Molecular Evolutionary Analysis for Estimating the Strength of Fluctuating Selection among Individuals (FSI)

Xun Gu[1,2,3*]

[1]Department of Genetics, Development and Cell Biology, [2] The Laurence H. Baker Center in Bioinformatics on Biological Statistics, [3]Program of Ecological and Evolutionary Biology, Iowa State University, Ames, IA 50011, USA.

*Corresponding author: xgu@iastate.edu

Abstract

Fluctuating selection among individuals (FSI) refers to any mutation that exhibits different fitness effects among individuals. Thus, the selection nature of a mutation (deleterious, neutral, or beneficial) should be interpreted by the means of the population average. For instance, a neutral mutation on average could be slightly deleterious in some individuals, and slightly beneficial in others. It has been recently demonstrated that the effect of FSI is important in molecular evolution especially when the effective population size (N_e) is not small. Intriguingly, a novel pattern of molecular evolution called 'selection duality', i.e., mutations that are statistically slightly beneficial are subject to a negative selection, emerges under the condition that selective advantage is less than FSI. While FSI sheds some lights on the long-term neutralist-selectionist debate, an immediate question is how to calculate the strength of FSI-genetic drift relative to the N_e-genetic drift. In this article we develop a statistical method the relative FSI strength (F): if F is close to 0, the N_e-genetic drift is dominant; whereas the FSI-genetic drift is dominant if F is close to 1. One may tentatively set F=0.5 as an empirical criterion to weigh between those two genetic drifts. Our case study showed that the relative FSI-strength F is over 0.5 in most species, suggesting that the FSI-genetic drift, rather than the N_e-drift, plays a major role in metazoan genome evolution.

Introduction

Galtier (2024) provided a historical perspective of the neutralist-selectionist debate in molecular evolution (Kimura 1968; Kern and Hahn 2018; Jensen et al. 2019), and highlighted how the debate has evolved from its origins to influence various aspects of modern evolutionary biology. Meanwhile, de Jong et al. (2024) evaluated the validity of both neutralist and selectionist arguments, finding that many of the tested predictions were sensitive to the underlying assumptions and can be consistent with both viewpoints. Thus, this central debate has casted some doubts whether the current wisdom of population genetics that underlies molecular evolution is theoretically sufficient; a recently-growing consensus between neutralists and selectionists. It has been indeed highly desirable to perform existed model examination and new model development (Hahn 2008; Hughes 2008; Wagner 2008; Nei et al. 2010; Munoz-Gomez et al. 2021; Gu 2021).

In population genetics, fixation models attempt to describe how a new mutation can spread in a population, ultimately toward being fixed rather than being lost, due to either selection or drift (Kimura 1962; 1983); one may see McCandlish and Stoltzfus (2014) for a comprehensive review. An implicit assumption of the fixation model is that any mutation has the same fitness among those individuals with the same genotype (homozygotes or heterozygotes). That said, a neutral mutation is selectively neutral for all individuals who carry the mutation, and so forth a deleterious or beneficial mutation. It should be noticed that human genetics revealed that mutations frequently exhibit different effects on individuals, challenging the fixed view of the mutational effects. In some studies, those effects were attributed to modifier genes (Nadeau 2001; Riodan and Nadeau 2017). For instance, incomplete penetrance occurs when a mutation shows deleterious effects in some individuals but not others; indeed, not all individuals with the same mutation show the disease phenotype (Riordan and Nadeau 2017; Eldar et al. 2009; Raj et al. 2010; Jensen et al. 2025). Moreover, Eldar et al. (2009) found the effect of partial penetrance on developmental evolution in bacteria.

In particular, fluctuating selection among individuals (FSI) refers to any mutation that exhibits different fitness effects among individuals. In other words, the selection nature of a mutation should be interpreted by the means of the population average. For instance, one may anticipate that a neutral mutation could be slightly deleterious in some individuals, and slightly beneficial in others, but, on average, it is neutral. The population genetics of fluctuating selection between generations (FSG) has been extensively studied around 1970's (Karlin and Lerikson 1974; Karlin and Lieberman 1974); one may see Gu (2024) for a recent commentary, but the issue of FSI was largely unexplored. The reason is mainly due to the presumption that the effect of FSI must be negligible in molecular evolution, as nearly-neutral mutations (slightly deleterious or beneficial) have played a major role on shaping the pattern.

Gu (2021) studied this FSI-problem, demonstrating that the effect of FSI is important in molecular evolution especially when the effective population size (N_e) is not small. Yet, the model of FSI in Gu (2021) was found inaccurate. Recently, Gu (2025) addressed this issue, and conducted an extensive population genetics analysis to evaluate the effects of FSI on sequence divergence between species and genetic diversity within population. Intriguingly, a novel pattern of molecular evolution called 'selection duality' emerges: mutations that are statistically slightly beneficial are subject to negative selection, revealing a slow pace of molecular evolution in large population, and a signal of positive selection by the MK test simultaneously. Note that the former is the landmark evidence for the neutralist-view, whereas the latter is that for the selectionist-view. The condition for the emergence of selection-duality, intuitively speaking, is that the degree of selective advantage must be less than the effect of FSI. It means that in the case of no FSI, the selection-duality would be reduced to the strict neutrality.

Overall, it is theoretically plausible that FSI sheds some lights on the long-term neutralist-selectionist debate. An immediate question, apparently, is how to calculate the relative effect of FSI to the classical N_e-genetic drift, which determines the range broadness of the selection duality. In the case of no FSI or N_e is very small, selection-duality would vanish as it is reduced to a single point of strict neutrality.

In this article, I formulate a FSI-model of molecular evolution that established an inverse relationship between the ratio of substitution rate to mutation rate (λ/v) and the logarithmic effective population size ($\ln N_e$). Compared with the classical nearly-neutral models of Ohta (1973), the FSI-model have three advantages: (i) it covers selection-duality mutations and deleterious mutations; (ii) it considers the effect of invariable sites; and (iii) it includes a non-zero rate component in a very large population. After using the conventional d_N/d_S (the ratio of nonsynonymous rate to synonymous rate) as a proxy of λ/v, and π_S (the genetic diversity of synonymous mutations) as a measure of N_e, we formulated the $d_N/d_S \sim \ln \pi_S$ analysis across a number of species, based upon which one can predict the strength of FSI for each species. Our results are exemplified by a genome-wide $d_N/d_S \sim \ln \pi_S$ analysis in metazoans and the relative importance of FSI-genetic drift to the N_e-genetic drift in molecular evolution is discussed.

