Evolutionary Framework for Protein Sequence Evolution and Gene Pleiotropy

Xun Gu¹

Department of Genetics, Development and Cell Biology, Center for Bioinformatics and Biological Statistics, Iowa State University, Ames, IA 50011

Manuscript received October 5, 2006
Accepted for publication January 8, 2007

ABSTRACT

In this article, we develop an evolutionary model for protein sequence evolution. Gene pleiotropy is characterized by K distinct but correlated components (molecular phenotypes) that affect the organismal fitness. These K molecular phenotypes are under stabilizing selection with microadaptation (SM) due to random optima shifts, the SM model. Random coding mutations generate a correlated distribution of K molecular phenotypes. Under this SM model, we further develop a statistical method to estimate the “effective” number of molecular phenotypes (K_e) of the gene. Therefore, for the first time we can empirically evaluate gene pleiotropy from the protein sequence analysis. Case studies of vertebrate proteins indicate that K_e is typically ~6–9. We demonstrate that the newly developed SM model of protein evolution may provide a basis for exploring genomic evolution and correlations.

Although the wide availability of high-throughput data has greatly facilitated the study of evolutionary genomics, the foundation of molecular evolution has been relatively “static” (Kimura 1983). Basically, the evolutionary fate of a mutant is largely determined by the prespecified fitness value (the coefficient of selection) and the effective population size. Although this classical treatment has successfully removed the complexity of genotype-phenotype association, problems may appear when the association itself has been one of the research highlights. The recent debate about genomic correlations among evolutionary rate, interaction, expression, and dispensability has indicated the limitation of current evolutionary models (reviewed in Pal et al. 2006). Alternatively, researchers have to use the partial-correlation or principle-component methods to analyze major determinants governing protein evolution (e.g., Drummond et al. 2005; Wall et al. 2005; Wolf et al. 2006), because these approaches are not model dependent.

It is desirable to develop a novel framework that can provide an integrated view of protein evolution, but an unsolved key problem is how to deal with the sophisticated relationship between protein sequence and organismal fitness. We recognized that the substantial genotype-phenotype literature can be used to address this issue (e.g., Fisher 1930; Lande 1980; G. Wagner 1989; Waxman and Peck 1998; A. Wagner 2000; Poon and Otto 2000; Zhang and Hill 2003). As mutations are raw materials for evolution, the distribution of effects of mutations, f(s), has been a central issue in the genotype-phenotype study (Lynch et al. 1999; Bataillon 2000; Shaw et al. 2002). Under Fisher’s geometric model (e.g., Fisher 1930; Hartl and Taubes 1998; Poon and Otto 2000) or multivariate model (e.g., Lande 1980; Welch and Waxman 2003; Martin and Lenormand 2006), f(s) can be predicted on the basis of selective and mutational assumptions. The dimensionality in each framework has been interpreted as the number of loci affecting a quantitative trait (reviewed in Kingsolver et al. 2001), or the number of affected phenotypic traits of a mutation, estimated from mutational-accumulation experiments (reviewed in Elena and Lenski 2003). On the other hand, an ad hoc gamma distribution for f(s) was introduced in the early days of the neutral theory of molecular evolution (Ohta 1973, 1977; Kimura 1983), which has been found useful to estimate the proportion of neutral or deleterious mutations (e.g., Nielsen and Yang 2003; Piganeau and Eyre-Walker 2003; Eyre-Walker et al. 2006).

Here we extend our recent work (Gu 2006). The principle behind our present work in modeling protein sequence evolution is as follows: First, multifunctionality of a gene (pleiotropy) is characterized by K distinct components in the fitness (called molecular phenotypes). Second, these K molecular phenotypes are under stabilizing selection (Fisher 1930; Waxman and Peck 1998) with microadaptations due to random shifts of the fitness optimum (Gillespie 1991). Third, random mutations in the coding region of the gene generate a mutational distribution for K molecular phenotypes. Correlated structures among K molecular phenotypes are taken into account at both selectional and mutational levels. More importantly, we develop a statistical
method to estimate the “effective” number of molecular phenotypes ($K_e$) of the gene. Case studies of vertebrate proteins demonstrate that for the first time we can empirically evaluate gene pleiotropy only from phylogenetic sequence analysis.

THE MODEL

Molecular phenotypes and gene pleiotropy: Pleiotropy of a gene (protein) function can be modeled by multiple ($K$) molecular phenotypes, denoted by ($y_1, \ldots, y_K$). Each molecular phenotype $y_i$ represents a nontrivial component of genetic variation, contrasting with organismal fitness as a result of a specific (yet unknown) biological process (Figure 1). At one extreme, molecular phenotypes could correspond to subcomponents of protein function, regardless of the biological processes. At another extreme, molecular phenotypes could be determined mainly by distinct physiological processes. Since these underlying biological processes are usually intractable, the concept of molecular phenotypes based on fitness effects avoids this difficulty.

Following the widely used procedure (e.g., Lande 1980; Turelli 1985; Waxman and Peck 1998), we assume that the fitness function of molecular phenotypes $y = (y_1, \ldots, y_K)'$ is Gaussian-like; i.e.,

$$w(y) = \exp\left[ -\frac{(y - \mu)'\Sigma^{-1}_w(y - \mu)}{2} \right],$$

where $\mu$ is the optimum of the fitness function, and $\Sigma_w$ is a (positive definite) symmetric matrix characterizing correlated stabilizing selection on $K$ molecular phenotypes. The $i$th diagonal element $\sigma^2_{y_i}$ measures the strength of stabilizing selection on the $i$th molecular phenotype, while the $ij$th nondiagonal element $\sigma_{y_i y_j}$ measures the correlated stabilizing selection on $y_i$ and $y_j$.

Let $y_0$ be the population mean of molecular phenotypes of a gene, which may not necessarily be at the optimum ($\mu$). By definition (Fisher 1930), the coefficient of selection for the molecular phenotype $y$ is given by

$$p(y \mid y_0; \mu) = w(y)/w(y_0) - 1.$$  \hspace{1cm} (2)

Classical stabilizing selection (Turelli 1985; Waxman and Peck 1998) assumes that the population mean is always at the fitness optimum, i.e., $y_0 = \mu$, leading to purifying selection; that is, $p(y \mid y_0 = \mu) \leq 0$ always holds.

Next, we consider the effects of mutations on molecular phenotypes. A random mutation in the coding region of a gene may affect molecular phenotypes in a correlated fashion (Figure 1). Such a pleiotropic mutational effect can be described by a multivariate normal distribution

$$p(y) = N(y; y_m; \Sigma_m),$$

where $y_m$ is the mean mutational effect on molecular phenotypes, and the covariance matrix $\Sigma_m$ measures the correlated mutational effects. The coefficient of selection $p(y \mid y_0, \mu)$ and the distribution of mutational effects $p(y)$ provide a connection between organismal fitness and protein sequence evolution through molecular phenotypes $y$ (Figure 1). Below we discuss the classical stabilizing selection model and the newly developed stabilizing selection with microadaptation (SM) model. See Figure 2 for illustration.

