Stochastic Demography and the Neutral Substitution Rate in Class-Structured Populations

Laurent Lehmann[1]
Department of Ecology and Evolution, University of Lausanne, Sorge, Le Biophore, CH -1015 Lausanne, Switzerland

ABSTRACT The neutral rate of allelic substitution is analyzed for a class-structured population subject to a stationary stochastic demographic process. The substitution rate is shown to be generally equal to the effective mutation rate, and under overlapping generations it can be expressed as the effective mutation rate in newborns when measured in units of average generation time. With uniform mutation rate across classes the substitution rate reduces to the mutation rate.

CONSIDER a haploid population of constant size N without overlapping generations, where neutral mutant alleles are substituted sequentially at a given locus through the constant input of mutations at rate μ per gene. Then, the rate k of allelic substitution is equal to the mutation rate; that is, $k = \mu$ (Kimura 1971). Does this simple result extend to more realistic demographic scenarios? In a population with overlapping generations where individuals reproduce at discrete time points and mutations arise only in newborns (the standard assumption), the substitution rate is $k = \mu_0/T$, where μ_0 is the rate at which a newborn accumulates mutations at the gene locus under focus and T is the generation time (Pollak 1982, equation 11). This is the average age of the parent of a randomly sampled individual among the N_0 surviving newborns. When measured in units of T time steps, the substitution rate thus depends only on μ_0 (Pollak 1982, equation 15; Charlesworth 1994, p. 94).

There is, however, an even simpler interpretation of k in terms of mutation rate under overlapping generations. This can be reached by observing that the total population size N necessarily satisfies $N = TN_0$, since the generation time is the number of units of time the population takes to produce N newborns when only N_0 are produced per unit time (an observation implied by Felsenstein 1971, equation 1). With this, $k = \overline{\mu}$, where $\overline{\mu} = \mu_0 N_0/N$ is the mutation rate at a randomly sampled gene from the population, since mutations arise only in newborns and they form a fraction N_0/N of the population.

How do these two different interpretations of k in terms of mutation rate generalize to stochastic demography, where population sizes can fluctuate over time? This article develops a full stochastic demographic model of neutral evolution in a class-structured population (*e.g.*, by sex, age, stage). The model shows that the substitution rate is precisely the effective mutation rate μ_e, which is the rate at which a gene lineage accumulates mutations (Rousset 2004, p. 158), and can be different from the mutation rate at a randomly sampled gene from a randomly sampled individual in the population. In an age-structured population the effective mutation rate is shown to reduce to the effective mutation rate in newborns when measured in units of average generation time, which can itself be expressed in terms of class reproductive values. The use of such reproductive values turns out to be central in structuring the connections between substitution rate, average mutation rate, effective mutation rate, and generation time.

Model

Biological assumptions

Demography: We now consider a population structured into a finite number of classes and evolving in discrete time. This class structure could, for instance, result from the presence of males and females, of age classes, of groups of individuals located at different positions in the habitat, or a combination of these or other factors. The number of individuals in class $i = 0, 1, 2, \ldots$ is written N_i and the vector $\mathbf{n} \equiv (N_0, N_1, N_2, \ldots)$ denotes a state of the population, which gives the

doi: 10.1534/genetics.114.163345
Manuscript received December 19, 2013; accepted for publication February 27, 2014; published Early Online March 3, 2014.
[1]Address for correspondence: Department of Ecology and Evolution, UNIL Sorge, Le Biophore, CH -1015 Lausanne, Switzerland. E-mail: laurent.lehmann@unil.ch

Table 1 List of symbols

Symbol	Definition
k	Substitution rate.
N_i	No. class i individuals.
$\overline{N}$	Average no. gene copies in the population.
$\Pr(\mathbf{n})$	Stationary probability that the population is in state $\mathbf{n}$.
$\Pr(\mathbf{n}'\|\mathbf{n})$	Forward transition probability from state $\mathbf{n}$ to state $\mathbf{n}'$.
g_i	Ploidy of a class i individual.
t_{ij}	Probability that a gene randomly sampled in a class-i individual was transmitted by a class-j individual.
μ_{ij}	Mutation rate at a gene transmitted by a class-j individual to a class-i individual.
$\overline{\mu}$	Average mutation rate in the stationary demographic regime.
μ_e	Effective mutation rate.
$\mu_{e,0}$	Effective mutation rate in newborns.
T	Average generation time (or mean age of the parents of a newborn).
$T(\mathbf{n})$	Average generation time in population state $\mathbf{n}$.
$w_{ij}(\mathbf{n}', \mathbf{n})$	Expected no. class-i individuals in a population in state $\mathbf{n}'$ descending from a single class-j individual in population state $\mathbf{n}$.
$f_{ij}(\mathbf{n}', \mathbf{n})$	Probability that a gene sampled in a class-i individual when the population is in state $\mathbf{n}'$ is a copy of a gene of a class-j individual given the population was in state $\mathbf{n}$ in the parental generation.
$a_{ij}(\mathbf{n}', \mathbf{n})$	Probability that a gene sampled in a class-i individual when the population is in state $\mathbf{n}'$ is a copy of a gene of a class-j individual and the population was in state $\mathbf{n}$ in the parental generation.
$\pi_i(\mathbf{n})$	Fixation probability of a single mutant residing in a class-i individual.
$\overline{\pi}$	Average fixation probability of a single mutant.

realized number of individuals in each class i at a census point (see Table 1 for a list of symbols).

The change in the demographic state of the population is assumed to follow a discrete-time stochastic process and we denote by $\Pr(\mathbf{n}'|\mathbf{n})$ the transition probability per unit time that a population in state $\mathbf{n}$ will be in state $\mathbf{n}'$ in the next time step. This defines a homogeneous Markov chain, which may be driven by both endogenous and exogenous factors. This demographic process is further assumed to have a finite state space and to be ergodic (irreducible and aperiodic), conditional on the nonextinction of the population. In force of these assumptions, the distribution of the process will approach the stationary probability $\Pr(\mathbf{n})$ of being in state $\mathbf{n}$ (Karlin and Taylor 1975; Grinstead and Snell 1997), which determines the stationary demographic regime.

