In a scenario where drift is very strong (Figure 8, N_e = 50), the effect of migration in phylogenetic signal is very low. As we reduce the action of drift, the effect of migration increases (Figure 8, N_e = 100 and N_e = 1000) and the loss of phylogenetic signal is more abrupt through phylogenetic distance classes at higher migration rates.
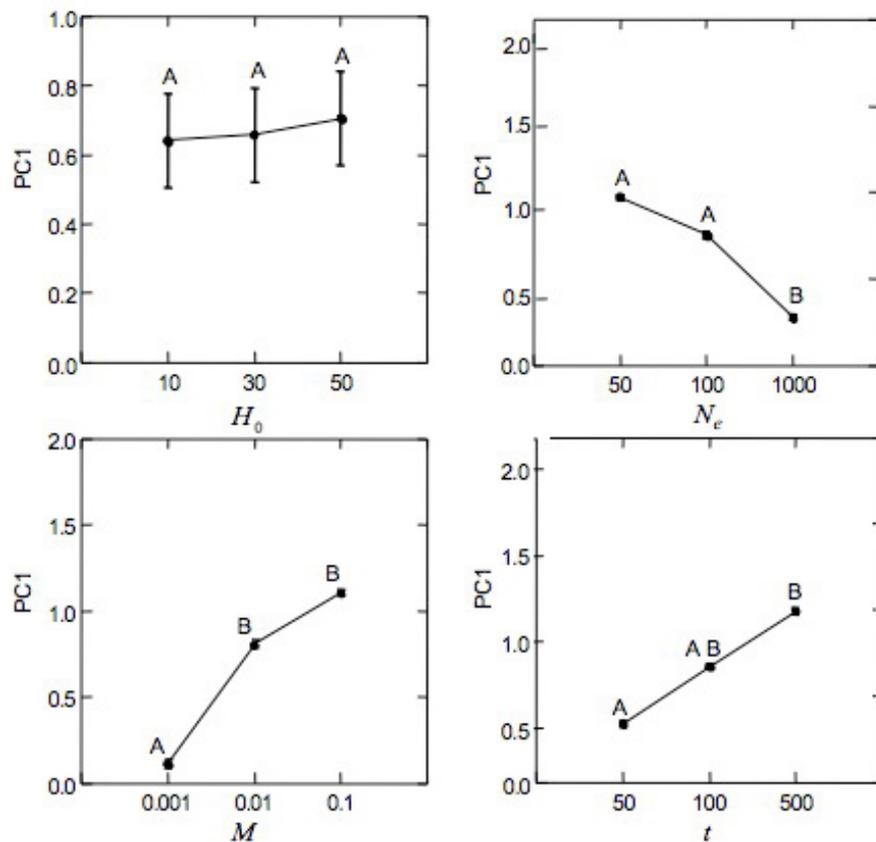

Figure 7. PC1 scores obtained for each parameter value used in simulations of genetic drift *vs* migration model. The different letters over the PC1 data on scores distribution represent significant differences at the 0.05 level (ANOVA). For abbreviations, see legend to Table 1.
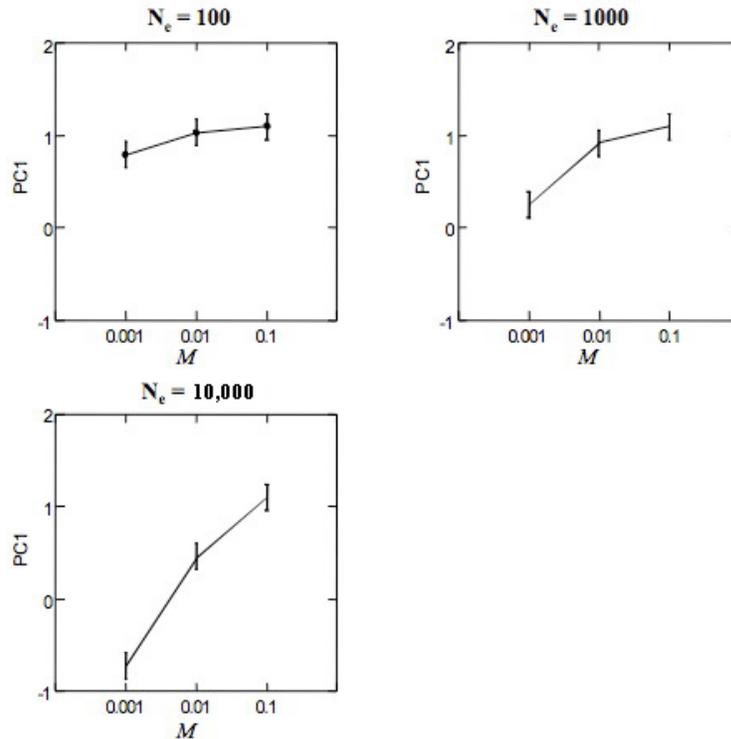
Figure 8. Interaction effect of effective population size (N_e) with migration rate (M) in simulations of genetic drift *vs* migration model.