Stabilizing selection model: Under the classical stabilizing model, the population mean of molecular phenotypes ($y_0$) is always fixed at the optimum ($\mu$). Without loss of generality, we assume that $y_0 = \mu = 0$. From Equation 2 one can show that $p_0(y) = p(y \mid y_0 = \mu = 0) \approx -\frac{1}{2}y'\Sigma_w^{-1}y/2 < 0$, reflecting the stabilizing (purifying) selection against deleterious mutations, which necessarily results in a deviation from the optimum. A consequence of this model is that sequence evolution is dominated by the fixation of very slightly deleterious mutations (Ohta 1973; Kimura 1979, 1983).

For mathematical convenience, we assume that the mutational distribution, $p(y)$ in Equation 3, is centered at the population mean, namely, $y_m = y_0 = 0$. In the following we consider mainly the selection intensity $S(y) = 4N_c p_0(y) = -2N_c(y'\Sigma_w^{-1}y)$, where $N_c$ is the effective population size. Given Equation 3 of $p(y)$, the distribution of $S$, $f(S)$, can be uniquely determined by the probability theory (see the appendix). Although the analytical form of $f(S)$ is available only for a few special cases, the moments of $S$ can be obtained readily. For instance, the mean of $S$ can be computed by

$$\bar{S} = -2N_c \int y'\Sigma_w^{-1}yp(y)dy,$$

which is always negative.

Stabilizing selection with microadaptation model: Consider the case when the fitness optimum ($\mu$) of a
gene is no longer fixed during the evolution. Rather, $\mu$ can be shifted either by environmental changes or by internal physiological perturbations. Each shift of $\mu$ results in a microadaptation toward a new fitness optimum (Gillespie 1991). Since the direction and strength of the $\mu$-shift are unobservable, we treat $\mu$ as a ($K$-dimensional) random variable. Similar to the above, we assume that the population mean and the mutation mean satisfy $y_0 = y_m = 0$. Then, under the SM model, the coefficient of correlation defined by Equation 2 can be specified as follows:

$$
\rho(y | \mu) = \exp[\gamma y^{\mu} - y^{\mu} y/2] - 1.
$$

(5)

After assuming that $\mu$ follows a multivariate normal distribution $\phi(\mu)$ with mean $0$ and covariance matrix $\Sigma_{\mu}$, one can integrate out the unobservable $\mu$ by $\rho(y) = \int \rho(y | \mu) \phi(\mu) d\mu$. Therefore, one can show that $\rho(y)$, the coefficient of selection of $y$ averaged over $\mu$-shifts during the evolution, is given by

$$
\rho(y) = \exp[\gamma y^{\mu} - y^{\mu} y/2] - 1 \\
\approx -y^{\mu} U^{-1} y/2,
$$

(6)

where the matrix $U = \Sigma_{w}^{-1} - \Sigma_{w}^{-1} \Sigma_{m} \Sigma_{w}^{-1}$ characterizes the correlated nature of molecular phenotypes in fitness after stabilizing selection and microadaptation. It follows that the selection intensity defined by $S(y) = 4N_e \rho(y)$ is given by

$$
S(y) = -2N_e y^{\mu} U^{-1} y.
$$

(7)

Apparently, the stabilizing selection model is a special case when $\Sigma_{\mu} = 0$ so that $U = \Sigma_{w}$. Moreover, unlike the stabilizing selection model that $S(y)$ is always negative, the selection intensity $S(y)$ under the SM model can be positive when the $\mu$-shift becomes huge. See ANALYTICAL RESULTS for more details.

**Evolutionary rate of protein sequence**: Suppose that a protein carrying one (nonsynonymous) mutation has the molecular phenotype vector $y$. The selection intensity for this mutation, $S(y)$, is given by Equation 7. Hence, the well-known formula for the evolutionary rate (Kimura 1983) depends on the molecular phenotypes $y$; that is,

$$
\lambda(y) = v \frac{S(y)}{1 - e^{-S(y)}}.
$$

(8)

where $v$ is the mutation rate. At the gene level, molecular phenotypes $y$ are random variables, resulting from random mutations at nucleotide sites (or amino acid residues). Therefore, the evolutionary rate is a random variable that varies among sites. Under the SM model, the mean evolutionary rate of a protein sequence is given by

$$
\bar{\lambda} = v \int \frac{S(y)}{1 - e^{-S(y)}} p(y) dy,
$$

(9)

given the distribution of mutational effects on $y$, $p(y)$.

**ANALYTICAL RESULTS**

Here we present analytical results under the SM model, which are useful to study the underlying mechanisms governing protein sequence evolution and to develop a statistical method for estimating important parameters such as $K$, the number of molecular phenotypes of a gene.

**Single molecular phenotype ($K = 1$)**: When $K = 1$, the fitness function in Equation 1 can be reduced as $w(y) = \exp(-y^{2} / 2\sigma_{w}^{2})$ and the selection coefficient in Equation 6 as $\rho(y) = \exp(y^{2} / 2\sigma_{w}^{2}) - 1$, where $\sigma_{w}^{2}$ measures the strength of single stabilizing selection. Similarly, the shift of $\mu$ follows a single normal distribution $\phi(\mu)$ with mean $0$ and $\sigma_{\mu}^{2}$. Together, Equation 8 can be simplified as

$$
S(y) = -\frac{2N_e}{\sigma_{w}^{2}} \left(1 - \frac{\sigma_{\mu}^{2}}{\sigma_{w}^{2}}\right) y^{2} = \frac{h}{2} y^{2}.
$$

(10)

Given the distribution of mutational effects $p(y) = N(y; 0, \sigma_{w}^{2})$, one can show the mean of selection intensity, $\bar{S} = \int S(y) p(y) dy$, as follows:

$$
\bar{S} = -2N_e \frac{\sigma_{m}^{2}}{\sigma_{w}^{2}} \left(1 - \frac{\sigma_{\mu}^{2}}{\sigma_{w}^{2}}\right).
$$

(11)