Genetics: Neutral evolution will be investigated at a given locus in the backdrop of the stationary demography. Neutrality means that the number of successful "offspring" (number of descendants over one time step) of all individuals in the same class j are an exchangeable random variable, and this holds for each class. This entails that the forward transition probability $\Pr(\mathbf{n}'|\mathbf{n})$ and the probability $a_{ij}(\mathbf{n}', \mathbf{n})$ that a gene randomly sampled in a class-i individual in population state $\mathbf{n}'$ descends from a class-j individual and from population state $\mathbf{n}$ in the previous time step do not depend on the genetic state (allele frequency distribution) of the population ($\sum_{\mathbf{n}}\sum_j a_{ij}(\mathbf{n}', \mathbf{n}) = 1$).

Individuals in different classes, like males and females, may have different ploidies and g_i will denote the ploidy of an individual of class i at the locus of interest. Mutation may also differ across classes and μ_{ij} will denote the mutation rate at a gene transmitted by a class-j individual to a class-i individual. This allows us to evaluate the rate at which a randomly sampled gene in a randomly sampled individual of class i in population state $\mathbf{n}'$ accumulates mutations, which is

$$\mu_i(\mathbf{n}') = \sum_{\mathbf{n}}\sum_j a_{ij}(\mathbf{n}', \mathbf{n})\mu_{ij}. \tag{1}$$

This generalizes to any class and to stochastic demography, the standard expression for the rate at which a randomly sampled gene in a newborn accumulates mutations in models of overlapping generations with constant size (Pollak 1982, equations 1 and 2, Charlesworth 1994, equation 2.49b).

From Equation 1, we can define the following average mutation rate,

$$\overline{\mu} = \frac{1}{\overline{N}}\sum_{\mathbf{n}}\sum_j \mu_j(\mathbf{n})g_j N_j \Pr(\mathbf{n}), \tag{2}$$

which gives the ratio of the average number of mutations produced in the population to the average number $\overline{N} = \sum_{\mathbf{n}}\sum_j g_j N_j \Pr(\mathbf{n})$ of genes in the population (average size of the gene pool). The mutation rate $\overline{\mu}$ gives the rate at which a randomly sampled gene in a randomly sampled individual accumulates mutations, since over replicates of the evolutionary process the probability of sampling an individual in demographic state $\mathbf{n}$ is $N(\mathbf{n})\Pr(\mathbf{n})/\overline{N}$, where $N(\mathbf{n}) = \sum_j g_j N_j$, in which case the probability this is a class-j individual is $g_j N_j/N(\mathbf{n})$. Hence, the probability of sampling a mutant individual of class j in state $\mathbf{n}$ is proportional to the size of its class and the probability of occurrence of that state.

The aim of this note is to express the neutral substitution rate k in terms of a mutation rate averaged over the class

specific mutation rates, the $\mu_j(\mathbf{n})$'s, and we will see that in general this quantity is not equal to $\overline{\mu}$.

Substitution rate

Forward process: The substitution rate k is the expected number of mutations that will fix in the population per unit time (Kimura 1971; Pollak 1982). Under the above assumptions, this can be written as

$$k = \overline{N}\overline{\mu}\,\overline{\pi}, \tag{3}$$

where

$$\overline{\pi} = \sum_{\mathbf{n}} \sum_{j} \pi_j(\mathbf{n}) \frac{g_j N_j \mu_j(\mathbf{n}) \Pr(\mathbf{n})}{\sum_s \sum_h g_h N_h \mu_h(\mathbf{s}) \Pr(\mathbf{s})} \tag{4}$$

is the average fixation probability of a single mutant allele, which depends on the fixation probability $\pi_j(\mathbf{n})$ of a mutant arising as a single copy in a class-j individual in population state $\mathbf{n}$ and on the probability that the mutation arises in this state (see *Appendix A* for a proof). This latter quantity is given in Equation 4 as the ratio of the number of mutations arising in class-j individuals in population state $\mathbf{n}$ to the total number of mutations arising per unit time.

Three points are now made about Equation 3. First, this is exactly of the same form as the classical expression for the substitution rate (Kimura 1971, equation 4.2), but where all quantities are averages. Second, if the underlying mutation model at the locus of interest is the infinite-allele model (Kimura and Crow 1964; Kimura 1971), then the average mutation rate $\overline{\mu}$ must be very small so that novel mutations occur in homoallelic populations; otherwise the population will hardly ever fixate for an allele. By contrast, under the infinite-site mutation model (Kimura 1969, 1971), free recombination among sites and a very small mutation rate per site entail that mutants arise at homoallelic sites, so that several mutations can segregate simultaneously and independently in the population. Then, Equation 3 gives the number of mutants fixing at different sites per unit time and this is the classical situation envisioned for the substitution process (Kimura 1971, p. 183). Third, k does generally not reduce to $\overline{\mu}$, since $\overline{\pi}$ depends on the mutation distribution and simplifies to $1/\overline{N}$ only under special cases. For instance, this is the case when the mutation rate is the same in all classes $\mu_{ij} = \mu$ for all i and j, which entails that $\mu_j(\mathbf{n}) = \mu$. Then $\overline{\pi} = 1/\overline{N}$ (Lehmann 2012, appendix A), and the substitution rate becomes equal to the mutation rate

$$k = \mu. \tag{5}$$

Backward process: To further simplify the expression for k, we now express the fixation probability $\pi_i(\mathbf{n})$ in terms of class reproductive values (*e.g.*, Taylor 1990; Rousset 2004), which allows us to look at neutral evolution in terms of a backward genealogical process, much like coalescent theory does for effective population size (*e.g.*, Wakeley 2008).

Imagine we trace backward in time the ancestral lineage of a randomly sampled gene in the population in the present, at time $h = 0$. Then, the probability $\gamma_{j,\,h}(\mathbf{n})$ that this lineage resides in an individual of class j and in population state $\mathbf{n}$ at time h in the past satisfies the recursion

$$\gamma_{j,h+1}(\mathbf{n}) = \sum_{\mathbf{n}'} \sum_{i} \gamma_{i,h}(\mathbf{n}') a_{ij}(\mathbf{n}', \mathbf{n}). \tag{6}$$

Assuming that every gene position (class of individual and population state) can be reached in the long run, the circulation of the ancestral gene lineage among the classes of individuals and population states determined by the $a_{ij}(\mathbf{n}', \mathbf{n})$ coefficients defines an ergodic Markov chain that will eventually reach a stationary distribution [given by the left unit eigenvector of the ergodic Markov matrix with transition probabilities $a_{ij}(\mathbf{n}', \mathbf{n})$ (Karlin and Taylor 1975; Grinstead and Snell 1997)].