Results

Wright-Fisher model under FSI

Consider a random mating population of a monoecioys diploid organism. In a finite population, each individual produces a large number of offspring and that exactly N of those survive to maturity. Let a and A be the mutant and wild-type alleles at a particular locus, respectively, whose fitness effects are additive. The FSI model (fluctuating selection between individuals) postulates that the fitness effect of mutant a is fluctuating among individuals, whereas that of wild-type A remains a constant. Thus, the relative fitness of genotype AA, Aa, or aa is, on average, given by 1, $1+\bar{s}$ and $1+2\bar{s}$, respectively; $\bar{s}$ is the mean selection coefficient (s) of mutant a. Let $V_{\bar{s}}$ be the variance of $\bar{s}$ and $Var(s)$ be the variance of s, respectively. As the number of mutant a is $2Nx$, where x is the frequency of mutant a, by the basic statistics we obtain

$$V_{\bar{s}} = \frac{Var(s)}{2Nx} = \frac{\varepsilon_{mut}^2}{x} \quad (1)$$

where ε_{mut}^2 is called the FSI-coefficient; a large values means a strong FSI and *vice versa*; the subscript indicates mutant.

Gu (2025) developed a diffusion model to study the Wright-Fisher model under FSI, where the infinitesimal mean $\mu(x)$ and the infinitesimal variance $\sigma^2(x)$ are, respectively, given by

$$\begin{aligned} \mu(x) &= (\bar{s} - 2\varepsilon_{mut}^2)x(1-x) \\ \sigma^2(x) &= \left[\frac{1}{2N_e} + \varepsilon_{mut}^2(1-x)\right] x(1-x) \end{aligned} \quad (2)$$

where N_e is the effective population size. While $\mu(x)$ describes how selection factors influence the mean frequency change, $\sigma^2(x)$ describes the strength of genetic drifts. Two features are mentioned. First, the FSI-coefficient ε_{mut}^2 may negatively influence the mean change of gene frequency. Second, in addition to the N_e-genetic drift, the FSI-genetic drift may provide a second resource of random effects in molecular evolution.

Consequently, the diffusion model of FSI takes three forces into account: (*i*) the direct selection, measured by $\bar{s}$, the mean selection coefficient of the mutant; (*ii*) the N_e-genetic drift, the well-known Wright-Fisher sampling inversely related to N_e; and (*iii*) the FSI-genetic drift characterized by ε_{mut}^2, the FSI-coefficient, which measures the over-dispersed progeny variation among those mutant-carrying individuals. One may further anticipate that FSI-genetic drift should be dominant in a large population where N_e-genetic drift becomes negligible.

The rate of nucleotide substitution under FSI

The rate of nucleotide substitution (λ) plays a central role in the study of molecular evolution. Under fixation models, mutations are usually described as random events that have a rate of occurrence and a contribution to fitness. Suppose we have a diploid population of size N (the census population size) and we assume that mutations appear with a rate of v. Then the total rate that such mutations appear in the population as a whole is simply $2Nv$. Since only a fraction of these mutations will eventually go to fixation, to calculate the rate of molecular evolution (λ), we need to multiply $2Nv$ with the

fixation probability of a new mutation (Kimura 1962). Let $u(p)$ be the fixation probability of a mutation in the population, with the initial frequency p. Then, the formula of substitution rate can be concisely written as follows

$$\lambda = 2Nv \times u(1/2N) \quad (3)$$

where the initial frequency set by $1/(2N)$ is based on the assumption of rare, single *de novo* mutation event. In addition to the initial frequency, the fixation probability is dependent of the selection coefficient (s) of the mutation: it becomes high when the mutation is beneficial ($s>0$) or low when the mutation is deleterious ($s<0$).

The key step to solve Eq.(2) is to derive the analytical form of fixation probability. As detailed in Methods, we showed that, in addition to the mutation rate (v), the substitution rate (λ) is determined by two variables. The first one is the selection-FSI ratio ρ, defined by the (two-fold) ratio of the (FSI-adjusted) selection coefficient ($\tau = \bar{s}$-$2\varepsilon_{mut}^2$) to the FSI-coefficient (ε_{mut}^2), that is,

$$\rho = \frac{2\tau}{\varepsilon_{mut}^2} = \frac{2\bar{s}}{\varepsilon_{mut}^2} - 4 \quad (4)$$

Intuitively speaking, the selection-FSI ratio ρ measures the relative magnitude of selection to the fluctuation.

FSI emerges as a new resource of genetic drift, measured by the FSI-strength, $2N_e\varepsilon_{mut}^2$. Note that a large value of ε_{mut}^2 indicates a strong FSI, or *vice versa*. Moreover, reading the FSI-strength by $\varepsilon_{mut}^2 \div (\frac{1}{2N_e})$, one may claim a dominant FSI-genetic drift when $2N_e\varepsilon_{mut}^2>1$, or a dominant N_e-genetic drift when $2N_e\varepsilon_{mut}^2<1$. Sometimes it is concise to use a relative measure of FSI-strength (F), as given by

$$F = \frac{2N_e\varepsilon_{mut}^2}{1 + 2N_e\varepsilon_{mut}^2} \quad (5)$$

Apparently, F increase from 0 at $2N_e\varepsilon_{mut}^2 = 0$ (no-FSI), which approaches to 1 when $2N_e\varepsilon_{mut}^2 \gg 1$. It follows that the basic form of the substitution rate (λ) can be concisely written by

$$\lambda = v\frac{F(\rho+1)}{1-(1-F)^{\rho+1}} \quad (6)$$

Fig.1 illustrates the impact of FSI on the substitution rate. It appears the substitution rate decreases with the increasing of the relative strength of FSI. Moreover, when $F \to 1$, the substitution rate reaches a positive value in the case of -1<$\rho < 0$), implying that the rate is nonzero even in a very large population.

Eq.(6) extends the golden standard of neutrality when $\varepsilon^2_{mut}>0$: the neutrality ($\lambda=v$) holds when $\rho=0$, as well as the positive selection ($\lambda>v$) when $\rho>0$ or the negative selection ($\lambda<v$) when $\rho<0$, respectively. In particular, the classical formula of molecular evolution (Kimura 1962) is a special case of $\varepsilon^2_{mut}\to 0$ whereas $\rho\varepsilon^2_{mut} \to 2s$, leading to

$$\lambda = v\frac{4N_e s}{1-e^{4N_e s}} \quad (7)$$

Putting together, under FSI ($\varepsilon^2_{mut}>0$) the substitution rate-mutation rate ratio (λ/v) is determined by two functionally-independent variables, ρ and F, which converge into a single variable $4N_e s$ as $\varepsilon^2_{mut}\to 0$.