Several interesting results need to be mentioned. First, $\bar{S}$ is determined by the mutational variance ($\sigma_{m}^{2}$) and the
variance of microadaptation ($\sigma_w^2$) relative to the coefficient of stabilizing selection ($\sigma_m^2$). Hence, the scale of molecular phenotype is not important. Second, $\bar{S} < 0$ if $\sigma_m^2 < \sigma_w^2$, and vice versa. In particular, the microadaptation model is reduced to the stabilizing selection model when $\sigma_m^2 = 0$ (Gu 2007). Third, microadaptation reduces the sequence conservation by decreasing the absolute values of $\bar{S}$, although $\bar{S}$ may remain negative. In other words, microadaptation may increase the evolutionary rate of protein sequence. And fourth, from Equation 15, one can show that $\bar{S}$ follows a negative gamma distribution with the shape parameter $\frac{1}{2}$. For instance, when $\sigma_m^2 < \sigma_w^2$, $\bar{S}$ follows a negative gamma distribution

$$f(S) = \frac{e^{-S/\bar{S}}\bar{S}^{1/2}}{\Gamma(1/2)} S^{-1/2} e^{\bar{S} S}.$$  \hfill (12)

**$K$ molecular phenotypes:** For $K \geq 2$, the distribution of selection intensity $\bar{S}$, $f(\bar{S})$, is generally not analytically tractable. Nevertheless, one can derive the analytical forms of moments on the basis of the theory of multiple normal distribution. These results are crucial for estimating $K$ from the sequence data.

**Mean and variance of $\bar{S}$—general results:** Denote matrix $A = U^{-1} \Sigma_m = Z - \Omega$, where $Z = \Sigma_w^{-1} \Sigma_m$ and $\Omega = \Sigma_w^{-1} \Sigma_m Z$. Briefly, under stabilizing selection, matrix $Z$ describes correlated microadaptations, while $\Omega$ describes correlated micromutations. Together, matrix $A$ characterizes the net effects of correlated mutations on fitness under the SM model. In the appendix (also see Martin and Lenormand 2006 for phenotypic data), we show that the joint effects of all selective, microadaptive, and mutational covariances can be reduced to $K$ eigenvalues of matrix $A, \alpha_1, \ldots, \alpha_K$. Consequently, the mean and variance of $\bar{S}$ are given by

$$\bar{S} = -2N_e \sum_{i=1}^{K} \alpha_i$$  \hfill (13)

and

$$\text{Var}(\bar{S}) = 8N_e^2 \sum_{i=1}^{K} \alpha_i^2,$$  \hfill (14)

respectively. In short, each $\alpha_i$ corresponds to an independent molecular phenotypic direction, on which mutation, microadaptation, and stabilizing selection act with an average effect $-2N_e \alpha_i$ on the mean selection intensity ($\bar{S}$) of the gene.

**Canonical representation of $\bar{S}$:** Although Equation 13 is mathematically elegant, we wish to find an equivalent representation of $\bar{S}$ that is biologically more interpretable. Due to the arbitrary nature of the original $K$ molecular phenotypes, without loss of generality one may choose a convenient coordinate system. Here we adopt a canonical form of $K$ molecular phenotypes, to define independent phenotypes experiencing stabilizing selection. This leads to a diagonal matrix $\Sigma_w$; each diagonal element $\sigma_{w,i}^2$ measures the independent stabilizing selection on the $i$th (canonical) molecular phenotype. Let $\sigma_{m,i}^2$ be the $i$th diagonal element of matrix $\Sigma_m$ or the mutational variance for the $i$th (canonical) molecular phenotype. We have shown that the canonical presentation for the mean of selection intensity is given by

$$\bar{S} = -2N_e \sum_{i=1}^{K} \sigma_{m,i}^2 \left(1 - \gamma_i\right),$$  \hfill (15)

where the parameter $\gamma_i$ measures the net effect of microadaptation on the $i$th (canonical) molecular phenotype. In particular, when the mutational effect or microadaptation is independent among the canonical molecular phenotypes, we have $\gamma_i = \sigma_{m,i}^2 / \sigma_{w,i}^2$. It should be noted that the pleiotropic nature of mutations affects the variance of $\bar{S}$, $\text{Var}(\bar{S})$ (see the appendix), while $\bar{S}$ is not affected, as shown in Equation 15.

**Model classification:** As the stabilizing selection acts against nucleotide substitutions (negative selection), microadaptation may provide an opposite force (positive selection) to increase the evolutionary rate. The mean selection intensity given by Equations 13 and 15 can be used to classify the pattern of sequence conservation under the SM model:

1. Strong stabilizing selection with weak microadaptation ($\text{SM}_m$): In this case, the magnitude of $\mu$-shift of molecular phenotypes during the evolution is small, relative to the strength of stabilizing selection. It indicates a dominant purifying selection in the protein sequence evolution. The $\text{SM}_m$ model can be defined as having matrix $A$ positive definite, such that each eigenvalue $\alpha_i > 0$ and $\bar{S} < 0$. In the canonical form of Equation 15, it means that $\gamma_i < 1$ for any $i = 1, \ldots, K$. The pure stabilizing selection is a special case of $\gamma_i = 0$ (no microadaptation). Moreover, $\bar{S} < 0$ implies that the ratio of nonsynonymous to synonymous substitutions ($d_K / d_s$) is $< 1$. Therefore, the $\text{SM}_m$ model may describe the general pattern in protein sequence evolution, evidenced by genome-wide analyses that have shown $d_K / d_s < 1$ for most genes.

2. Episodic microadaptation under strong stabilizing selection ($\text{SM}_E$): Some genes may have experienced episodic adaptive processes in a few molecular phenotypes, resulting in negative eigenvalues of matrix $A$ ($i.e.$, $\alpha_i < 0$ for some $i$’s), but the mean selection intensity remains negative, $\bar{S} < 0$. That is, matrix $A$ is no longer positive definite but a positive trace ($\sum_{i=1}^{K} \alpha_i > 0$) remains. In the canonical form, some of $\gamma_i$ are $> 1$, although $\bar{S} < 0$. Apparently, the $d_K / d_s$ ratio under the $\text{SM}_E$ model may be increased by episodic microadaptations, but $d_K / d_s < 1$ remains. A special case is the neutral-like behavior of protein sequence evolution when $\sum_{i=1}^{K} \alpha_i \approx 0$, resulting in
$\bar{S} \approx 0$. That is, the observed neutral-like evolution may reflect the zero net effect on fitness, canceling each other between episodic microadaptation and stabilizing selection.

3. Strong microadaptation under stabilizing selection (SM$_s$): In a few genes, positive selection and adaptation may dominate the evolution in many molecular phenotypes. In these cases, $\sum_{i=1}^{K} \alpha_i > 0$ results in $\bar{S} > 0$ and the ratio $d_k/d_1 > 1$. An extreme case occurs if $\alpha_i < 0$ or $\gamma_i > 1$ (the canonical form) for all $i = 1, \ldots, K$. In this case, matrix $A$ is negative definite, indicating the overwhelming adaptation forces for virtually all of the substitutions.