The probability that a gene lineage resides in class i and in population state $\mathbf{n}$ under the stationary distribution is denoted $\gamma_i(\mathbf{n})$, which is the reproductive value of class $(i, \mathbf{n})$ and is equal to $\alpha_i(\mathbf{n})\Pr(\mathbf{n})$, where $\alpha_i(\mathbf{n})$ is the probability that the ancestral lineage of a randomly sampled gene in the population was in class i, given population state $\mathbf{n}$ in the distant past [the reproductive value of class i (Rousset 2004, p. 181)]. With this, we can write the fixation probability of a single mutant allele entering the population in state $\mathbf{n}$ as

$$\pi_i(\mathbf{n}) = \frac{\alpha_i(\mathbf{n})}{g_i N_i}, \tag{7}$$

where $1/[g_i N_i]$ is the frequency of a single mutant in class i (see *Appendix B* for a proof).

Substituting Equation 4 and Equation 7 into Equation 3 and rearranging produces

$$k = \sum_{\mathbf{n}} \sum_{i} \gamma_i(\mathbf{n}) \mu_i(\mathbf{n}), \tag{8}$$

which is a reproductive value weighted average mutation rate.

Substitution rate as effective mutation rate

The effective mutation μ_e is defined as the average rate at which a gene lineage accumulates mutations (Rousset 2004, equation 9.36). This is precisely an interpretation that can be given to the right-hand side of Equation 8. In effect, $\gamma_i(\mathbf{n}')$ can be interpreted as the probability that the ancestral lineage of a randomly sampled gene was in class i in population state $\mathbf{n}'$ in the distant past, $a_{ij}(\mathbf{n}', \mathbf{n})$ is the probability that this gene descends from an individual in class j in population state $\mathbf{n}$ in the previous time step, and μ_{ij} is the mutation rate during such a transition. Hence,

$$k = \mu_e, \tag{9}$$

which shows that the neutral substitution rate is the effective mutation rate regardless of the genetic and

demographic assumptions behind the model. This correspondence is not surprising given the definition of μ_e, but Equation 8 makes precise that while k has originally been defined in terms of a forward in time looking process (fixation probabilities, see Equation 3), it can be interpreted in terms of a meaningful class-reproductive value weighted average mutation rate, which defines a backward in time looking process (*e.g.*, Equation 6). The substitution rate is thus also the rate at which a randomly sampled gene from a randomly sampled common ancestor of the population accumulates mutations, since with probability $\gamma_i(\mathbf{n})$ the common ancestor descends from state $(i, \mathbf{n})$, in which case it carries mutations at rate $\mu_i(\mathbf{n})$.

The substitution rate k will generally depend on population size fluctuations (Balloux and Lehmann 2012) as the reproductive values generally depend on such fluctuations (Whitlock and Barton 1997; Rousset 2004). Further, these are actually very hard to evaluate explicitly and most of the time can be expressed implicitly only as the solution of a very large set of simultaneous equations involving population sizes (*e.g.*, Equation 6). But if

$$\alpha_i(\mathbf{n}) = \frac{g_i N_i}{\overline{N}}, \tag{10}$$

then $\pi_i(\mathbf{n}) = 1/\overline{N}$ and from Equation 7 we have

$$k = \overline{\mu}. \tag{11}$$

When will the reproductive value of a class be proportional to the number of genes in that class so that this result holds?

If evolution occurs in a patch-structured haploid population of constant size in which migration does not affect group size, migration is said to be conservative and Equation 10 holds (Nagylaki 1998, p. 1600). More generally, however, migration will result in population size fluctuations and density-dependent competition can result in very large populations (or patches) producing only very few individuals. In these cases, the reproductive value of a class is unlikely to be proportional to the number of genes in that class, because the size of a demographic class is not necessarily indicative of its asymptotic contribution to the gene pool. For instance, under good environmental conditions a small group of individuals may contribute a disproportionally larger share to the gene pool than a big group of individuals in poor environmental conditions. Then, μ_e will differ from $\overline{\mu}$ if the rate at which mutations accumulate in individuals from different classes differs [the $\mu_i(\mathbf{n})$'s].

A simple situation where the class reproductive values are not proportional to the number of genes in a class is a population with separate sexes with a constant number N_f of females and N_m of males (constancy of demography entails that we can also drop the dependence on the state $\mathbf{n}$ in all quantities). Then, we have $\gamma_f = \alpha_f = t_{mf}/(t_{mf} + t_{fm})$, where t_{ij} is the probability that a gene randomly sampled in a class-i offspring has been transmitted by a class-j parent (see *Appendix D*, Equation D17 for a proof; $t_{mf} = t_{fm} = 1/2$ for diploids, while $t_{mf} = 1$ and $t_{fm} = 1/2$ for diploid females and haploid males). With this, the effective mutation rate is

$$\mu_e = \frac{t_{mf}}{t_{mf} + t_{fm}}\mu_f + \frac{t_{fm}}{t_{mf} + t_{fm}}\mu_m, \tag{12}$$

while the average mutation rate is

$$\overline{\mu} = \frac{g_f N_f}{g_f N_f + g_m N_m}\mu_f + \frac{g_m N_m}{g_f N_f + g_m N_m}\mu_m. \tag{13}$$

Thus, if the mutation rate is different in the sexes, then μ_e and $\overline{\mu}$ can markedly differ and this stems from the fact that even if the number of females is much larger than the number of males, every offspring has both a father and a mother. Thus, each sex contributes to the gene pool in proportion to its ploidy and not according to the number of its representative individuals (*e.g.*, Taylor 1990).

Overlapping generations and generation time

Haploid reproduction: We now focus on overlapping generations under haploid reproduction, so that $i = 0, 1, 2, \ldots$ indexes the age classes. An individual will be said of age i if it is between i and $i + 1$ units of age, where $i = 0$ stands for newborns. Reproduction is assumed to occur at the end of a time period, that is, when individuals of age i have reached $i + 1$ units of age.