The importance of species lifetime

In simulations of the three different models, one pattern was common to the parameters μ, M and t: the larger the parameter value, the faster the phylogenetic signal was lost. This reduction of phylogenetic signal is a simple consequence of the fact that higher values for parameters drive the system to equilibrium heterozygosity within species in a few generations and almost all OTUs of a branch end with very similar heterozygosity values.

As an example, we can see the correlogram and the distribution of heterozygosity among phylogenetic distance classes in Figure 9 that correspond to a drift *vs* mutation simulation with $N_e = 1000$ and $t = 1000$ and $H_0 = 0.1$. For $\mu = 10^{-3}$, we had a rapid increase in heterozygosity from 0.1 to 0.8 in the first tree bifurcation. Since the maximum value for heterozygosity in this case is 0.83 (the equilibrium point), the evolution in subsequent lineages had just a small possibility of varying between lineages, and groups became very homogeneous very rapidly, leading to the fixation of the value of heterozygosity in the distance 4 class. In this class, the phylogenetic signal disappears. For $\mu = 10^{-3}$, the initial increase in heterozygosity is high, from 0.1 to 0.46, but it still leaves more space for variation until the maximum value, 0.52, is reached. In this case, there was variation in means between nodes of all classes of distance, and the phylogenetic signal just disappears only in class five.
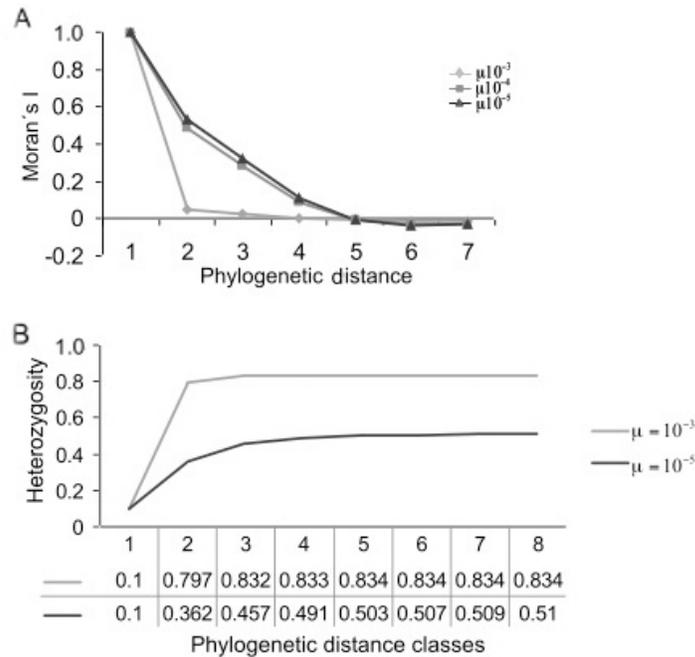
Figure 9. Results obtained in simulation of drift *vs* mutation model with parameter values of $H = 0.1$, $N_e = 100$ and $t = 1000$. **A.** Correlogram and **B.** Distribution of mean heterozygosity by phylogenetic distance class. Class 8 represents tip values.

DISCUSSION

Phylogenetic comparative analysis is usually applied to phenotypic traits that have a quantitative genetic base (such as Brownian motion and the O-U process) and its evolution is usually approximated by general models of evolution (Felsenstein, 1985, 1988; Hansen and Martins, 1996). The use of evolutionary models specifically constructed for the genetic basis of the evolution of genetic variability certainly gave us more powerful and precise results.

In contrast to that observed for statistical models such as Brownian motion (Gittleman and Kot, 1990; Diniz-Filho, 2001; Martins et al., 2002), our models based on real biological properties of the trait do not lead to linear decreases in phylogenetic signal. In our simulations, it depends on the parameter values. This non-linear decrease was also observed for some neutral models of phenotypic trait evolution studied by Hansen and Martins (1996).