Emergence of selection-duality

One unexpected result of Eq.(6) is that when $\bar{s}=0$ under FSI, we obtain $\rho=-4$ and so the substitution rate (λ) is always less than the mutation rate (v). That is, a generically neutral mutation, i.e., the selection coefficient is zero on average, is subject to a negative selection. On the other hand, a mutation is subject to a neutral evolution ($\lambda=v$) only when $\rho=0$, or $\bar{s} = 2\varepsilon^2_{mut}$, which means that this mutation must be, on average, slightly beneficial. These facts challenged the traditional wisdom of neutral theory. One may call the case of $\rho=0$ (or $\bar{s} = 2\varepsilon^2_{mut}>0$) *the fixation neutrality*, where the substitution rate equals to the mutation rate ($\lambda=v$), corresponding to the neutrality in molecular evolution. Meanwhile, *the generic neutrality* refers to the case of $\rho=-4$ (or $\bar{s}=0$), corresponding to the neutrality in population genetics.

Intriguingly, a novel phenomenon called *selection-duality*, i.e., $-4 < \rho < 0$, emerges between the generic neutrality and the fixation neutrality: a slightly beneficial

mutation defined by $0 < \bar{s} < 2\varepsilon_{mut}^2$ is subject to a negative selection ($\lambda<v$). The arising of selection-duality is due to the statistical nature of FSI, where its distribution can be divided into the negative side ($s<0$) and the positive side ($s>0$). The generic neutrality ($\bar{s}=0$) simply means an equal proportion for both sides. The effect of negative side to reduce the fixation probability is stronger than that of positive side to increase, resulting in a negative selection on generic neutrality. Under fixation neutrality, the proportion of positive side is slightly larger than the proportion of negative side so that the net effect on fixation probability is just zero.

In addition to the generic neutrality and the fixation neutrality as boundaries for selection-duality, Gu (2025) showed the middle-point of selection-duality at $\bar{s} = \varepsilon_{mut}^2$ or ρ=-2, called FSI-neutrality, may play a pivotal role in population genetics: it has been demonstrated that FSI-neutrality is the most appropriate null hypothesis for many population genetics tests.

The rate- ln[N_e] inverse relationship and the $d_N/d_S \sim \ln \pi_S$ analysis

The inverse relationship between the evolutionary rate (λ) and the effective population size (N_e) is the landmark prediction of the nearly-neutral theory that claims virtually no adaptive evolution (Ohta 1973). Later, Ohta (1993) tested this inverse relationship by using the rate ratio of nonsynonymous to synonymous substitutions (d_N/d_S) of encoding genes, a proxy of the ratio of substitution rate to the mutation rate (λ/v). Based on a limited number of genes, the author showed about two-fold high of the mean d_N/d_S in the human lineage as that in the mouse lineage. Since the effective population size of humans is much smaller than that of mice, it has been widely cited as the textbook evidence for supporting the nearly-neutral theory (Li 1997; Lynch 2006; Lynch et al. 2011; Kosiol et al. 2008; Ellegren 2009).

Gu (2021) pointed out a scale-problem of such λ-N_e inverse relationship. For instance, Galtier et al. (2016) reported a genome-wide analysis in metazoans. Using synonymous genetic diversity (π_S) as a proxy of N_e, they showed that the genome-wide d_N/d_S ratio is negatively correlated with the log of π_S, rather than π_S predicted by the

nearly-neutral theory. Noticeably, the substitution rate under FSI, as shown by Eq.(6), does predict an asymptotic $\lambda/v \sim \ln \pi_S$ relationship when F>0.5 (or $2N_e\varepsilon_{mut}^2 > 1$). However, a direct application of Eq.(6) on real data analysis is problematic because of the heterogeneity of selection pressures in different nonsynonymous nucleotide sites.

We address the issue of site heterogeneity as follows. In the case of no adaptive evolution, the substitution rate at a nonsynonymous nucleotide site follows Eq.(6) under FSI, where the selection-FSI ratio defined by Eq.(4) is always negative, i.e., $\rho<0$. Suppose that the pattern of FSI measured by the FSI-coefficient ε^2_{mut} is the same across all nonsynonymous nucleotide sites. Therefore, the ρ-variation reflects that the mean selection coefficient ($\bar{s}$) varies among nonsynonymous sites in the range of $\bar{s} < 2\varepsilon_{mut}^2$ such that the substitution rate is always less than the mutation rate. Let λ_g be the substitution rate of a gene. To simplify the notation, one may rewrite $\rho = -z|\rho_g|$, where $|\rho_g|$ is the absolute value of the mean of ρ over nucleotide sites, and z is a positive, random variable that follows a pre-specified distribution with one unit of the mean, denoted by $\phi(z)$.

From Eq.(6), it appears that the mean substitution rate of a gene, denoted by λ_g, is given by

$$\lambda_g = v\int_0^\infty \frac{F(1-|\rho_g|z)}{1-(1-F)^{1-|\rho_g|z}}\phi(z)dz \qquad (8)$$

Next we implement an Ohta-Gu model for $\phi(z)$: a proportion (θ_{inv}) of nucleotide sites are invariable, at which all mutations are lethal, and others are variable, at which mutations follow an exponential distribution. Apparently, the Ohta model is a special case when θ_{inv}=0. As shown by Methods in details, the substitution rate of Eq.(8) can be approximately derived as follows

$$\lambda_g \approx \frac{v(1-\theta_{inv})}{1+|\rho_g|\ln(1+2N_e\varepsilon_{mut}^2)} + v\Omega \qquad (9)$$

The first term on the right hand of Eq.(9) presents the inverse-relationship between λ_g and N_e, that is, the inverse $\lambda_g \sim \ln N_e$ when $2N_e\varepsilon_{mut}^2$>1, or the $\lambda_g \sim N_e$ inverse when

$2N_e\varepsilon_{mut}^2$<1. The second term represents an asymptotic effect of mutations with ρ in the range of $-1 \leq \rho < 0$: $\Omega = 0$ when $2N_e\varepsilon_{mut}^2$=0, but quickly increases toward a saturated, N_e-independent value as $2N_e\varepsilon_{mut}^2$ increases; one may see Eq.(M6) in Methods for a mathematical description. Intuitively speaking, Eq.(9) claims that the substitution rate can be divided two components: one is strongly N_e-dependent and the other is weakly N_e-dependent. That said, FSI theory proposed a new mechanism for molecular evolution in a large population (Gillespie 2000, 2001; Lnfear et al. 2014). Note that Eq.(9) implies that notion of molecular clock (Zuckerkandl and Pauling 1965) may vary locally due to the variation of effective population size but the scale of variation tends to be logarithmic.