Following standard formulations (Pollak 1982, equations 1 and 2; Charlesworth 1994, equation 2.49b), we also assume that mutations occur only during gametogenesis and therefore only when newborns are produced. Hence, $\mu_{ij} = 0$ for $i \neq 0$, and we have $\mu_{0j} \geq 0$ for all j. From Equation 1, the probability that a newborn carries a mutation in demographic state $\mathbf{n}'$ is then given by

$$\mu_0(\mathbf{n}') = \sum_{\mathbf{n}}\sum_j a_{0j}(\mathbf{n}', \mathbf{n})\mu_{0j}, \tag{14}$$

whereby the substitution rate (Equation 8) can be written as

$$k = \sum_{\mathbf{n}} \gamma_0(\mathbf{n})\mu_0(\mathbf{n}) \ . \tag{15}$$

We can now define an effective mutation rate $\mu_{e,0}$ for newborns, by averaging the mutation rate $\mu_0(\mathbf{n})$ over the probabilities $\gamma_0(\mathbf{n})/\left[\sum_{\mathbf{s}}\gamma_0(\mathbf{s})\right]$ that the ancestral lineage of a randomly sampled gene will be in a newborn. This yields

$$\mu_{e,0} = \sum_{\mathbf{n}} \frac{\gamma_0(\mathbf{n})}{\sum_{\mathbf{s}}\gamma_0(\mathbf{s})}\mu_0(\mathbf{n}), \tag{16}$$

whereby $k = \mu_{e,0}\left[\sum_{\mathbf{s}}\gamma_0(\mathbf{s})\right]$, where the term in square brackets is related to the average generation time in the next section.

Generation time: The generation time is now introduced through the quantity

$$T(\mathbf{n}') = \sum_{\mathbf{n}} \sum_{i} (i+1) a_{0i}(\mathbf{n}', \mathbf{n}). \tag{17}$$

This is the mean age of the parent of an individual of age class zero randomly sampled in population state $\mathbf{n}'$ because $a_{0i}(\mathbf{n}', \mathbf{n})$ is the probability that a newborn sampled in population state $\mathbf{n}'$ descends from a parent of class i living in population state $\mathbf{n}$, in which case the parent has lived $i + 1$ units of time. For a population with constant size Equation 17 reduces to the equation for the generation time used in population genetics (Felsenstein 1971, p. 583; Pollak 1982, p. 91) and is a direct extension of that case to fluctuating demography.

To evaluate the average generation time, one needs to take into account the probability of sampling a newborn in a given demographic state. Hence, the average generation time is

$$T = \sum_{\mathbf{n}} \frac{\gamma_0(\mathbf{n})}{\sum_{\mathbf{s}} \gamma_0(\mathbf{s})} T(\mathbf{n}). \tag{18}$$

In *Appendix C*, it is proved that

$$\sum_{\mathbf{n}} \gamma_0(\mathbf{n}) T(\mathbf{n}) = 1, \tag{19}$$

which implies that

$$T = \frac{1}{\sum_{\mathbf{s}} \gamma_0(\mathbf{s})} \tag{20}$$

and shows that the generation time can be expressed in terms of class reproductive values.

Substitution rate in terms of generation time: From Equations 15, 16, and 20, we can now write

$$k = \frac{\mu_{\mathrm{e},0}}{T}. \tag{21}$$

Measured in units of average generation time T, the substitution rate is equal to the effective mutation rate in newborns. If the mutation rate is the same in all individuals, then $\mu_{0j} = \mu$ in which case $\mu_{\mathrm{e},0} = \mu$. In this case, when measured in units of average generation time, the substitution rate is equal to the mutation rate, and this then removes the effect of population size on the substitution rate observed in previous stochastic models of neutral evolution (Balloux and Lehmann 2012), a point that was suggested by Lanfear *et al.* (2014).

This result (Equation 21) also applies to a population with separate sexes (diploid or haplodiploid), in which case the summations in $\mu_{\mathrm{e},0}$ and T must also be taken over the different sexes (*Appendix D*, Equation D4). For a constant demography, the effective mutation rate in newborns can then be simplified to

$$\mu_{\mathrm{e},0} = \frac{t_{\mathrm{mf}}}{t_{\mathrm{mf}} + t_{\mathrm{fm}}} \mu_{\mathrm{f}0} + \frac{t_{\mathrm{fm}}}{t_{\mathrm{mf}} + t_{\mathrm{fm}}} \mu_{\mathrm{m}0}, \tag{22}$$

which is of the same form as Equation 12 and where μ_{g0} denotes the rate at which a gene randomly sampled in a newborn of sex $g \in \{\mathrm{f}, \mathrm{m}\}$ accumulates mutations. With this, k (Equation 21) reduces to the standard expression for the neutral substitution rate under overlapping generations for diploid reproduction (Pollak 1982, equation 8) (see *Appendix D* for the connection). This shows that the average mutation rate in newborns from previous models is the effective mutation in newborns, where the sex-specific mutation rate is weighted by conditional reproductive values, which give the proportion of time a gene lineage will spend in males and females given that it resides in a newborn.

Discussion

The rate k of neutral allelic substitution at a given locus has been analyzed under the assumptions that the evolving population can be structured into classes (*e.g.*, by sex, age, geography, etc.) and that it follows a stationary stochastic demography. Two main results were found. First, the substitution rate is equal to the effective mutation rate: $k = \mu_{\mathrm{e}}$, which is the rate at which a randomly sampled gene from a randomly sampled common ancestor of the population accumulates mutations. Second, in the presence of overlapping generations, the substitution rate can also be expressed as the product $k = \mu_{\mathrm{e},0}/T$, where $\mu_{\mathrm{e},0}$ is the effective mutation rate in newborns and T is the average generation time (or the mean age of the parents of a newborn), which itself can be expressed entirely in terms of class reproductive values (Equation 20).