Inheriting genetic variability

The process that retains phylogenetic signal in genetic variability can be called a "phylogenetic inheritance" of this variability because it is consequence of the ancestor-descent relationships between individuals, populations and species. The phylogenetic inheritance of genetic vari-

ability could occur by at least three main processes: i) during speciation a part of genetic variability can be maintained from ancestor to descent species; ii) similar mutation rates in phylogenetically close species can help to maintain the same level of variability throughout the species life, and iii) effects of phylogenetic signal in life history traits (Blomberg et al., 2003) (such as dispersion and environmental tolerance) that influence populations demographically and, as a consequence, the amount of variation can also create a phylogenetic signal in genetic variability.

Predictions for natural species

Population genetic traits have never been used as a phylogenetic character and analyzed in a phylogenetic comparative way since the recent paper of Duminil et al. (2007). Duminil et al. reevaluated the relationships between genetic structure and species traits in seed plants. The phylogenetic signal in the genetic structure was estimated by using taxonomy as an approximation of the phylogenetic relationships, and the estimates were up to 79% of variance in genetic structure, explained by the taxonomic hierarchy. Although the phylogenetic signal may be overestimated by the use of taxonomy instead of phylogeny, the results of Duminil et al. (2007) point to the need to study and analyze population genetic traits in a phylogenetic comparative way.

Considering the results we obtained, we can make some considerations and predictions of the phylogenetic signal that can be observed in natural species. For the simulated models of evolution and for the parameter values we used, the heterozygosity in a species will almost reach the equilibrium value in a species lifetime. However, we allowed variation just in effective population size, which easily varies between species and seem to be related to heterozygosity in nature (Frankham, 1996). In natural populations and species, mutation rate, migration rate and lifetime must also vary between species, increasing heterogeneity between species. The presence of other forces such as natural selection and demographic events will also help to increase the dynamics of heterozygosity change in time (Crow and Kimura, 1970).

Besides the main importance population geneticists have attributed to the stochastic process for the distribution of genetic variability, many population genetic works, including some reviews, gave natural selection large importance on the maintenance of genetic variability, either in microevolution (Nevo, 1978; Hilbish and Koehn, 1985; Hedrick, 1986; Kreitman and Akashi, 1995; Eanes, 1999; Panova and Johannesson, 2004) or in macroevolution (Lloyd and Gould, 1993; Lieberman and Vrba, 2005). Simulations of the O-U process for phenotypic traits are used to represent evolution by stabilizing selection and usually retain lower phylogenetic signal than stochastic models (Felsenstein, 1988). As we reported here for mutation and migration, the higher the strength of selection, the lower the retained phylogenetic signal was (Hansen and Martins, 1996; Diniz-Filho, 2001; Martins et al., 2002).

The dynamics of microevolutionary forces will define the change of heterozygosity in time and its equilibrium point, which can be the loss of heterozygosity within the species lifetime, as we observed in some simulations. In conditions where heterozygosity reaches an equilibrium point within the species lifetime, we can predict the estimate of lower phylogenetic signal simply because there is just a little or no variation between species to test. However, in a natural scenario the action of forces other than the ones we tested here must occur and an equilibrium point for heterozygosity must be improbable to be reached. Therefore, we predict that in natural species the phylogenetic signal estimates may be generally higher than the estimates for the simulated species.

The pattern produced by the three models we used can be used further to compare the expected phylogenetic signal for a given phylogeny and what is observed from data of real species. This procedure is common for phenotypic traits using the Brownian motion as evolutionary neutral model (Felsenstein, 1981; Martins, 1994). For genetic variability, we are now able to make the same comparisons using precise models and to obtain more powerful results.

Phylogenetic signal in genetic variability

In spite of the variation in phylogenetic signal retained under different evolutionary models, we have to emphasize that it must exist for genetic variability and must not be ignored. There are some consequences for the existence of a phylogenetic signal in genetic variability, for different levels of biological organization, the level of populations and the level of species, and also for statistic analysis of genetic variability data. At the population level, genetic variability must evolve from a starting point determined by the biological processes acting at the species level. It means that not all the variability values we observe within a species can be attributed to its life-history traits and that between populations of the same species we could also have some non-independence for genetic variability.