We apply Eq.(9) to analyze nonsynonymous substitutions in encoding genes. Let d_N/d_S be the ratio of nonsynonymous distance (d_N) to synonymous distance (d_S) of an encoding gene between two closely-related species, and π_S be the genetic diversity of synonymous mutations in the population. Following the common practice, the d_N/d_S ratio can be used as a proxy of the ratio of the (nonsynonymous) substitution rate to the mutation rate (λ_g/v). Meanwhile, the synonymous diversity π_S is used as a measure of effective population size (N_e) by the formula $\pi_S = 4N_e v$. Note that both arguments are based on the assumption that synonymous mutations are strictly neutral. In the case when FSI is strong, we assume that the weakly N_e-dependent component (Ω) is close enough to its limit $\tilde{\omega} = \lim_{N_e\to\infty} \Omega$ that is N_e-independent. Replacing λ_g/v by d_N/d_S, N_e by $\pi_S/(4v)$, and Ω by $\tilde{\omega}$ in Eq.(9), we obtain

$$\frac{d_N}{d_S} = \frac{1-\theta_{inv}}{1+|\rho_g|\ln(1+\pi_S/\eta)} + \tilde{\omega} \qquad (10)$$

where $\eta = 2v/\varepsilon_{mut}^2$, the (two-fold) ratio of mutation rate to the FSI coefficient. Thus, $\tilde{\omega}$ is the predicted d_N/d_S value when the effective population size is very large.

Computational procedure and case-study

Suppose we have a number of species under study, called focus species. For each focus species, the d_N/d_S ratio is estimated based on a closely-related species, and the synonymous genetic diversity (π_S) is obtained by population genetics screening. The ultimate goal of $d_N/d_S \sim \ln \pi_S$ analysis is to predict $2N_e\varepsilon_{mut}^2$, the FSI-strength of nonsynonymous mutations of each focus species: one may easily verify the following relationship

$$2N_e\varepsilon_{mut}^2 = \frac{4N_e v}{\eta} = \frac{\pi_S}{\eta} \quad (11)$$

Since the synonymous diversity (π_S) is known, the prediction is straightforward after η is estimated by the $d_N/d_S \sim \ln \pi_S$ analysis across multiple species.

Galtier et al. (2016) reported a genome-wide analysis in metazoans, a genome-wide set of protein-coding genes in 44 pairs including 18 vertebrate and 26 invertebrate pairs. Each pair consisted of two closely-related species to estimate the synonymous distances (d_S) and the nonsynonymous distance (d_N), respectively. Moreover, the synonymous genetic diversity (π_S) in the focal species of each specie pair was used as a proxy of the effective population size. Fig.2 represents the plotting of d_N/d_S to $\ln \pi_S$, showing a strong negative correlation between them; the straight line (*p*-value <0.001). This inverse $d_N/d_S \sim \ln \pi_S$ provides a strong evidence for supporting the notion of FSI-driven molecular evolution when $2N_e\varepsilon_{mut}^2$>1. By fitting the data with the Ohta-Gu model (Fig.1), we estimated the mutation-to-fluctuation ratio η=0.00402 to 0.00433 as the proportion of invariable site ranges from θ_{inv}=0 to θ_{inv}=0.5, suggesting that the estimate of η is virtually independent of θ_{inv}.

Table 1 represents the list of predicted $2N_e\varepsilon_{mut}^2$ as well as *F* in 44 metazoans. Surprisingly, we found that 41 species show $2N_e\varepsilon_{mut}^2$>1 or *F*>0.5. Those species with $F \leq 0.5$ are *Camponotus ligniperdus* (Carpet ant), *Reticulitermes flavipes* (Eastern subterranean termite), and *Homo sapiens* (Human). Our finding suggests that, during the long-term metazoan evolution, the FSI-coefficient (ε_{mut}^2) roughly remains a constant, whereas the effective population size varies in magnitudes among species, resulting in up to three-fold changes of the substitution rate of genome evolution. We further conclude

that FSI-genetic drift may play a more importance role than N_e-genetic drift in molecular evolution of most metazoans.

An intriguing question is what is a typical value of ε_{mut}^2. To address First we mention that the estimate of η ≈0.004 indicates that ε_{mut}^2 is about 5×10^2 folds greater than the mutation rate (v). If we assume $v \approx 2 \times 10^{-7}$per generation, we obtain $\varepsilon_{mut}^2 \approx 1.0 \times 10^{-4}$. This result is consistent with the estimate based on the assumption of $N_e = 5000$ in the human population (Table 1), where $\varepsilon_{mut}^2 \approx 0.8 \times 10^{-4}$. In conclusion, we believe that ε_{mut}^2 should be roughly at the magnitude of 10^{-4}.

Discussion

The role of FSI-genetic drift in molecular evolution

The FSI (fluctuating selection among individuals) theory of molecular evolution is based upon the premise that fitness may fluctuate among individuals that carry a mutation. The relative importance of FSI-genetic drift to the conventional N_e-genetic drift can be measured by the relative FSI strength (F): if F is close to 0, the N_e-genetic drift is dominant; whereas the FSI-genetic drift is dominant if F is close to 1. One may tentatively set F=0.5 as an empirical criterion to weigh between those two genetic drifts. As shown by Table 1, our case study showed that the relative FSI-strength F is over 0.5 in most species, suggesting that the FSI-genetic drift, rather than the N_e-drift, plays a major role in metazoan genome evolution. Nevertheless, the classical nearly-neutral theory is appropriate when the effective population size is small. Given ε_{mut}^2=10^{-4}, it appears that F<0.5 when N_e ~ 5,000. So, roughly, the N_e-genetic drift becomes dominant when N_e is less than 2,500, whereas the FSI-genetic drift becomes dominant when N_e is greater than 10,000. Recall that the effective population size of humans is about 5,000, suggesting that two types of genetic drifts should be taken into account in the human evolution.

It should be noticed that background FSI, denoted by FSI_{bg}, the stochastic variation of offspring number commonly for all individuals, has been integrated into the definition of effective population size. Technically, it is convenient to postulate that the mutant is subjecting to FSI, whereas the wildtype has no FSI. Under a more general framework, we propose a general coefficient $\psi = FSI_{mut} - FSI_{bg}$, and so the FSI coefficient is given by $\varepsilon_{mut}^2 = \psi$ if $\psi \geq 0$. Thus, ε_{mut}^2=0 means that FSI of a mutant equals to the background FSI, and $\varepsilon_{mut}^2 > 0$ means that FSI of a mutant is greater than the background FSI. One may call anti-FSI if $\psi < 0$, i.e., the FSI of a mutant is less than the background FSI of wildtype individuals. That is, anti-FSI refers to those mutations that reduce level of FSI, which are, intriguingly, derive a positive selection. We shall further this issue in the future study.