These results entail that the effective mutation rate μ_{e} (Equation 8) is not necessarily equal to the rate $\overline{\mu}$ at which a randomly sampled gene in a randomly sampled individual from the population accumulates mutations (Equation 2). Care must thus be taken in interpreting k as a mutation rate. For instance, it is sometimes said that k is equal to the mutation rate under almost any conceivable complication (Lanfear *et al.* 2014, box 3). But if this mutation rate is supposed to be that in a randomly sampled gamete, as in the original formulation of the neutral substitution process (Kimura 1971), then the substitution rate can be different, unless mutations occurs at the same rate in every class of individuals (Equation 5). But if mutations occur with different magnitude in different classes of individuals, then even if k is measured in units of generation time, it can be different from the rate at which a randomly sampled gene in a newborn accumulates mutations. This occurs because the effective mutation rate in newborns ($\mu_{\mathrm{e},0}$) is a reproductive value weighted average of the rate at which newborns accumulate mutations (Equation 16), which thus weights different classes of individuals according to their asymptotic contribution to the gene pool. Different classes of newborns may be obtained when individuals are born under different types of demographic or environmental conditions (or are of different sexes) and where mutation rates can vary during gametogenesis, owing, for instance, to the fact that irradiation can vary in space and time.

In summary, with uniform mutation across classes, the substitution rate is the mutation rate under stochastic demography and class structure. With nonuniform mutation, how different classes of individuals contribute to the ancestry of the population can matter (*e.g.*, Equations 12 and 13), in which case the substitution rate is the average rate at which the ancestors of the population accumulate mutations. This then justifies the interpretation of the substitution rate as the effective mutation rate.

Acknowledgments

I thank F. Balloux and especially C. Mullon for useful discussions. The two reviewers of this article provided constructive comments that allowed me to improve it; many thanks to them. This work was supported by Swiss National Science Foundation grant PP00P3-123344.

Literature Cited

Balloux, F., and L. Lehmann, 2012 Substitution rates at neutral genes depend on population size under fluctuating demography and overlapping generations. Evolution 66: 605–611.

Charlesworth, B., 1994 *Evolution in Age-Structured Populations*, Ed. 2. Cambridge University Press, Cambridge, UK.

Emigh, T. H., and E. Pollak, 1979 Fixation probabilities and effective population numbers in diploid populations with overlapping generations. Theor. Popul. Biol. 15: 86–107.

Felsenstein, J., 1971 Inbreeding and variance effective numbers in populations with overlapping generations. Genetics 68: 581–597.

Grinstead, C. M., and J. L. Snell, 1997 *Introduction to Probability*, Ed. 2. American Mathematical Society, Providence, RI.

Hill, W., 1972 Probability of fixation of genes in populations of variable size. Theor. Popul. Biol. 3: 27–40.

Karlin, S., and H. M. Taylor, 1975 *A First Course in Stochastic Processes*. Academic Press, San Diego.

Kimura, M., 1969 The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics 61: 893–903.

Kimura, M., 1971 Theoretical foundation of population genetics at the molecular level. Theor. Popul. Biol. 2: 174–208.

Kimura, M., and J. F. Crow, 1964 The number of alleles that can be maintained in a finite population. Genetics 49: 725–738.

Lanfear, R., H. Kokko, and A. Eyre-Walker, 2014 Population size and the rate of evolution. Trends Ecol. Evol. 29: 33–40.

Lehmann, L., 2012 The stationary distribution of a continuously varying strategy in a class-structured population under mutation-selection-drift balance. J. Evol. Biol. 25: 770–787.

Leturque, H., and F. Rousset, 2002 Dispersal, kin competition, and the ideal free distribution in a spatially heterogeneous population. Theor. Popul. Biol. 62: 169–180.

Nagylaki, T., 1998 The expected number of heterozygous sites in a subdivided population. Genetics 149: 1599–1604.

Pollak, E., 1982 The rate of mutant substitution in populations with overlapping generations. Genet. Res. 40: 89–94.

Rousset, F., 2004 *Genetic Structure and Selection in Subdivided Populations*. Princeton University Press, Princeton, NJ.

Taylor, P., 1990 Allele-frequency change in a class-structured population. Am. Nat. 135: 95–106.

Wakeley, J., 2008 *Coalescent Theory: An Introduction*. Roberts and Company, Greenwood Village, CO.

Whitlock, M. C., and N. H. Barton, 1997 The effective size of a subdivided population. Genetics 146: 427–441.

Communicating editor: L. M. Wahl

Appendix A

Neutral Substitution Rate

Here, we provide a proof of Equation 3. To that end, we first note that the neutral substitution rate can be expressed as the expectation over all demographic states of the expected number $N_{\mathrm{Fix}}(\mathbf{n})$ of mutants produced per unit time in a population in state $\mathbf{n}$ and that will ultimately fix in the population,

$$k = \sum_{\mathbf{n}} N_{\mathrm{Fix}}(\mathbf{n})\Pr(\mathbf{n}), \tag{A1}$$

where

$$N_{\mathrm{Fix}}(\mathbf{n}) = \sum_{\mathbf{n}'}\sum_{i}\sum_{j} \pi_i(\mathbf{n}')w_{ij}(\mathbf{n}', \mathbf{n})g_i t_{ij}\mu_{ij}N_j \Pr(\mathbf{n}'|\mathbf{n}) \tag{A2}$$

(Lehmann 2012, equations A1 and A2, assuming no selection). Here, $\pi_i(\mathbf{n}')$ is the fixation probability of a mutation arising as a single copy in a class-i individual in population state $\mathbf{n}'$, $w_{ij}(\mathbf{n}', \mathbf{n})$ is the expected number of class-i individuals in population state $\mathbf{n}'$ produced by a single class-j individual in population state $\mathbf{n}$ (the fitness of an individual of class j through class-i offspring), g_i is the ploidy of a class-i individual, t_{ij} is the probability that a gene randomly sampled in a class-i offspring has been transmitted by a class-j individual, and N_j is the number of class-j individuals in population state $\mathbf{n}$. Hence, $w_{ij}(\mathbf{n}', \mathbf{n})g_i t_{ij}\mu_{ij}N_j$ gives the number of mutations in population state $\mathbf{n}'$ when the population was in state $\mathbf{n}$ in the previous time step.