The $W$-model for $S$-gamma distribution: Under some strong assumptions, Gu (2007) showed that $\bar{S}$ is distributed as a (negative) gamma. A more general condition for $S$-gamma distribution exists when stabilizing selection, mutations, and microadaptation have very similar correlation structures among molecular phenotypes. Indeed, when the three corresponding matrices can be written as $\Sigma_w = \alpha_w^2 W$, $\Sigma_m = \alpha_m^2 W$, and $\Sigma_\mu = \alpha_\mu^2 W$, respectively, where $W$ is a (positive-definite) symmetric matrix, the eigenvalues of matrix $A$ become identical; i.e., $\alpha_1, \ldots, \alpha_K = \alpha$, as given by

$$\alpha = \frac{\sigma_m^2}{\sigma_w^2} \left( 1 - \frac{\sigma_\mu^2}{\sigma_w^2} \right).$$

(16)

Then, one can easily show that when $\alpha > 0$ (the SM$_w$ model), $f(S)$ follows a negative gamma distribution

$$f(S) = \frac{h^{K/2}}{\Gamma(K/2)} (-S)^{K/2-1} e^{-hS},$$

(17)

where the parameter $h = 1/(4N_\epsilon \alpha)$. Although it is rare, when $\alpha < 0$ (the SM$_m$ model), $f(S)$ follows a (positive) gamma distribution. In both situations, the gamma shape parameter of $S$ is the (half) number of molecular phenotypes ($K/2$) (Gu 2007).

ESTIMATION AND EXAMPLES

The newly developed SM model is useful in practice only if the underlying parameters can be estimated. But this is difficult to accomplish using conventional statistical approaches, since the number of unknown parameters may be huge. Instead, we attempt to focus on two important parameters, $K$, critical for understanding gene pleiotropy, and $\bar{S}$, the measure for overall sequence conservation. Our purpose here is to estimate $K$ and $\bar{S}$, without specifying other more detailed parameters.

Moments of evolutionary rate: Under the SM model, the mean evolutionary rate of a protein sequence ($\lambda$) is given by Equation 9. Therefore, any $k$th moment of the evolutionary rate is given by

$$\lambda_k \approx \frac{1}{u_k} \int_0^\infty e^{-kS(y)} (1 + cS(y))^k p(y) dy$$

(18)

To apply Equation 18 to estimate unknown parameters, we use the approximation $S/(1 - e^{-S}) \approx e^{-S}(1 + cS)$, where $c \approx 0.5772$ is Euler’s constant. As shown by Figure 3, this approximation is satisfactory for $S \leq 0$. Thus, under the SM$_w$ model (stabilizing selection with weak microadaptation), which assures $S(y) = -2N_\epsilon y^T U^{-1} y < 0$, the $k$th moment of the evolutionary rate $\lambda$ can be approximated as follows:

$$\lambda_k \approx \frac{1}{u_k} \int_0^\infty e^{-kS(y)} (1 + cS(y))^k p(y) dy$$

(19)

As shown in the Appendix, the mean evolutionary rate ($k = 1$) is given by

$$\lambda \approx \frac{1}{u} = \prod_{i=1}^{K} \left[ 1 + 4N_\epsilon \alpha_i \right]^{-1/2} \left( 1 + c \sum_{i=1}^{K} \frac{2N_\epsilon \alpha_i}{1 + 4N_\epsilon \alpha_i} \right)$$

(20)

and the second moment of $\lambda$ ($k = 2$) by...
\[
\bar{\lambda}^2 / \nu^2 = \prod_{i=1}^{K} [1 + 8N_c \alpha_i]^{-1/2} \\
\times \left(1 + 2c \sum_{i=1}^{K} \frac{2N_c \alpha_i}{1 + 8N_c \alpha_i} + c^2 \sum_{i=1}^{K} \frac{8N_c^2 \alpha_i^2}{(1 + 8N_c \alpha_i)^2} \right) \\
+ c^2 \left[\sum_{i=1}^{K} \frac{2N_c \alpha_i}{1 + 8N_c \alpha_i}\right]^2.
\] (21)

**Effective number (\(K_e\)) of molecular phenotypes:** In general, the mean evolutionary rate depends on \(K\) distinct eigenvalue parameters \((N_c \alpha_i)\), while the second moment depends on \(N_c \alpha_i\)'s and \(N_c \alpha_i^2\), and so forth. To develop a simple method for estimating \(K\) on the basis of Equations 20 and 21, we invoke the assumption of \(N_c \alpha_i \gg 1\), leading to

\[
\bar{\lambda} / \nu \approx \prod_{i=1}^{K} [4N_c \alpha_i]^{-1/2}(1 + cK/2)
\] (22)

and

\[
\bar{\lambda}^2 / \nu^2 \approx \prod_{i=1}^{K} [8N_c \alpha_i]^{-1/2}(1 + cK/2 + c^2 K/8 + c^2 K^2/16),
\] (23)

respectively. Notably, the first and second moments of evolutionary rate depend only on two parameters, \(K\) and the product \(\prod_{i=1}^{K} 4N_c \alpha_i\). Further, we found that the ratio of the second moment \(\bar{\lambda}^2 / \nu^2\) to the first moment \((\bar{\lambda} / \nu)\), denoted by \(g_K = [\bar{\lambda}^2 / \nu^2] / [\bar{\lambda} / \nu]\), is given by

\[
g_K = 2^{-K/2} \left[\frac{1 + (c \bar{\lambda} + 2K + 2)}{1 + cK/2}\right].
\] (24)

which is only \(K\) dependent. As shown in Figure 3, \(g_K\) decreases when \(K\) increases; \(g_K = 1\) when \(K = 0\), and \(g_K \to 0\) when \(K \to \infty\).

Equations 22–24 provide the theoretical foundation for estimating \(K\) when \(N_c \alpha_i \gg 1\). In this case, the parameter \(K\) is interpreted as the number of molecular phenotypes that have experienced strong stabilizing selection. Thereafter we refer to \(K_e\) as the effective number of molecular phenotypes, which is less than the true number of molecular phenotypes. If the ratio \(g_K\) is known, \(K_e\) can be estimated according to Equation 24.