Biological mechanisms of FSI: intrinsic randomness versus local adaptation

From the evolutionary view, the biological mechanisms of FSI can be roughly classified into two categories: the intrinsic randomness versus the local adaptation. In essence, the intrinsic randomness reflects the stochastic nature of genotype-phenotype mapping (Dowell et al. 2010; Lehner 2013; Binder et al. 2015), e.g., regulatory or epigenetic effects triggered by stochastic fluctuations in development stages (Ozbudak et al. 2002; Elowitz et al. 2002; Suel et 2007; Chang et al. 2008). Intrinsic randomness can arise through a variety of mechanisms, including epistasis (genetic interactions) (Mullis et al, 2018; Taylor and Ehrenreich 2014), genetic background (Chandler et al. 2013), or epigenetics (Maamar et al. 2007; Raj et al. 2008). In development, variable expressivity occurs when a mutation shows quantitatively different phenotypic effects among individuals (Vu et al. 2015). Differences in gene expression between affected and unaffected individuals were frequently used to explain the randomness of penetrance (Raj et al. 2010; Hume 2000; Wernet et al 2006).

The second category of biological mechanisms for FSI is due to the change of environmental cues (Vu et al. 2015; Khourt et al. 1988). It usually involves the

sophisticated genetic-environmental interactions underlying the so-called phenotype plasticity (Lai et al. 2024; Wang et al. 2023; Lee et al. 2022; Sommer et al. 2020), which possibly is shaped by the local adaptation (He et al. 2023). Though the detail may have been beyond the scope of current study, we pinpoint out that FSI could be the result of local adaptation, but not necessary.

Ohta-Gu model versus Kimura model

Our current study invokes the Ohta-Gu model: a proportion (θ_{inv}) of nucleotide sites are invariable, at which mutation is lethal, and at the rest of variable sites, the variable ($-\rho$) of mutations follow an exponential distribution. Kimura (1979) argued that the exponential model (Ohta 1973) was insufficient to account for the selection heterogeneity among mutations. Instead, he suggested a gamma distribution with the shape parameter γ=1/2, implying a nontrivial portion of mutations that are highly deleterious. Note that the exponential model, a special case of gamma distribution with γ=1, implies that highly deleterious mutations are statistically negligible. Under Kimura model, however, the patterns of near-neutral mutations and highly deleterious mutations are determined by the shape parameter, which is biologically unrealistic. Our preliminary analysis has demonstrated that the Ohta-Gu model that allows a proportion of invariable sites (i.e., mutations occurred at those sites are highly deleterious or even lethal) is biologically more plausible than the gamma model of Kimura (1979). Moreover, it has been shown that the prediction of FSI-strength is virtually unaffected by the proportion of invariable sites.

Estimation of $2N_e\varepsilon_{mut}^2$ based on life history traits (LHT)

Figuet et al. (2016) studied whether life history traits (LHT) can be taken as proxies of N_e in the molecular evolutionary analysis of amniotes, including 38 mammals, 47 birds and 21reptiles. An interesting problem is whether one can predict $2N_e\varepsilon_{mut}^2$ by the d_N/d_S-LHT analysis. While the detailed analysis will be published elsewhere, here

we present a concise description. First we invoke a simple power relationship between the effective population size (N_e) and the LHT, that is,

$$N_e = N_0 \times LHT^{\alpha} \quad (12)$$

where α is the power-component and N_0 is a constant. After replacing N_e by LHT according to Eq.(12), Eq.(M10) can be rewritten as follows

$$\begin{aligned} \frac{d_N}{d_S} &= \frac{1-\theta_{inv}}{1+|\rho_g|\ln(1+2N_0 LHT^{\alpha}\varepsilon_{mut}^2)} + \tilde{\omega} \\ &\approx \frac{(1-\theta_{inv})/|\rho_g|}{\ln(2N_0\varepsilon_{mut}^2)+\alpha\ln(LHT)} + \tilde{\omega} \\ &= \frac{k'}{b'+\ln(LHT)} + \tilde{\omega} \end{aligned} \quad (13)$$

where two parameters k' and b' are given by

$$\begin{aligned} k' &= \frac{1-\theta_{inv}}{\alpha|\rho_g|} \\ b' &= \frac{\ln(2N_0\varepsilon_{mut}^2)}{\alpha} \end{aligned} \quad (14)$$

By virtually the same procedure described in Methods, one can estimate k' and b' through the $d_N/d_S \sim \ln LHT$ analysis. Therefore, the $2N_e\varepsilon_{mut}^2$ of a species can be predicted by

$$2N_e\varepsilon_{mut}^2 = 2N_0\varepsilon_{mut}^2 \times LHT^{\alpha} = e^{\alpha\hat{b'}}LHT^{\alpha} \quad (15)$$

The remaining question is to determine the power-component α. Figuet et al. (2016) used a subset of species with observed synonymous genetic diversity π_S. Under the assumption that synonymous mutations are strictly neutral, we have $\pi_S=4N_e v$. Applying Eq.(12) again, we have the following linear regression

$$\ln[\pi_S] = \alpha\ln[LHT] + \ln[4N_0 v] \quad (16)$$

to estimate α.

Analogy to Kimura's basic formula of molecular evolution

While we have shown that the classical formula of molecular evolution (Kimura 1962) is a special case of Eq.(6), it is desirable to have a form of formula that is intuitively analogous to the classical formula. To this end, we introduce a new variable Q, the FSI-selection intensity, by $Q = 2N_e\varepsilon_{mut}^2\rho = 4N_e\tau$, as well as the f-scale of FSI by

$$f = \frac{\ln(2N_e\varepsilon_{mut}^2 + 1)}{2N_e\varepsilon_{mut}^2} \qquad (17)$$

Putting together, the substitution rate of Eq.(6) can be written as follows

$$\lambda = v\frac{Q + 2N_e\varepsilon_{mut}^2}{1 - e^{-Qf} + 2N_e\varepsilon_{mut}^2} \qquad (18)$$

Intuitively speaking, in addition to the mutation rate (v), the substitution rate (λ) is determined by the FSI selection intensity (Q), the intensity ($2N_e\varepsilon^2_{mut}$) or the f-scale of FSI. One can easily verify that, when ε^2_{mut}=0, we have f=1 and $Q = 4N_e\bar{s}$ such that Eq.(18) reduced to the formula of Kimura (1962).