The probability $a_{ij}(\mathbf{n}', \mathbf{n})$ that a gene randomly sampled in a class-i individual in population state $\mathbf{n}'$ descends from a class-j individual and in population state $\mathbf{n}$ in the previous time step can now be written as

$$a_{ij}(\mathbf{n}',\mathbf{n}) = f_{ij}(\mathbf{n}',\mathbf{n})\frac{\Pr(\mathbf{n}'|\mathbf{n})\Pr(\mathbf{n})}{\Pr(\mathbf{n}')}. \tag{A3}$$

The second term is the backward transition probability that a population in state $\mathbf{n}'$ derives from a population in state $\mathbf{n}$ one time step earlier, while

$$f_{ij}(\mathbf{n}',\mathbf{n}) = \frac{w_{ij}(\mathbf{n}',\mathbf{n})t_{ij}N_j}{N_i} \tag{A4}$$

is the probability that a class-i individual descends from a class-j individual given state $\mathbf{n}'$ in the offspring generation and state $\mathbf{n}$ in the parental generation (Charlesworth 1994, p. 81; Rousset 2004, equation 11.2; Lehmann 2012, equation A15), which is obtained as the ratio of the number of genes in class i descending from class j to the total number of genes in class i.

Substituting Equation A3 into Equation A2 yields

$$\begin{aligned} k &= \sum_{\mathbf{n}}\sum_{\mathbf{n}'}\sum_{i}\sum_{j}\pi_i(\mathbf{n}')g_iN_ia_{ij}(\mathbf{n}',\mathbf{n})\mu_{ij}\Pr(\mathbf{n}') \\ &= \sum_{\mathbf{n}'}\sum_{i}\pi_i(\mathbf{n}')\mu_i(\mathbf{n}')g_iN_i\Pr(\mathbf{n}') \\ &= \overline{N}\underbrace{\sum_{\mathbf{n}}\sum_{j}\frac{\mu_j(\mathbf{n})g_jN_j\Pr(\mathbf{n})}{\overline{N}}}_{\overline{\mu}}\underbrace{\sum_{\mathbf{n}'}\sum_{i}\pi_i(\mathbf{n}')\frac{\mu_i(\mathbf{n}')g_iN_i\Pr(\mathbf{n}')}{\sum_{\mathbf{n}}\sum_{j}\mu_j(\mathbf{n})g_jN_j\Pr(\mathbf{n})}}_{\overline{\pi}}, \end{aligned} \tag{A5}$$

which displays k as in Equation 3.

Appendix B

Fixation Probability and Reproductive Value

Here, we provide a proof of Equation 7. To that end, we first write the recursion at steady state for the γ reproductive values, which from Equation 6 reads

$$\gamma_j(\mathbf{n}) = \sum_{\mathbf{n}'}\sum_{i}\gamma_i(\mathbf{n}')a_{ij}(\mathbf{n}',\mathbf{n}). \tag{B1}$$

Let us now denote by $p_{i,t}(\mathbf{n})$ the expectation (over replicates of the evolutionary process) of the average frequency of a mutant allele in class i at time t, conditional on the demographic state being $\mathbf{n}$ at that time and conditional on some initial mutant frequency distribution at $t = 0$. This satisfies the recursion

$$p_{i,t+1}(\mathbf{n}') = \sum_{\mathbf{n}}\sum_{j}a_{ij}(\mathbf{n}',\mathbf{n})p_{j,t}(\mathbf{n}). \tag{B2}$$

We can collect the $p_{i,t}(\mathbf{n})$ elements into a vector $\mathbf{p}_t$, which satisfies the recursion

$$\mathbf{p}_{t+1} = \mathbf{A}\mathbf{p}_t, \tag{B3}$$

where $\mathbf{A}$ is the row stochastic transition matrix collecting the $a_{ij}(\mathbf{n}', \mathbf{n})$ elements $\left[\sum_{\mathbf{n}}\sum_{j}a_{ij}(\mathbf{n}',\mathbf{n}) = 1\right]$.

Since the fixation probability of a mutant is its asymptotic frequency (*e.g.*, Hill 1972; Emigh and Pollak 1979; Rousset 2004), the fixation probability in a given class i, conditional on the demographic state being $\mathbf{n}$, is $\lim_{t\to\infty}p_{i,t}(\mathbf{n})$. The vector $\boldsymbol{\pi}$ of fixation probabilities in each class is then given by $\boldsymbol{\pi} = \lim_{t\to\infty}\mathbf{A}^t\mathbf{p}_0$, where $\mathbf{p}_0$ is the initial mutant frequency distribution across classes. Note that each element of the vector $\boldsymbol{\pi}$ will be the same, as the mutant either fixes in the total population and thus in each class or goes extinct. By standard results for finite ergodic Markov chains, each row of $\lim_{t\to\infty}\mathbf{A}^t$ is equal to the left unit eigenvector of the transition matrix $\mathbf{A}$ (Grinstead and Snell 1997), which is precisely the vector satisfying the system of recursions displayed in Equation B1: $\boldsymbol{\gamma} = \boldsymbol{\gamma}\mathbf{A}$. Hence, using $\gamma_j(\mathbf{n}) = \alpha_j(\mathbf{n})\Pr(\mathbf{n})$, the fixation probability of a mutant given initial frequency distribution $\mathbf{p}_0$ is

$$\sum_{\mathbf{n}}\sum_{j}\gamma_j(\mathbf{n})p_{j,0}(\mathbf{n})=\sum_{\mathbf{n}}\sum_{j}\alpha_j(\mathbf{n})\Pr(\mathbf{n})p_{j,0}(\mathbf{n}). \tag{B4}$$

When a single mutant arises initially in class i and in population state $\mathbf{n}$, this further reduces to $\alpha_i(\mathbf{n})p_{i,0}(\mathbf{n})$ (since with probability 1 the state is $\mathbf{n}$), which is precisely Equation 7 since $p_{i,0}(\mathbf{n}) = 1/[g_iN_i]$. This latter equation was also given in Leturque and Rousset (2002, p. 178) for a class-structured population of constant size.