A widely used measure for \(\bar{\lambda} / \nu\) is \(d_{\bar{\lambda}} / d_{\nu}\), the ratio of nonsynonymous to synonymous substitutions, under the assumption that synonymous substitutions are almost neutral. We have realized that the second moment of rate \(\bar{\lambda}^2 / \nu^2\) is related to the \(H\)-measure (Gu et al. 1995) for the rate variation among sites (Uzzell and Corbin 1971), defined by \(H = 1 - (\bar{\lambda})^2 / \bar{\lambda}^2\). Ranging from 0 to 1, a higher value of \(H\) means a greater rate variation among sites and vice versa. One can easily verify

\[
\bar{\lambda}^2 / \nu^2 = \left(\frac{d_{\bar{\lambda}} / d_{\nu}}{1 - H}\right)^2.
\] (25)

Therefore, the effective number of molecular phenotypes \((K_e)\) is the solution of the following equation:

\[
\left(\frac{1}{1 - H}\right) d_{\bar{\lambda}} = g_K = 2^{-K/2} \left[1 + \left(\frac{c \bar{\lambda} + 2K + 2}{1 + cK/2}\right)^{1/2}\right].
\] (26)

**Effective selection intensity \((S_e)\):** Equation 13 indicates that the mean selection intensity \(\bar{S}\) can be written as \(\bar{S} = K \times S_0\), where \(S_0 = -\sum_{i=1}^{K} 2N_c \alpha_i / K\) is the (arithmetic) average of selection intensity over molecular phenotypes. However, without knowing each \(N_c \alpha_i\) it is difficult to estimate \(\bar{S}\). After realizing that the product \(\prod_{i=1}^{K} 2N_c \alpha_i\) is actually estimable, as implied by Equation 22, we define the effective selection intensity as \(S_e = K \times S_0\), where \(S_0\) is the geometric average of selection intensity per molecular phenotype; that is,

\[
S_e = -K \times \left[\prod_{i=1}^{K} 2N_c \alpha_i\right]^{1/K}.
\] (27)

In general, \(S_e \geq \bar{S}\), and \(\bar{S} \approx \bar{S}\) when all \(\alpha_i\)'s are roughly the same. At any rate, both geometric and arithmetic means of selection intensity have virtually the same effect on the rate of protein evolution. Replacing \(\bar{\lambda} / \nu\) by \(d_{\bar{\lambda}} / d_{\nu}\) in Equation 22 and \(K\) by \(K_e\), the effective selection intensity can be estimated by

\[
S_e = \frac{K_e}{2} \left[\frac{d_{\bar{\lambda}} / d_{\nu}}{d_{\bar{\lambda}} / d_{\nu}}\right]^{2/K_e}.
\] (28)

**Estimation procedure:** We propose an analytical pipeline to estimate the effective number of molecular phenotypes \((K_e)\) and the effective selection intensity \((S_e)\). Although it needs to be refined both statistically and computationally, our procedure consists of the following steps:

1. Infer the phylogenetic tree from a multiple alignment of homologous protein sequences. Although there is no methodological preference, we require that the inferred tree topology should be roughly the same over methods.
2. Estimate the nonsynonymous to synonymous ratio \((d_{\bar{\lambda}} / d_{\nu})\) from closely related coding sequences that satisfy \(d_{\bar{\lambda}} / d_{\nu} < 1\).
3. Estimate the \(H\)-measure for rate variation among sites: Use the method of Gu and Zhang (1997) to infer the (corrected) number of changes at each site, under the inferred phylogeny.

Let \(\bar{x}\) and \(\nu(x)\) be the mean and variance of number of changes over sites, respectively. Assuming a Poisson process at each site, we obtain the mean evolutionary \(\bar{\lambda} = \bar{x} / T\), where \(T\) is the total evolutionary time along the tree. Similarly, the variance of evolutionary rate
among sites is given by $V(\lambda) = [V(x) - \bar{x}] / T^2$. Then, $H$ can be estimated by

$$
\hat{H} = \frac{V(\lambda)}{V(\lambda) + (\bar{x})^2} = \frac{V(x) - \bar{x}}{V(x) + \bar{x}(x - 1)}. \quad (29)
$$

And finally, one can estimate $K_s$ and $S_e$ according to Equations 26 and 28, respectively.

**Examples:** For case studies, we compiled 20 vertebrate protein families. Each family includes eight complete sequences from human, mouse, dog, cow, chicken, Xenopus, fugu, and zebrafish, respectively, which were downloaded from Ensembl EnMar (http://www.ensembl.org/Multi/martview) (KASPRZYK et al. 2004). The synonymous distance ($d_s$) and the nonsynonymous distance ($d_N$) between mammalian orthologs were estimated by *codeml* of the PAML package (YANG 1997). We used the same phylogeny (Figure 4) to estimate the $H$-index of rate variation among sites; using an alternative tree gave a very similar result (not shown).

Table 1 shows the results. The $d_N/d_s$ ratio varies substantially among genes (from 0.001 to 0.274), and the $H$-index ranges from 0.286 to 0.886. Consequently, the estimated effective number of molecular phenotypes $K_e$ ranges from 2.3 to 20.4, with an average of 8.77 (the median is 8.23). Our analysis suggests that most genes are highly pleiotropic. That is, there are on average eight distinct components in the fitness that may be related to a gene. Roughly, ~73% of the among-gene variation in the $d_N/d_s$ ratio can be explained by the variation in $K_e$ among genes. In other words, the level of gene pleiotropy associated with a gene may be the dominant factor for predicting sequence conservation, as predicted by FISHER (1930). The effective mean selection intensity ($S_e$) ranges from 5.39 to 23.63, with a median of 12.6. Note that $d_N/d_s$ ratios in Table 1 were estimated from the human and mouse orthologous genes; using other mammals or taking an average would lead to ~5% difference in the $K_e$ estimation.

**DISCUSSION**

**Estimation of gene pleiotropy:** Although the theory of gene pleiotropy, the capacity of a gene to affect multiple phenotypic characters, has been used to explain many biological phenomena (FISHER 1930; WRIGHT 1968; BARTON 1990; WAXMAN and PECK 1998; OTTO 2004; MACLEAN et al. 2004; DUDLEY et al. 2005), the characteristic level of pleiotropy remains largely unknown. Many phenotype-genotype models (e.g., FISHER 1930; WRIGHT 1968; HARTL and TAUBES 1996, 1998; WAXMAN and PECK 1998; POON and OTTO 2000) have linked gene pleiotropy to the dimensionality of the fitness function of the organism. The major contribution of this article is to formulate a statistical framework for estimating the capacity of a gene to significantly affect distinct components in the fitness, the effective number of molecular phenotypes ($K_e$). Thus, for the first time we have an empirical measure of gene pleiotropy. Examples in Table 1 show that $K_e$ is typically ~6-9, the number of distinct fitness components a gene may affect, which has been further confirmed by a large-scale data analysis (not shown).

