ELSEVIER

Gene 205 (1997) 269–278

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

Codon bias evolution in *Drosophila*. Population genetics of mutation–selection drift

Hiroshi Akashi *

Section of Evolution and Ecology, University of California, Davis, CA 95616, USA

Accepted 2 July 1997

Abstract

Although non-random patterns of synonymous codon usage are a prominent feature in the genomes of many organisms, the relatives roles of mutational biases and natural selection in maintaining codon bias remain a contentious issue. In some species, patterns of codon bias and empirical findings on the biology of translation suggest 'major codon preference', a balance among mutation pressure, genetic drift, and weak selection in favor of translationally superior codons. Population genetics theory makes testable predictions to distinguish such a model from a strictly mutational model of codon bias. Major codon preference predicts two fitness classes of synonymous DNA changes: 'preferred' mutations from non-major to major codons and 'unpreferred' changes in the opposite direction. An extension of current statistical methods is employed to reveal differences in the within and between species dynamics of preferred and unpreferred silent mutations in *Drosophila simulans*. In this lineage, codon bias appears to be maintained under roughly equal magnitudes of natural selection and genetic drift. In the sibling species, *D. melanogaster*, however, a reduction in $N_e s$, the product of effective population size and selection coefficient, appears to have allowed a genome-wide reduction in codon bias. © 1997 Elsevier Science B.V.

Keywords: Codon usage; Molecular evolution; Weak selection; Nearly neutral theory

1. Introduction

In *Escherichia coli* and *Saccharomyces cerevisiae*, synonymous codon usage may enhance the efficiency of protein synthesis [reviewed in Andersson and Kurland (1990) and Sharp et al. (1993)]. Codon usage in these organisms is biased toward 'major' codons that generally encode the most abundant tRNA(s) for each amino acid. Among codons recognized by the same tRNA, the codon that forms the natural Watson–Crick pairing with the tRNA anticodon is generally favored (Ikemura, 1981, 1982; Bennetzen and Hall, 1982; Grosjean and Fiers, 1982). During polypeptide chain elongation in *E. coli*, the arrival time of a cognate tRNA is inversely proportional to its abundance (Varenne et al., 1984; Curran and Yarus, 1989). Major codons may confer fitness benefits by enhancing translational elongation rates, by lowering the energetic cost of proofreading (rejecting non-cognate tRNAs), or by reducing the rate of amino acid misincorporation (Bulmer, 1988). In *E. coli* and yeast, the degree to which codon usage is biased varies among genes and correlates strongly with protein abundance (Bennetzen and Hall, 1982; Gouy and Gautier, 1982; Ikemura, 1985; Sharp and Cowe, 1991). Because the fitness advantage to encoding a translationally superior codon will be a function of the rate at which it is translated, selection intensity for codon bias should be stronger in highly expressed genes. Finally, silent DNA divergence between *E. coli* and *Salmonella typhemurium* is inversely related to codon usage bias, consistent with a relationship between the strength of selection and levels of codon bias [Sharp and Li (1987) and Berg and Martelius (1995); but see Eyre-Walker and Bulmer (1995) for contrary evidence].

Patterns of codon usage and silent DNA evolution in *D. melanogaster* appear to be similar to those found in *E. coli* and yeast (Shields et al., 1988; Sharp and Li, 1989). Codon usage is biased toward a subset of synonymous codons for each amino acid. In multicellular animals, tRNA abundances and gene expression levels can be tissue- and developmental-stage-specific and are thus difficult to quantify. However, major codons correspond to abundant tRNA's for the three amino acids

* Tel: +1 916 7524253; Fax: +1 916 7521449;
e-mail: hakashi@ucdavis.edu

PII S0378-1119(97)00400-9

for which data are available (Shields et al., 1988). In addition, anecdotal evidence suggests a relationship between expression levels and codon bias; highly expressed genes encoding ribosomal proteins and glycolytic enzymes show greater codon bias than genes with limited or low expression such as those encoding developmental regulatory proteins. Similarly to prokaryotes, silent divergence between *Drosophila* species is inversely related to codon usage bias (Sharp and Li, 1989; Carulli et al., 1993). Finally, an association between codon bias and recombination rate in the *D. melanogaster* genome is consistent with theoretical predictions that the efficacy of natural selection is a function of the rate of genetic exchange (Kliman and Hey, 1993).

The major codon preference model is a form of mutation–selection drift at silent sites (Sharp and Li, 1986; Bulmer, 1988). Major codons confer fitness benefits, but the magnitude of selection is small enough that non-major codons persist through mutation pressure and genetic drift. Such a model can be distinguished from a simple mutational model of codon bias by contrasting the evolutionary dynamics of classes of silent changes; major codon preference is maintained by positive selection for preferred mutations from non-major to major codons and selection against unpreferred mutations in the opposite direction (Akashi, 1995). The following sections will first examine mutational and selection models of codon bias evolution. Then, patterns of silent DNA divergence will be compared in the sibling species *Drosophila melanogaster* and *D. simulans* to determine whether codon bias has evolved at a steady state. Finally, population genetic analyses will be employed to confirm the role of selection in maintaining codon bias in the *D. simulans* lineage and to identify a reduction in selection intensity as the cause of a genome-wide decline of codon bias in *D. melanogaster*.