At the species level, the existence of a phylogenetic signal in genetic variability points to a heritability of genetic variability between species. This heritability would be a basis for a species selection process. Lloyd and Gould (1983) and Gould (2002) treat variability as a case of species selection in the broad sense but a genuine case under the emergent fitness approach. This is a controversial case, as Lieberman and Vrba (2005) argued that genetic variability could be a case of species selection in the narrow sense if there is heritability at the species level. However, they conclude that "most differences in variability between clades are evanescent and either are not passed on to descent species or have no causal bearing on species sorting". It has been argued that the nonrandom species sort of genetic variability would be a by-product of other traits heritable at the species level (Lieberman and Vrba, 2005).

Our results give a theoretical basis for the existence of an ancestral to descent maintenance of genetic variability, producing a heritability of genetic variability itself, making possible the species selection guided by this trait in a narrow sense as an emergent species-level character. The causal bearing on species sorting is still considered controversial by Lieberman and Vrba (2005) in spite of the evidence for influence of genetic variability on species extinction (Spielman et al., 2004).

In analyzing data statistically, the fact that actual species are connected by an underlying genealogical process increases the type I error. Species are not independent units as supposed by classic statistical methods; they have a degree of dependence that directly reflects their phylogenetic relationships (Felsenstein, 1985). In comparing characteristics between species, the use of any phylogenetic comparative method gives more precise results than the use of traditional statistics (Martins et al., 2002).

In previous analyses of the distribution of genetic variability among species, statistical errors could have occurred not only as a consequence of the phylogenetic signal in genetic variability but also as a consequence of phylogenetic signal in the life-history traits used in comparisons (Blomberg et al., 2003). The effect of phylogeny on life-history traits related to genetic variability would be one of the factors for the phylogenetic inheritance of genetic

variability itself, and can overestimate these relationships in traditional analyses (Blomberg et al., 2003; Garland et al., 2005).

Part of the associations between genetic variability and life-history traits found in the studies of Nevo (1978), Nevo et al. (1983), Mukherjee et al. (1987), Ward et al. (1992), Frankham (1996, 1998), Spielman et al. (2004), and Garner et al. (2005) could be attributed to the phylogenetic relationships connecting the species analyzed. Even in researchs that had taken care to divide species into taxonomic groups for analysis we can identify two problems: taxonomy does not always correspond to species phylogeny, and the highest phylogenetic signal must be between close-related species for which taxonomy is poorly informative (Felsenstein, 1985).

Our results presented evidence not only for a phylogenetic signal in genetic variability but also for differences in correlogram profiles produced by different evolutionary models and by different combinations of their genetic and demographic parameters. Considering all studies in the literature that have reported on phylogenetic comparative analysis and the results we obtained here, we suggest that previous analyses on the distribution of genetic variability among species and its association with other ecological and life-history traits must be viewed with caution. Future analysis could use the results we obtained to compare the correlogram profile for genetic variability data of a specific group of organisms with those we obtained, by using similar statistical analysis.

ACKNOWLEDGMENTS

REFERENCES

Altukhov YP and Salmenkova EA (2002). DNA polymorphism in population genetics. *Russ J. Genet.* 38: 989-1008.

Blomberg SP and Garland T Jr (2002). Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods. *J. Evol. Biol.* 15: 899-910.

Blomberg SP, Garland T Jr and Ives AR (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57: 717-745.

Cheverud JM, Dow MM and Leutenegger W (1985). The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body weight among primates. *Evolution* 39: 1335-1351.

Crow JF and Kimura M (1970). An Introduction to Population Genetics Theory. Harper and Row, New York.

Diniz-Filho JA (2001). Phylogenetic autocorrelation under distinct evolutionary processes. *Evolution* 55: 1104-1109.

Diniz-Filho JA, de Sant'ana CER and Bini LM (1998). An eigenvector method for estimating phylogenetic inertia. *Evolution* 52: 1247-1262.

Dobzhansky T (1939). Genetics and the Origin of Species. Columbia University Press, New York.

Duminil J, Fineschi S, Hampe A, Jordano P, et al. (2007). Can population genetic structure be predicted from life-history traits? *Am. Nat.* 169: 662-672.

Eanes WF (1999). Analysis of selection on enzyme polymorphisms. *Annu. Rev. Ecol. Syst.* 30: 301-326.

Felsenstein J (1981). Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. *Evolution* 35: 1229-1242.

Felsenstein J (1985). Phylogenies and the comparative method. *Am. Nat.* 125: 1-15.

Felsenstein J (1988). Phylogenies and quantitative characters. *Annu. Rev. Ecol. Syst.* 19: 445-471.