Methods

Formulation of the $d_N/d_S \sim \ln \pi_S$ analysis

To simplify the notation, we rewrite $\rho = -z|\rho_g|$, where $|\rho_g|$ is the absolute value of the mean of ρ over (nonsynonymous) sites, and z is a positive, random variable that follows an exponential distribution, with a mean of 1, that is,

$$\phi(z) = e^{-z} \qquad (M1)$$

Under the Ohta-Gu model, the mean substitution rate of a gene is then calculated by

$$\lambda_g = (1-\theta_{inv})\int_0^\infty \lambda\phi(z)dz \qquad \text{(M2)}$$

To obtain a close-form of λ_g, we first notice that the substitution rate given by Eq.(6) has two distinct patterns when $\rho < 0$. As F increases to 0 to 1, λ decreases from v toward 0 in the case of $\rho < -1$, or toward $1 - |\rho|$ in the case of $-1 \leq \rho < 0$. After careful numerical analyses, we found that Eq.(6) can be well approximated by

$$\frac{\lambda}{v} \approx \begin{cases} (1+2N_e\varepsilon)^\rho + [1-(1+2N_e\varepsilon)^\rho](1-|\rho|) & \text{if } -1 \leq \rho < 0 \\ (1+2N_e\varepsilon)^\rho & \text{if } \rho < -1 \end{cases} \qquad \text{(M3)}$$

Replacing $\rho = -z|\rho_g|$ in Eq.(M3) and then plugging into Eq.(M2), we have

$$\begin{aligned} \frac{\lambda_g}{v} &\approx (1-\theta_{inv})\int_0^{1/|\rho_g|} \frac{\phi(z)dz}{(1+2N_e\varepsilon_{mut}^2)^{z|\rho_g|}} \\ &+ (1-\theta_{inv})\int_0^{1/|\rho_g|} \left[1-\frac{1}{(1+2N_e\varepsilon_{mut}^2)^{z|\rho_g|}}\right](1-z|\rho_g|)\phi(z)dz \\ &+ (1-\theta_{inv})\int_{1/|\rho_g|}^\infty \frac{\phi(z)dz}{(1+2N_e\varepsilon_{mut}^2)^{z|\rho_g|}} \end{aligned} \qquad \text{(M4)}$$

After some integral rearrangements, Eq.(M4) turns out to be

$$\begin{aligned} \frac{\lambda_g}{v} &= (1-\theta_{inv})\int_0^\infty \frac{\phi(z)dz}{(1+2N_e\varepsilon_{mut}^2)^{z|\rho_g|}} + \Omega \\ &= \frac{1-\theta_{inv}}{1+|\rho_g|\ln(1+2N_e\varepsilon_{mut}^2)} + \Omega \end{aligned} \qquad \text{(M5)}$$

where Ω is given by

$$\Omega = (1-\theta_{inv})\int_0^{1/|\rho_g|} \left[1-\frac{1}{(1+2N_e\varepsilon_{mut}^2)^{z|\rho_g|}}\right](1-z|\rho_g|)\phi(z)dz \qquad \text{(M6)}$$

Note that Ω is usually small in practice. When $2N_2\varepsilon_{mut}^2 >> 1$, one can show that

$$\Omega \rightarrow \tilde{\omega} = (1-\theta_{inv})\int_0^{1/|\rho_g|}(1-z|\rho_g|)\phi(z)dz \qquad \text{(M7)}$$

Estimation procedure of $2N_e\varepsilon_{mut}^2$

The main goal of $d_N/d_S \sim \ln \pi_S$ analysis is to statistically predict $2N_e\varepsilon_{mut}^2$ for each focus species. Consider a number of species under study, called focus species. For each focus species, the d_N/d_S ratio is estimated based on a closely-related species, and the synonymous genetic diversity (π_S) is obtained by population genetics screening. There are four unknown parameters. By some numerical analysis of Eq.(10), we found that the proportion of invariable sites (θ_{inv}) may affect the estimation of $|\rho_g|$, but that effect on the estimation of η or $\tilde{\omega}$ is minimal. We therefore treat the proportion of invariable sites (θ_{inv}) as a known constant; one may choose a value from 0 to 0.5. We then implement a standard non-linear regression algorithm based on Eq.(10) to estimate the rest of three parameters, denoted by $|\hat{\rho}_g|$ and $\hat{\eta}$, as well as, $\tilde{\omega}$, respectively.

Moreover, we developed an approximate θ_{inv}-independent method to estimate η because it is the key step to predict the FSI-strength. Under the assumption of $|\rho_g| \ln\left(1 + \frac{\pi_S}{\eta}\right) > 1$, Eq.(10) can be approximated by

$$\frac{d_N}{d_S} \approx \frac{(1-\theta_{inv})/|\rho_g|}{\ln(\pi_S) - \ln(\eta)} + \tilde{\omega} = \frac{k}{b + \ln(\pi_S)} + \tilde{\omega} \qquad \text{(M8)}$$

where $k = (1-\theta_{inv})/|\rho_g|$ and $b = -\ln \eta$. All these methods are implemented in the *R*-package.

References

Binder BJ, Landman KA, Newgreen DF, Ross JV (2015) Incomplete penetrance: The role of stochasticity in developmental cell colonization. J Theor Biol

Chandler et al. (2013) Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. Trends Genet

Chang et al (2008) Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature 453:544–547.

de Jong et al. (2024) Moderating the neutralist–selectionist debate: exactly which propositions are we debating, and which arguments are valid? Biological Reviews

Dowell et al. (2010) Genotype to phenotype: a complex problem. Science 328:469.

Elowitz et al. (2002) Stochastic Gene Expression in a Single Cell. Science

Eldar A et al. (2009) Partial penetrance facilitates developmental evolution in bacteria. Nature 460:510–514.

Ellegren H (2009) A selection model of molecular evolution incorporating the effective population size. Evolution 63:301–305.

Figuet et al. (2016). Life history traits, protein evolution, and the nearly neutral theory in amniotes. Molecular Biology and Evolution

Galtier, N. (2016). Adaptive protein evolution in animals and the effective population size hypothesis. PLoS Genetics.

Galtier (2024) Half a Century of Controversy: The Neutralist/Selectionist Debate in Molecular Evolution. GBE

Gillespie JH (2000) The neutral theory in an infinite population. Gene 261:11–18. 32.

Gillespie JH (2001) Is the population size of a species relevant to its evolution? Evolution 55:2161–2169.

Gu, X (2021) Random penetrance of mutations among individuals: a new type of genetic drift in molecular evolution. Phenomics

Gu, X (2024) Models of fluctuating selection between generations: a solution for the theoretical inconsistency. Journal of Molecular Evolution.

Gu, X(2025) Population genetics and molecular evolution of fluctuating selection among individuals (FSI). BioRxiv.

Hahn MW (2008) Toward a selection theory of molecular evolution. Evolution 62:255–265.