Our SM model provides an approach to understanding the role of gene pleiotropy in a gene network. Although a detailed analysis will be published elsewhere, we test the hypothesis that gene pleiotropy underlies genomic correlations related to the evolutionary rate of protein sequence (HE and ZHANG 2006; PAL et al. 2006; WOLF et al. 2006). That is, (i) evolutionary rate is

**TABLE 1**

<table>
<thead>
<tr>
<th>Genes</th>
<th>$d_N/d_s$ (H–M)</th>
<th>$H$</th>
<th>$K_e$</th>
<th>$S_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adiponectin receptor protein 2</td>
<td>0.109</td>
<td>0.785</td>
<td>2.30</td>
<td>-12.32</td>
</tr>
<tr>
<td>GTP cyclohydrolase 1</td>
<td>0.188</td>
<td>0.586</td>
<td>2.68</td>
<td>-7.16</td>
</tr>
<tr>
<td>F-box protein 34</td>
<td>0.274</td>
<td>0.355</td>
<td>2.90</td>
<td>-5.39</td>
</tr>
<tr>
<td>RNA</td>
<td>0.221</td>
<td>0.422</td>
<td>3.26</td>
<td>-6.18</td>
</tr>
<tr>
<td>methyltransferase-like 1</td>
<td>0.148</td>
<td>0.393</td>
<td>4.79</td>
<td>-7.63</td>
</tr>
<tr>
<td>Galectin-3</td>
<td>0.047</td>
<td>0.729</td>
<td>5.97</td>
<td>-11.65</td>
</tr>
<tr>
<td>Glia maturation factor (GMF)β</td>
<td>0.076</td>
<td>0.514</td>
<td>6.31</td>
<td>-9.94</td>
</tr>
<tr>
<td>Metalloprotease 1</td>
<td>0.094</td>
<td>0.650</td>
<td>6.64</td>
<td>-11.34</td>
</tr>
<tr>
<td>Nucleolar GTP-binding protein 1</td>
<td>0.063</td>
<td>0.286</td>
<td>8.18</td>
<td>-10.80</td>
</tr>
<tr>
<td>Sperm tail protein SHIPPO1</td>
<td>0.042</td>
<td>0.524</td>
<td>8.23</td>
<td>-11.96</td>
</tr>
<tr>
<td>Transaldolase</td>
<td>0.030</td>
<td>0.654</td>
<td>8.23</td>
<td>-12.93</td>
</tr>
<tr>
<td>CGTHBA protein</td>
<td>0.024</td>
<td>0.582</td>
<td>9.64</td>
<td>-13.77</td>
</tr>
<tr>
<td>Rhomboid family 1</td>
<td>0.020</td>
<td>0.641</td>
<td>9.80</td>
<td>-14.38</td>
</tr>
<tr>
<td>Nucleoredoxin</td>
<td>0.023</td>
<td>0.573</td>
<td>9.85</td>
<td>-13.93</td>
</tr>
<tr>
<td>Myosin lc (Myosin Iβ)</td>
<td>0.023</td>
<td>0.566</td>
<td>9.94</td>
<td>-13.97</td>
</tr>
<tr>
<td>Exocyst complex component 5</td>
<td>0.034</td>
<td>0.343</td>
<td>9.97</td>
<td>-12.89</td>
</tr>
<tr>
<td>Plakophilin-3</td>
<td>0.014</td>
<td>0.360</td>
<td>12.74</td>
<td>-15.83</td>
</tr>
<tr>
<td>Homeobox protein OTX2</td>
<td>0.001</td>
<td>0.886</td>
<td>15.83</td>
<td>-23.63</td>
</tr>
<tr>
<td>Cornichon homolog (TGAM77)</td>
<td>0.001</td>
<td>0.777</td>
<td>17.77</td>
<td>-23.61</td>
</tr>
<tr>
<td>14-3-3 protein epsilon (14-3-3E)</td>
<td>0.001</td>
<td>0.507</td>
<td>20.40</td>
<td>-24.27</td>
</tr>
</tbody>
</table>

The $d_N/d_s$ ratio was calculated from the human (H) and mouse (M) orthologous genes.
inversely related to gene pleiotropy \( K \), (ii) gene pleiotropy determines gene dispensability, and (iii) gene pleiotropy underlies “hub” genes. In short, estimation of gene pleiotropy \( K_e \) provides a model-based data-analysis approach that avoids the interpretative difficulty in principle-component or partial-correlation analysis.

In the future, we shall refine the analytical pipeline for estimating \( K_e \). We shall examine how statistical properties of \( ds/dk \) and \( H \) would affect the estimation efficiency of \( K_e \). In addition to the delta method, another method to estimate the sampling variance of the estimated \( K_e \) is nonparametric bootstrapping. Under the \( W \) model when the selection intensity \( S \) follows a (negative) gamma distribution, a likelihood approach can be developed. Since the method of Nielsen and Yang (2003) is applicable only for closely related sequences, we shall develop a protein sequence-based likelihood approach.

Gene pleiotropy and allelic pleiotropy: It should be noted that the effective number of molecular phenotypes \( K_e \) measures the functional pleiotropy of a gene, rather than a single mutation. For instance, \( K_e = 8.23 \) for trans-aldolase is the estimated number of distinct fitness dimensions implied by all random mutations in the coding sequence. Although the estimated \( K_e \) indicates that most genes may be highly pleiotropic, what it exactly means is very much gene specific.

The number of fitness dimensions affected by a single mutation \( \Omega \) could be much less than \( K_e \), say, \( \sim 2–4 \) (Dudley et al. 2005). Waxman and Peck (1998) argued that a single optimal gene sequence may become common when three or more characters (as molecular phenotypes here) are affected by each mutation, \( \langle \Omega \rangle \geq 3 \). The problem of estimating \( \Omega \) effectively from sequence data remains a question that requires further study.

Several studies (Nielsen and Yang 2003; Piganeau and Eyre-Walker 2003; Eyre-Walker et al. 2006) estimated the shape parameter of gamma distribution for \( S \). Nielsen and Yang (2003) obtained the shape parameter of 3.22 from the primate mitochondrial genome sequences. Under the SM framework, it can be interpreted as \( K_e/2 \); see the above section for the \( W \) model. That is, \( K_e = 2 \times 3.22 = 6.44 \) for primate mitochondria proteins, consistent with our results (Table 1). In contrast, Eyre-Walker et al. (2006) obtained a much smaller value (0.23) of the shape parameter from the human SNP data. We speculate that the estimate from population genetics data perhaps reflects the allelic pleiotropy \( \Omega \). Hence, Eyre-Walker et al. (2006) may provide a rough estimate for \( \Omega = 2 \times 0.23 = 0.46 \).

Assumptions and alternative models: The SM model involves several simplifying assumptions, since an exhaustive consideration of all possible mutational effects on molecular phenotypes is intractable. Many are shown with previous phenotype–genotype models. One may refer to Martin and Lenormand (2006) for a recent summary about rationales and criticisms. The first assumption is the single fitness optimum. Under this condition we assume that disruptive selection with alternative optima following an adaptive process (reviewed in Kingsolver et al. 2001 and Elena and Lenski 2003) occurs rarely at the molecular level. Moreover, we used the Gaussian fitness function (see Equation 1) to measure the fitness consequences of a deviation from the optimum. Lande (1980) showed that when the population mean is close to the optimum, a Gaussian-like function can be a local kernel approximation for many arbitrary fitness functions.