Frankham R (1996). Relationship of genetic variation to population size in wildlife. *Conserv. Biol.* 10: 1500-1508.

Frankham R (1998). Inbreeding and extinction in island populations. *Conserv. Biol.* 12: 665-675.

Garland T Jr, Bennett AF and Rezende EL (2005). Phylogenetic approaches in comparative physiology. *J. Exp. Biol.* 208: 3015-3035.

Garner A, Rachlow JL and Hicks JE (2005). Patterns of genetic diversity and its loss in mammalian populations. *Conserv. Biol.* 19: 1215-1221.

Gittleman JL and Kot M (1990). Adaptation: Statistics and a null model for estimating phylogenetic effects. *Syst. Zool.* 39: 227-241.

Gould SJ (2002). The Structure of Evolutionary Thought. Harvard University Press, Cambridge.

Hansen TF and Martins E (1996). Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution* 50: 1404-1417.

Hedrick PW (1986). Genetic polymorphism in heterogeneous environments: A decade later. *Annu. Rev. Ecol. Syst.* 17: 536-566.

Hedrick PW, Ginevan ME and Ewing EP (1976). Genetic polymorphism in heterogeneous environments. *Annu. Rev. Ecol. Syst.* 7: 1-32.

Hilbish TJ and Koehn RK (1985). The physiological basis of natural selection at the Lap locus. *Evolution* 39: 1302-1317.

Kimura M (1983). The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge.

Kimura M and Crow JF (1964). The number of alleles that can be maintained in a finite population. *Genetics* 49: 725-738.

Kreitman M and Akashi H (1995). Molecular evidence for natural selection. *Annu. Rev. Ecol. Syst.* 26: 403-422.

Lewontin RC (1985). Population genetics. *Annu. Rev. Genet.* 19: 81-102.

Lewontin RC (2002). Directions in evolutionary biology. *Annu. Rev. Genet.* 36: 1-18.

Lieberman BS and Vrba ES (2005). Stephen Jay Gould on species selection: 30 years of insight. *Paleobiology* 31: 113-121.

Lloyd EA and Gould SJ (1993). Species selection on variability. *Proc. Natl. Acad. Sci. U. S. A.* 90: 595-599.

Martins EP (1994). Estimating rates of character change from comparative data. *Am. Nat.* 144: 193-209.

Martins EP, Diniz-Filho JA and Housworth EA (2002). Adaptive constraints and the phylogenetic comparative method: a computer simulation test. *Evolution* 56: 1-13.

Mukherjee M, Skibinski DO and Ward RD (1987). A simulation study of the neutral evolution of heterozygosity and genetic distance. *Heredity* 58 (Pt 3): 413-423.

Nei M (1972). Genetic distance between populations. *Am. Nat.* 106: 283-292.

Nevo E (1978). Genetic variation in natural populations: patterns and theory. *Theor. Popul. Biol.* 13: 121-177.

Nevo E, Beiles A and Ben-Shlomo R (1983). The Evolutionary Significance of Genetic Diversity: Ecological, Demographic and Life-history Correlates. In: Evolutionary Dynamics of Genetic Diversity (Mani GS, ed.). Springer-Verlag, Manchester, 13-213.

Nielsen R (2005). Molecular signatures of natural selection. *Annu. Rev. Genet.* 39: 197-218.

Panova M and Johannesson K (2004). Microscale variation in AAT (aspartate aminotransferase) is supported by activity differences between upper and lower shore allozymes of *Littorina saxatilis*. *Mar. Biol.* 144: 1157-1164.

Skibinski DOF and Ward RD (1982). Correlations between heterozygosity and evolutionary rate of proteins. *Nature* 298: 490-492.

Skibinski DOF and Ward RD (1998). Are polymorphism and evolutionary rate of allozyme proteins limited by mutation or selection? *Heredity* 81: 692-702.

Spielman D, Brook BW and Frankham R (2004). Most species are not driven to extinction before genetic factors impact them. *Proc. Natl. Acad. Sci. U. S. A.* 101: 15261-15264.

Ward RD and Skibinski DO (1985). Observed relationships between protein heterozygosity and protein genetic distance and comparisons with neutral expectations. *Genet. Res.* 45: 315-340.

Ward RD, Skibinski DOF and Woodwark M (1992). Protein heterozygosity, protein structure, and taxonomic differentiation. *Evol. Biol.* 26: 73-159.

Wright S (1931). Evolution in Mendelian populations. *Genetics* 16: 97-159.