He et al (Bin Su) (2023) Polygenic adaptation leads to a higher reproductive fitness of native Tibetan at high altitude. Current Biology

Hughes, A. L. (2008). Near-neutrality: the leading edge of the neutral theory of molecular evolution. Annals of the New York Academy of Sciences

Hume DA (2000) Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression. Blood 96:2323–2328

Jensen et al. (2019) The importance of the Neutral Theory in 1968 and 50 years on: A response to Kern and Hahn 2018. Evolution 73:111–114.

Jensen et al. (2025) Genetic modifiers and ascertainment drive variable expressivity of complex disorders. Cell

Karlin and Levikson. 1974. Temporal fluctuations in selection intensities: Case of small population size. Theoret. Popul. Biol. 6: 383-342

Karlin, S, and U. Lieberman, 1974 Random temporal variation in selection intensities: case of large population size. Theoret. Popul. Biol. 6: 355-382

Kern AD, Hahn MW (2018) The Neutral Theory in Light of Natural Selection. Mol Biol Evol 35:1366–1371.

Kimura M (1962) On the probability of fixation of mutant genes in a population. Genetics 47:713–719

Kimura M (1968) Evolutionary rate at the molecular level. Nature 217:624–626.

Kimura M (1979) Model of effectively neutral mutations in which selective constraint is incorporated. PNAS

Kimura M (1983) The Neutral Theory and Molecular Evolution. Cambridge University Press, New York

Khoury MJ, Flanders WD, Beaty TH (1988) Penetrance in the presence of genetic susceptibility to environmental factors.

Kosiol et al. (2008) Patterns of positive selection in six Mammalian genomes. PLoS Genet 4:e1000144.

Lai et al. (2024) Evolution of Phenotypic Variance Provides Insights into the Genetic Basis of Adaptation. GBE

Lanfear R et al (2014) Population size and the rate of evolution. Trends Ecol Evol 29:33–41.

Lee et al. (2022) Evolution and maintenance of phenotypic plasticity. Biosystems.

Lehner B (2013) Genotype to phenotype: lessons from model organisms for human genetics. Nat Rev Genet 14:168–178.

Li WH (1997) Molecular evolution. Sinauer Associates Incorporated, Sunderland, USA

Lynch M The Origins of Genome Architecture. Sinauer Associates, Inc, Sunderland, MA

Lynch M et al. (2011) The repatterning of eukaryotic genomes by random genetic drift. Annu Rev Genomics Hum Genet 12:347–366.

Maamar H, Raj A, Dubnau D (2007) Noise in gene expression determines cell fate in Bacillus subtilis. Science 317

McCandlish and Stoltzfus (2014) Modeling Evolution Using the Probability of Fixation: History and Implications. The Quarterly Review of Biology

Mullis et al (2018) The complex underpinnings of genetic background effects. Nat Commun 9:3548.

Munoz-Gomez et al (2021). Constructive neutral evolution 20 years later. JME

Nadeau JH (2001) Modifier genes in mice and humans. Nat Rev Genet 2:165–174.

Nei, M., Suzuki, Y. & Nozawa, M. (2010). The neutral theory of molecular evolution in the genomic era. Annual Review of Genomics and Human Genetics

Ohta T (1973) Slightly deleterious mutant substitutions in evolution. Nature 246:96–98.

Ohta T (1993) An examination of the generation-time effect on molecular evolution. Proc Natl Acad Sci U S A 90:10676–10680.

Ozbudak et al. (2002) Regulation of noise in the expression of a single gene. Nat Genet

Raj A et al. (2010) Variability in gene expression underlies incomplete penetrance. Nature 463:913–918.

Raj A, van Oudenaarden A (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. Cell

Riordan JD, Nadeau JH (2017) From Peas to Disease: Modifier Genes, Network Resilience, and the Genetics of Health. Am J Hum Genet 101:177–191.

Sommer (2020) Phenotypic Plasticity: From Theory and Genetics to Current and Future Challenges. Genetics

Süel (2007) Tunability and noise dependence in differentiation dynamics. Science

Taylor MB, Ehrenreich IM (2014) Genetic interactions involving five or more genes contribute to a complex trait in yeast. PLoS Genet

Vu V et al (2015) Natural Variation in Gene Expression Modulates the Severity of Mutant Phenotypes. Cell

Wagner, A. (2008). Neutralism and selectionism: a network-based reconciliation. Nature Reviews Genetics

Wang et al. (2023) Reproductive variance can drive behavioral dynamics. PNAS

Wernet et al. (2006) Stochastic spineless expression creates the retinal mosaic for colour vision. Nature 440:174–180.

Zuckerkandl E, Pauling L (1965) Evolutionary Divergence and Convergence in Proteins. In: Bryson V, Vogel HJ (eds) Evolving Genes and Proteins. Academic Press, pp 97–166

Table 1. List of predicted FSI strength $2N_e\varepsilon_{mut}^2$ and the relative FSI strength (*F*) for each species.

Focal_species	***Outgroup***	***Taxonomy***	$2N_e\varepsilon_{mut}^2$	***F***
Allolobophora_chlorotica_L2	*Aporrectodea_icterica*	Annelida	36.29	0.973
Armadillidium_vulgare	*Armadillidium_nasatum*	Arthropoda	13.96	0.933
Artemia_franciscana	*Artemia_sinica*	Arthropoda	9.18	0.902
Camponotus_ligniperdus	*Camponotus_aethiops*	Arthropoda	0.80	*0.444*
Culex_pipiens	*Culex_torrentium*	Arthropoda	34.14	0.972
Halictus_scabiosae	*Halictus_simplex*	Arthropoda	1.67	0.626
Melitaea_cinxia	*Melitaea_didyma*	Arthropoda	34.67	0.972
Messor_barbarus	*Messor_structor*	Arthropoda	4.64	0.823
Necora_puber	*Carcinus_aestuarii*	Arthropoda	4.22	0.808
Reticulitermes_grassei	*Reticulitermes_flavipes*	Arthropoda	1.00	*0.500*
Thymelicus_lineola	*Thymelicus_sylvestris*	Arthropoda	17.42	0.946
Thymelicus_sylvestris	*Thymelicus_lineola*	Arthropoda	14.70	0.936
Eunicella_cavolinii	*Eunicella_verrucosa*	Cnidaria	9.49	0.905
Abatus_cordatus	*Abatus_agassizi*	Echinodermata	2.81	0.738
Echinocardium_cordatum_B2	*Echinocardium_mediterraneum*	Echinodermata	33.84	0.971
Echinocardium_mediterraneum	*Echinocardium_cordatum_B2*	Echinodermata	10.88	0.916
Ophioderma_longicauda_L1	*Ophioderma_longicauda_L3*	Echinodermata	9.96	0.909
Chlorocebus_aethiops	*Macaca_mulatta*	Mammalia	1.83	0.647
Eulemur_coronatus	*Eulemur_mongoz*	Mammalia	1.63	0.620
Eulemur_mongoz	*Eulemur_coronatus*	Mammalia	1.43	0.588
Galago_senegalensis	*Nycticebus_coucang*	Mammalia	1.78	0.640
Homo_sapiens	*Pan_troglodytes*	Mammalia	0.85	*0.461*
Lepus_granatensis	*Lepus_americanus*	Mammalia	3.35	0.770
Macaca_mulatta	*Chlorocebus_aethiops*	Mammalia	1.83	0.647
Microtus_arvalis	*Microtus_glareolus*	Mammalia	7.25	0.879
Nycticebus_coucang	*Galago_senegalensis*	Mammalia	1.21	0.548
Pan_troglodytes	*Homo_sapiens*	Mammalia	3.07	0.754
Propithecus_coquereli	*Varecia_variegata_variegata*	Mammalia	4.59	0.821
Varecia_variegata_variegata	*Propithecus_coquereli*	Mammalia	2.89	0.743
Crepidula_fornicata	*Crepidula_plana*	Mollusca	20.93	0.954
Mytilus_galloprovincialis	*Mytilus_californianus*	Mollusca	21.76	0.956
Ostrea_edulis	*Ostrea_chilensis*	Mollusca	4.38	0.814
Physa_acuta	*Physa_gyrina*	Mollusca	21.89	0.956
Sepia_officinalis	*Sepiella_japonica*	Mollusca	1.63	0.620