Second, the effect of mutations on molecular phenotypes is assumed to be continuous and symmetric. The continuous property is consistent with some empirical evidence from random mutations of genes (Keightley 1994; Keightley and Caballero 1997; Imhof and Schlotterer 2001). The symmetric assumption implies that even if mutations bias toward higher or lower values of molecular phenotypes, the trend should be small compared to the mutational variance. In practice we used a multiple Gaussian distribution for mutational effects. In the non-Gaussian case, it is possible to reduce kurtosis after an appropriate transformation such that mutational effects on a “transformed” molecular phenotype are roughly Gaussian-like. In short, we conclude that in the absence of practically useful alternatives, single optimum, Gaussian fitness function, and continuous symmetric effect of mutations remain useful working assumptions. Further work will examine the effects of these assumptions using computer simulations.

A more important issue is whether the estimation procedure for the effective number of molecular phenotypes \( K_e \) is robust under various alternative models. The first alternative model is Fisher’s geometric model, in which adaptive change is represented by stepwise movement of a point toward the center of the hypersphere. In this case, the number of molecular phenotypes \( K \) is Fisher’s geometric dimension. Several studies (Hartl and Taubes 1996, 1998; Sella and Hirsh 2005) showed that at equilibrium, Fisher’s geometric model can be approximated by allowing the current population to vary randomly around the fixed optimum. A second alternative model is to define the fitness to be a function of flux through a metabolic pathway (Kacser and Burns 1979; Hartl and Taubes 1996), rather than a Gaussian-like function. Preliminary results have indicated that in both cases, estimation of \( K_e \) varies only slightly. It seems that \( K_e \) is a robust measure of gene pleiotropy, which may not be very sensitive to the specific evolutionary model. We shall address this issue more extensively in future studies.

The author is grateful to Adam Eyre-Walker and the reviewing editor Antony Long for constructive comments on the manuscript. Assistance in data analysis from Dongping Xu and Zhixi Su is appreciated.

**LITERATURE CITED**

Evolutionary Model and Gene Pleiotropy


Communicating editor: A. D. Long

APPENDIX

Mean and variances of $S$—derivation of Equations 13 and 14: For the distribution of $y$ that follows a multi ($K$)-variate normal distribution,

$$p(y) = \frac{1}{(\sqrt{2\pi})^K |\Sigma_m|^{1/2}} \exp \left[ -\frac{y^T \Sigma_m^{-1} y}{2} \right],$$

we note that for any quadratic function $y^T U^{-1} y$, the first and second moments can be expressed as

$$E(y^T U^{-1} y) = tr(U^{-1} \Sigma_m)$$

$$E[(y^T U^{-1} y)^2] = 2 tr((U^{-1} \Sigma_m)^2) + (tr(U^{-1} \Sigma_m))^2,$$

respectively, where the operator $tr[.]$ means the trace of a square matrix. From Equation 7 it is straightforward that

$$S = -2N tr(U^{-1} \Sigma_m) \text{ and } Var(S) = 8N^2 tr(U^{-1} \Sigma_m^2).$$

Denote the matrix $A = U^{-1} \Sigma_m$. After noting that the trace of the matrix is the sum of the eigenvalues, i.e., $\text{tr}[A] = \sum_{i=1}^K \alpha_i$, and $\text{tr}[A^2] = \sum_{i=1}^K \alpha_i^2$, we obtain Equations 13 and 14.

Canonical representation of $S$—derivation of Equation 15: The canonical representation will choose molecular phenotypes such that $\Sigma_w = \text{diag}(\sigma^2_{w,1}, \ldots, \sigma^2_{w,K})$. Recall that $A = U^{-1} \Sigma_m = Z - \Omega$, where $Z = \Sigma_w^{-1} \Sigma_m$.
and \( \Omega = \Sigma_{w}^{-1} \Sigma_{m} Z \). Then, one can verify \( \text{tr}[Z] = \text{tr}[\text{diag}(\alpha_{w,1}^{2}, \ldots, \alpha_{w,K}^{2})] = \sum_{i=1}^{K} \alpha_{m,i}^{2} / \alpha_{w,i}^{2} \), since the trace of a matrix is the sum of its diagonal elements. Next, matrix \( \Omega \) can be written as \( \Omega = \text{diag}(\alpha_{w,1}^{2}, \ldots, \alpha_{w,K}^{2}, K)^{-1} \Sigma_{m} \). It follows that \( \text{tr}[\Omega] = \sum_{i=1}^{K} \delta_{i}^{2} / \alpha_{w,i}^{4} \), where \( \delta_{i}^{2} \) is the \( i \)-th eigenvalue of matrix \( \Sigma_{m} \). From the relation \( \text{tr}[A] = \text{tr}[Z] - \text{tr}[\Omega] \), together we have

\[
\text{tr}[A] = \sum_{i=1}^{K} \frac{\delta_{i}^{2}}{\alpha_{w,i}^{4}} \left( 1 - \frac{\delta_{i}^{2}}{\alpha_{w,i}^{4}} \right) = \sum_{i=1}^{K} \frac{\alpha_{m,i}^{2}}{\alpha_{w,i}^{4}} \left( 1 - \frac{\alpha_{m,i}^{2}}{\alpha_{w,i}^{4}} \right) \tag{A3}
\]

leading to Equation 15, where \( \gamma_{i} = \delta_{i}^{2} / [\alpha_{w,i}^{2} \alpha_{m,i}^{2}] \).

The canonical form for the variance of \( S \) can be obtained similarly, but the algebra is somewhat tedious, largely due to the detail of \( A^{2} = Z^{2} + \Omega^{2} - 2Z\Omega \). Let \( M_{i}^{2} = \alpha_{m,i}^{2} + \sum_{j=1}^{K} \alpha_{m,j}^{2} \). One thus can show the trace \( \text{tr}[Z^{2}] = \sum_{i=1}^{K} \left( M_{i}^{2} / \alpha_{w,i}^{4} \right)^{2} \) and \( \text{tr}[\Omega^{2}] = \sum_{i=1}^{K} \frac{\delta_{i}^{2}}{\alpha_{w,i}^{4}} \).