Appendix C

Reproductive Value and Generation Time

Here, we provide a proof of Equation 20. For a haploid age-structured population, with age classes $i = 0, 1, 2, \ldots$, Equation B1 reduces to

$$\gamma_j(\mathbf{n})=\sum_{\mathbf{n}'}\Big[\gamma_0(\mathbf{n}')a_{0j}(\mathbf{n}',\mathbf{n})+\gamma_{j+1}(\mathbf{n}')a_{j+1j}(\mathbf{n}',\mathbf{n})\Big], \tag{C1}$$

where the first term in brackets is the contribution to the gene pool of class $(j, \mathbf{n})$ through production of newborns, while the second term is the contribution through survival. For an age-structured population, we also have

$$f_{j+1j}(\mathbf{n}',\mathbf{n})=\frac{w_{j+1j}(\mathbf{n}',\mathbf{n})N_j}{N_{j+1}}=1, \tag{C2}$$

owing to the fact that $w_{j+1j}(\mathbf{n}', \mathbf{n})$ is the survival of an individual of class j given demographic states $\mathbf{n}'$ and $\mathbf{n}$, so that necessarily $N_{j+1} = w_{j+1j}(\mathbf{n}', \mathbf{n})N_j$. This entails that $a_{j+1j}(\mathbf{n}', \mathbf{n}) = \Pr(\mathbf{n}'|\mathbf{n})\Pr(\mathbf{n})/\Pr(\mathbf{n}')$, whereby $\sum_{\mathbf{n}}a_{j+1j}(\mathbf{n}', \mathbf{n}) = \sum_{\mathbf{n}}\Pr(\mathbf{n}'|\mathbf{n})\Pr(\mathbf{n})/\Pr(\mathbf{n}') = \sum_{\mathbf{n}}\Pr(\mathbf{n}', \mathbf{n})/\Pr(\mathbf{n}') = 1$. Summing both sides of Equation C1 over $\sum_{\mathbf{n}}\sum_j(j + 1)$ then yields

$$\sum_{\mathbf{n}}\sum_{j}(j+1)\gamma_j(\mathbf{n})=\sum_{\mathbf{n}'}\Big[\gamma_0(\mathbf{n}')\sum_{\mathbf{n}}\sum_{j}(j+1)a_{0j}(\mathbf{n}',\mathbf{n})+\sum_{j}(j+1)\gamma_{j+1}(\mathbf{n}')\Big]. \tag{C3}$$

Using the definition of the generation time (Equation 17) and reindexing the last sum, we have

$$1+\sum_{\mathbf{n}}\sum_{j}j\gamma_j(\mathbf{n})=\sum_{\mathbf{n}'}\gamma_0(\mathbf{n}')T(\mathbf{n}')+\sum_{\mathbf{n}'}\sum_{h}h\gamma_h(\mathbf{n}'), \tag{C4}$$

whereby

$$\sum_{\mathbf{n}'}\gamma_0(\mathbf{n}')T(\mathbf{n}')=1. \tag{C5}$$

Appendix D

Separate Sexes

Stochastic demography

Here, we extend the result $k = \mu_{e,0}/T$ found for haploid reproduction (Equation 21) to a population with both males and females, which could be diploid, haplodiploid, or subject to other modes of ploidy. The coefficient $a_{0j}(\mathbf{n}', \mathbf{n})$ for haploids appearing in the previous section is now written as $a_{g0,g'j}(\mathbf{n}', \mathbf{n})$, which stems from the probability that a gene randomly sampled in a newborn of sex $g \in \{f, m\}$ in state $\mathbf{n}'$ descends from an individual of sex $g' \in \{f, m\}$ of age j and that was in state $\mathbf{n}$. Likewise, $\mu_{g0,g'j}$ denotes the mutation rate of a gene in an individual of sex g' of age j that is transmitted to a newborn of sex g.

With these definitions, the substitution rate (Equation 8) becomes

$$k=\sum_{\mathbf{n}}\sum_{g}\gamma_{g0}(\mathbf{n})\mu_{g0}(\mathbf{n}), \tag{D1}$$

where the sex-specific mutation probability in newborns is

$$\mu_{g'0}(\mathbf{n}') = \sum_{\mathbf{n}}\sum_{g}\sum_{j} a_{g'0,gj}(\mathbf{n}',\mathbf{n})\mu_{g'0,gj}. \tag{D2}$$

We can define the effective mutation rate among newborns as

$$\mu_{e,0} = \sum_{\mathbf{n}}\sum_{g} \frac{\gamma_{g0}(\mathbf{n})}{\sum_{\mathbf{s}}\sum_{y}\gamma_{y0}(\mathbf{s})}\mu_{g0}(\mathbf{n}), \tag{D3}$$

which generalizes Equation 14. Likewise, using the same weights, we can define the average generation time

$$T = \sum_{\mathbf{n}}\sum_{g} \frac{\gamma_{g0}(\mathbf{n})}{\sum_{\mathbf{s}}\sum_{y}\gamma_{y0}(\mathbf{s})}T_g(\mathbf{n}), \tag{D4}$$

which generalizes Equation 18 and where

$$T_g(\mathbf{n}') = \sum_{\mathbf{n}}\sum_{g'}\sum_{j}(j+1)a_{g0,g'j}(\mathbf{n}',\mathbf{n}). \tag{D5}$$

Using Equations D1–D5, we can write

$$k = \frac{\mu_{e,0}}{T}, \tag{D6}$$

provided that $T = 1/\left[\sum_{\mathbf{s}}\sum_{g}\gamma_{g0}(\mathbf{s})\right]$, which requires

$$\sum_{\mathbf{n}}\sum_{g}\gamma_{g0}(\mathbf{n})T_g(\mathbf{n}) = 1. \tag{D7}$$

We now prove this latter equality by applying the same argument as in the previous section. Namely, the reproductive value $\gamma_{gj}(\mathbf{n})$ satisfies the recursion

$$\gamma_{gj}(\mathbf{n}) = \sum_{\mathbf{n}'}\left[\left(\sum_{g'}\gamma_{g'0}(\mathbf{n}')a_{g'0,gj}(\mathbf{n}',\mathbf{n})\right) + \gamma_{gj+1}(\mathbf{n}')a_{gj+1,gj}(\mathbf{n}',\mathbf{n})\right], \tag{D8}$$

where

$$a_{gj+1,gj}(\mathbf{n}',\mathbf{n}) = \frac{w_{gj+1,gj}(\mathbf{n}',\mathbf{n})t_{gg}N_{gj}(\mathbf{n})}{N_{gj+1}(\mathbf{n}')}\frac{\Pr(\mathbf{n}'|\mathbf{n})\Pr(\mathbf{n})}{\Pr(\mathbf{n}')} = \frac{\Pr(\mathbf{n}'|\mathbf{n})\Pr(\mathbf{n})}{\Pr(\mathbf{n}')}. \tag{D9}$$