Caenorhabditis_brenneri	*Caenorhabditis_sp.10*	Nematoda	24.26	0.960
Lineus_longissimus	*Lineus_ruber*	Nemertea	10.26	0.911
Aptenodytes_patagonicus	*Aptenodytes_forsteri*	Sauropsida	3.04	0.753
Chelonoidis_nigra	*Chelonoidis_carbonaria*	Sauropsida	1.14	0.532
Emys_orbicularis	*Trachemys_scripta*	Sauropsida	3.34	0.770
Eudyptes_moseleyi	*Pygoscelis_papua*	Sauropsida	1.65	0.623
Parus_caeruleus	*Parus_major*	Sauropsida	4.92	0.831
Hippocampus_guttulatus	*Hippocampus_kuda*	Teleostei	1.14	0.532
Ciona_intestinalis_A	*Ciona_intestinalis_B*	Urochordata	8.62	0.896
Ciona_intestinalis_B	*Ciona_intestinalis_A*	Urochordata	38.90	0.975

Fig.1. The impact of FSI on the substitution rate in the case of no adaptive evolution ($\rho \leq 0$). Selection-duality is defined by the region between the solid line of fixation neutrality (ρ=0) and that of generic neutrality (ρ=-4). The Ω-region is defined by that between the solid line of fixation neutrality (ρ=0) and that of ρ=-1), in which the substitution rate is nonzero even in a very large population.

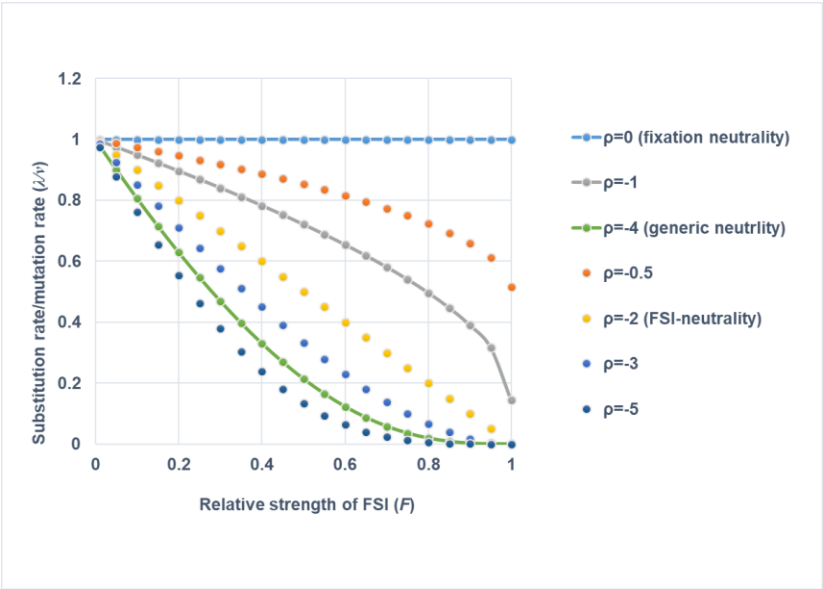
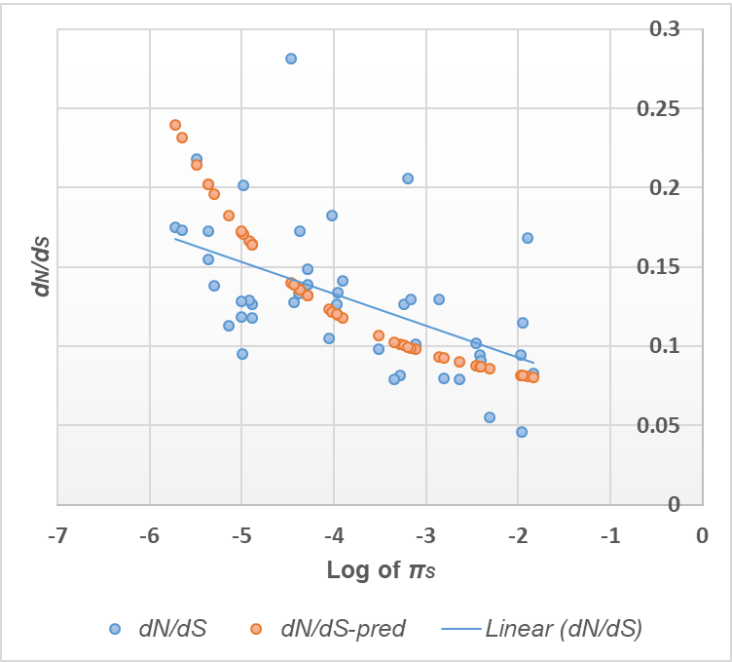
Fig.2. The potting of d_N/d_S against the log of synonymous genetic diversity (π_S). Data were from Galtier et al. (2016) including 18 vertebrate and 26 invertebrate pairs. See Table 1 for the name list of all species. Blue circles represent the observed d_N/d_S and π_S of each species, whereas the orange circles represent the predicted d_N/d_S, based on the estimated model parameters and the observed π_S.