Noting that \( Z\Omega = \text{diag}(\alpha_{w,1}^{2}, \ldots, \alpha_{w,K}^{2})\Sigma_{m} \), we obtain \( \text{tr}[Z\Omega] = \sum_{i=1}^{K} \beta_{i}^{2} / \alpha_{w,i}^{4} \), where \( \beta_{i}^{2} \) is the \( i \)-th eigenvalue of matrix \( \Sigma_{m} \). Therefore, from the fact \( \text{tr}[A^{2}] = \text{tr}[Z^{2}] + \text{tr}[\Omega^{2}] - 2 \text{tr}[Z\Omega] \), we have

\[
\text{Var}(S) = 8N^{2} \sum_{i=1}^{K} \left( \frac{M_{i}^{2}}{\alpha_{w,i}^{4}} \right)^{2} + \left( \frac{\delta_{i}^{2}}{\alpha_{w,i}^{4}} \right)^{2} - 2 \frac{\beta_{i}^{2}}{\alpha_{w,i}^{4}} \tag{A4}
\]

**Condition of \( S \)-gamma distribution:** Given \( p(y) \) characterized by the covariance \( \Sigma_{w} \), the distribution of selection intensity, \( f(S) \), can be uniquely determined by the probability theory that for any \( S^{y} \) and \( y^{y} \) that satisfy \( S^{y} = -2N_{c}(y^{y} \Sigma_{w}^{-1} y^{y}) \) in Equation 7, the relationship \( \int S^{y} f(S) dS = -2N_{c} \int y^{y} \Sigma_{w}^{-1} y^{y} p(y) dS \). Here, the analytical form of \( f(S) \) is not available, except for the case when the matrix \( U \) is in the quadratic form and the covariance in \( p(y) \), \( \Sigma_{m} \), satisfies \( A = \Sigma_{m} U^{-1} = A_{1} \). In other words, the matrix \( A \) is diagonal, having the same eigenvalues \( \alpha_{i} \).

**Moments of evolutionary rate—derivation of Equations 20 and 21:** Here we show several analytical integral results. First we consider \( I_{1} = \int e^{-2N_{c}yU^{-1}y} p(y) dy \), which is given by

\[
I_{1} = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} e^{-2N_{c}yU^{-1}y} \times \frac{\exp[-y^{1/2} \Sigma^{-1/2}] \Sigma^{-1/2}}{(\sqrt{2\pi})^{K} |\Sigma_{m}|^{1/2}} dy = \frac{|B|^{1/2}}{|\Sigma_{m}|^{1/2}} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp[-y^{1/2} \Sigma^{-1/2} B^{-1/2} \Sigma^{-1/2}] \times \frac{\exp[-y^{1/2} B^{-1/2} \Sigma^{-1/2}]}{(\sqrt{2\pi})^{K} |B|^{1/2} |\Sigma_{m}|^{1/2}} \tag{A5}
\]

where \( B = (\Sigma_{m}^{-1} + 4N_{c}U^{-1})^{-1} \). It follows that

\[
I_{1} = \frac{|\Sigma_{m}^{-1} + 4N_{c}U^{-1}|^{-1/2}}{|\Sigma_{m}|^{1/2}} = |I + 4N_{c}U^{-1} \Sigma_{m}|^{-1/2} = |I + 4N_{c}A|^{-1/2}. \tag{A6}
\]

In the same manner, we have

\[
I_{2} = \int e^{-4N_{c}yU^{-1}y} p(y) dy \quad \text{given by}
\]

\[
I_{2} = |I + 8N_{c} \Sigma_{w}^{-1} \Sigma_{m}^{-1/2}|^{-2} = |I + 8N_{c} A|^{-1/2}. \tag{A7}
\]

Next we consider \( I_{3} = \int [2N_{c}yU^{-1}y] e^{-2N_{c}yU^{-1}y} p(y) dy \). From the derivation of \( I_{1} \), one can verify

\[
I_{3} = \left| \frac{B^{1/2}}{|\Sigma_{m}|^{1/2}} \right| \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} (2N_{c}yU^{-1}y)^{\exp[-y^{1/2} B^{-1/2} \Sigma^{-1/2}]} (\sqrt{2\pi})^{K} |B|^{1/2} dy. \tag{A8}
\]

From Equation A2, we obtain

\[
I_{3} = \frac{|B|^{1/2}}{|\Sigma_{m}|^{1/2}} \times 2N_{c} \text{tr} [\Sigma_{w}^{-1} \Sigma_{m}^{-1/2}] = \frac{2N_{c} \text{tr} \left( \Sigma_{m}^{1/2} U + 4N_{c} I^{-1} \right)}{|I + 4N_{c} U^{-1} \Sigma_{m}|^{1/2}} = \frac{2N_{c} \text{tr} \left( A^{-1} + 4N_{c} I^{-1} \right)}{|I + 4N_{c} A|^{1/2}}. \tag{A9}
\]

In the same manner, we show that

\[
I_{4} = \left| \frac{B^{1/2}}{|\Sigma_{m}|^{1/2}} \right| \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} (2N_{c}yU^{-1}y)^{\exp[-y^{1/2} B^{-1/2} \Sigma^{-1/2}]} (\sqrt{2\pi})^{K} |B|^{1/2} dy.
\]

Finally, we consider the integral \( I_{5} = \int [2N_{c}yU^{-1}y]^{e^{-4N_{c}yU^{-1}y} p(y)} dy \). Similar to the derivations of \( I_{1}-I_{5} \), one can write

\[
I_{5} = \frac{|B|^{1/2}}{|\Sigma_{m}|^{1/2}} \times \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} (2N_{c}yU^{-1}y)^{\exp[-y^{1/2} B^{-1/2} \Sigma^{-1/2}]} (\sqrt{2\pi})^{K} |B|^{1/2} dy.
\]

where matrix \( B^{*} = (\Sigma_{m}^{-1} + 4N_{c}U^{-1})^{-1} \). Since \( U^{-1} B^{*} = (\Sigma_{m}^{1/2} U + 4N_{c} I^{-1})^{-1} \), and \( |B^{*} / \Sigma_{m}| = |I + 8N_{c} A| \), \( I_{5} \) can be written as follows:

\[
I_{5} = |I + 8N_{c} A|^{-1/2} 
\]

\[
	imes \left\{ 8N^{2} \text{tr} \left[ (A^{-1} + 8N_{c} I^{-1})^{-2} \right] + (2N_{c} \text{tr} [A^{-1} + 8N_{c} I^{-1}]) \right\}. \tag{A12}
\]

From Equation 19, one can show that \( \tilde{\alpha} / u = I_{1} + d_{3} \) and \( \tilde{\alpha}^{2} / u^{2} = I_{2} + 2d_{4} + c^{2}I_{5} \), both of which are determined a single matrix \( A \). Hence, we have derived Equations 20 and 21 on the basis of the matrix theory that, for a square matrix \( A \), the trace is the sum of the eigenvalues, the determinant is the production of the eigenvalues, and any eigenvalue of the matrix that is a function of \( A \) is the function of the corresponding eigenvalue of \( A \).