The last equality follows from the fact that necessarily $N_{gj+1} = w_{gj+1,gj}(\mathbf{n}', \mathbf{n})N_{gj}$ and $t_{gg} = 1$, since during survival no segregation of alleles occurs. If we now take the sum $\sum_{\mathbf{n}}\sum_{g}\sum_{j}(j+1)$ over Equation D8, we obtain

$$\begin{aligned}&\sum_{\mathbf{n}}\sum_{g}\sum_{j}(j+1)\gamma_{gj}(\mathbf{n})\\ &= \sum_{\mathbf{n}'}\left[\sum_{g'}\gamma_{g'0}(\mathbf{n}')\sum_{\mathbf{n}}\sum_{g}\sum_{j}(j+1)a_{g'0,gj}(\mathbf{n}',\mathbf{n}) + \sum_{\mathbf{n}}\sum_{g}\sum_{j}(j+1)\gamma_{gj+1}(\mathbf{n}')a_{gj+1,gj}(\mathbf{n}',\mathbf{n})\right],\end{aligned} \tag{D10}$$

and by reindexing the last sum, this can be written as

$$1 + \sum_{\mathbf{n}}\sum_{g}\sum_{j}j\gamma_{gj}(\mathbf{n}) = \sum_{\mathbf{n}'}\sum_{g'}\gamma_{g'0}(\mathbf{n}')T_{g'}(\mathbf{n}') + \sum_{\mathbf{n}'}\sum_{g}\sum_{h}h\gamma_{gh}(\mathbf{n}'), \tag{D11}$$

which shows that Equation D7 holds.

Constant demography

In a population of constant size, the effective mutation rate in newborns and the average generation time can be written

$$\begin{aligned}\mu_{\mathrm{e},0} &= \gamma^*_{\mathrm{f}0}\mu_{\mathrm{f}0} + (1-\gamma^*_{\mathrm{f}0})\mu_{\mathrm{m}0}\\ T &= \gamma^*_{\mathrm{f}0}T_{\mathrm{f}} + (1-\gamma^*_{\mathrm{f}0})T_{\mathrm{m}},\end{aligned} \tag{D12}$$

where

$$\begin{aligned}\mu_{g'0} &= \sum_g\sum_j a_{g'0,gj}\mu_{g'0,gj}\\ T_{g'} &= \sum_g\sum_j (j+1)a_{g'0,gj},\end{aligned} \tag{D13}$$

and

$$\gamma^*_{\mathrm{f}0} = \frac{\gamma_{\mathrm{f}0}}{\gamma_{\mathrm{f}0}+\gamma_{\mathrm{m}0}} \tag{D14}$$

is the probability that a gene lineage is in a female, given that it has been sampled in a newborn. This probability satisfies the recursion

$$\gamma^*_{\mathrm{f}0} = (1-t_{\mathrm{fm}})\gamma^*_{\mathrm{f}0} + t_{\mathrm{mf}}(1-\gamma^*_{\mathrm{f}0}), \tag{D15}$$

since, given that a gene lineage is a newborn female, it descends from a female with probability $t_{\mathrm{ff}} = 1 - t_{\mathrm{fm}}$ and, given that a gene lineage is a newborn male, it descends from a female with probability t_{mf}. At steady state, we have

$$\gamma^*_{\mathrm{f}0} = \frac{t_{\mathrm{mf}}}{t_{\mathrm{mf}}+t_{\mathrm{fm}}}. \tag{D16}$$

Assuming diploid reproduction, $t_{\mathrm{mf}} = t_{\mathrm{fm}} = 1/2$, whereby $\gamma^*_{\mathrm{f}0} = 1/2$ and the expression for the average generation time (Equation D13) can then be seen to be precisely the average of the mean age of the parents of a newborn defined by Pollak (1982, p. 91), since $a_{\mathrm{f}0,gj}$ ($a_{\mathrm{m}0,gj}$) corresponds to $2p^g_{j.}$ ($2p^g_{.j}$) in the notations of Pollak (1982). Likewise, $\mu_{\mathrm{e},0}$ in this case is precisely $\bar{v}_g$ in the notations of Pollak (1982, equations 1 and 2). With all this, we see that in a diploid population of constant size $k = \mu_{\mathrm{e},0}/T$ reduces precisely to equation 8 of Pollak (1982).

Semelparous populations

Here, we evaluate the class reproductive values and the fixation probability in a population of constant size in the absence of overlapping generations. Since, in this case, $w_{\mathrm{ff}} = 1$ and $w_{\mathrm{mf}} = N_{\mathrm{m}}/N_{\mathrm{f}}$, we have from Eqation A4 that $f_{\mathrm{ff}} = t_{\mathrm{ff}} = 1 - t_{\mathrm{fm}}$ and $f_{\mathrm{mf}} = t_{\mathrm{mf}}$. Then from Equations A3 and B1, the reproductive value of the female class satisfies

$$\gamma_{\mathrm{f}} = (1-t_{\mathrm{fm}})\gamma_{\mathrm{f}} + t_{\mathrm{mf}}(1-\gamma_{\mathrm{f}}), \tag{D17}$$

whereby

$$\gamma_{\mathrm{f}} = \frac{t_{\mathrm{mf}}}{t_{\mathrm{mf}}+t_{\mathrm{fm}}}. \tag{D18}$$

With this, noting that $\gamma_i = \alpha_i$ (population size is constant) and using Equation B4, the fixation probability of a single mutant allele that arises in females and males is, respectively, given by

$$\pi_{\mathrm{f}} = \frac{1}{g_{\mathrm{f}}N_{\mathrm{f}}}\left(\frac{t_{\mathrm{mf}}}{t_{\mathrm{mf}}+t_{\mathrm{fm}}}\right) \quad \text{and} \quad \pi_{\mathrm{m}} = \frac{1}{g_{\mathrm{m}}N_{\mathrm{m}}}\left(\frac{t_{\mathrm{fm}}}{t_{\mathrm{mf}}+t_{\mathrm{fm}}}\right). \tag{D19}$$