2. Quantitative models of codon bias evolution

The simplest model of major codon preference considers two-fold redundant codons in a haploid organism (Li, 1987; Bulmer, 1991). Mutations occur at rates v from non-major codons to major codons (preferred mutations) and u in the opposite direction (unpreferred mutations). Major codons confer selective advantage, s. This scenario is depicted in Eq. (1), where AAG is a major codon and AAA is a non-major codon.

$$\underset{+s}{\mathrm{AAG}} \overset{u}{\underset{v}{\rightleftarrows}} \underset{-s}{\mathrm{AAA}}. \quad (1)$$

Consider a 'locus' consisting of a number of such sites. The frequency of major codons at the locus is determined by the ratio of the mutation rates, u/v, and the product of effective population size and selection coefficient, $N_e s$. If these parameters remain relatively constant, then the frequency of major codons at the locus will reach a steady-state (the numbers of unpreferred and preferred substitutions will be equal).

Directional mutation pressure can bias equilibrium base composition in the absence of fitness differences (Freese, 1962; Sueoka, 1962, 1988). Under such a model ($s=0$), the equilibrium codon bias depends only on the ratio of the forward and backward mutation rates.

$$q = \frac{v}{u+v}. \quad (2)$$

At equilibrium, the per locus mutation rates in the forward and backward directions will be equal, and the evolutionary dynamics of both classes of mutations will be governed solely by genetic drift.

Under major codon preference, natural selection favors translationally superior codons for each amino acid ($s>0$), but selection is sufficiently weak so that mutation pressure and genetic drift allow minor codons to persist ($N_e s \approx 1$). In this case, the equilibrium frequency of major codons is a function of both u/v and $N_e s$. Setting the forward and backward substitution rates equal, such that $N_e u << 1$ and $N_e v << 1$, gives an expression for the steady-state proportion of major codons in a given gene (Li, 1987; Bulmer, 1991)

$$q = \frac{e^{2N_e s}}{e^{2N_e s} + u/v}. \quad (3)$$

$N_e s$ is assumed to be constant within a given gene and evolution at all sites is independent (no genetic linkage and independent effects on fitness). Note that the ratio of the mutation rates, rather than their absolute values, determines q. Eq. (3) also applies to diploid organisms under semi-dominant fitness effects by replacing $2N_e s$ with $4N_e s$. Although mutational biases and natural selection can produce similar patterns of codon bias, only major codon preference predicts that deterministic forces differentiate the evolutionary trajectories of preferred and unpreferred synonymous mutations.

3. Testing steady-state codon bias evolution in Drosophila

Comparing the evolutionary behavior of preferred and unpreferred mutations requires both the identification of candidates for major codons and inference of the direction of mutations (ancestral and derived states) in DNA. In *Drosophila*, both regional mutational biases and variation in selection intensity appear to contribute to variation in synonymous codon usage bias. Correlations between intron and silent-position base composition suggest that regional mutational biases explain about 10% of codon bias variation in *D. melanogaster* (Kliman and Hey, 1994). However, the G+C content of coding regions is higher than that of associ-

ated introns in 154 of 155 genes in the Kliman and Hey study; selection for codon bias may play a prominent role in determining base composition at silent sites.

Although tRNA abundances have not been quantified in *Drosophila*, candidates for major codons can be identified by examining variation in codon bias among genes (Akashi, 1995). Under major codon preference, selection intensity at silent sites is stronger in highly expressed genes than in low-expression loci. Major codons can be identified as those whose frequencies (within a synonymous family) show a positive correlation with the degree of bias at other codons in the same gene. Fig. 1 shows an example of such a relationship, which identifies AAG as a major codon for lysine. Candidates for major codons were identified for each amino acid in a similar manner. In *D. melanogaster*, at least one codon in each synonymous family shows a significant positive slope and all major codons are G- or C-ending [see Akashi (1995) for table of major codons].

Codon bias evolution was compared between the sibling species, *D. melanogaster* and *D. simulans*, since their split from a common ancestor. These species are morphologically almost indistinguishable (except for male genitalia) and are thought to have originated in Africa and become cosmopolitan in historical time (Lachaise et al., 1988). All genes for which multiple alleles have been sequenced in both species and at least one outgroup from within the *D. melanogaster* subgroup were examined. A total of eight genes, for which six alleles have been sequenced within *D. melanogaster* and five alleles have been sequenced in *D. simulans*, were available in the GenBank/EMBL databases or in the literature (Table 1). Parsimony assumptions and outgroup sequences were used to infer the lineage in which silent mutations have occurred (Fig. 2). For a given codon encoding different nucleotides in *D. melanogaster* and *simulans*, substitutions were assigned to minimize the number of changes in the phylogenetic tree.

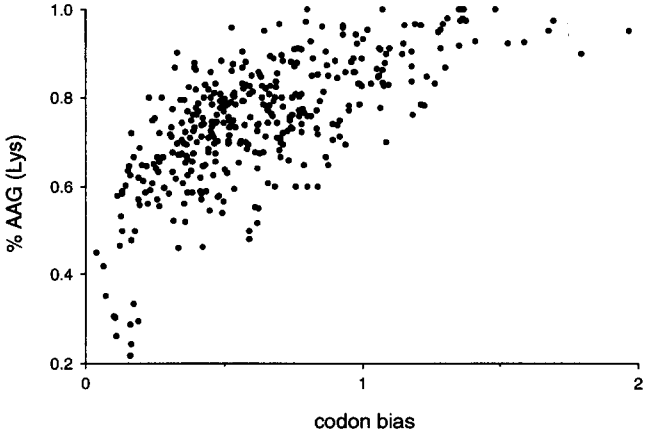

Fig. 1. Codon usage for lysine in *D. melanogaster*. Frequencies of AAG within lysine codons (AAA and AAG) are plotted against a measure of codon bias [from Akashi (1995)]. Data are shown for loci with a minimum of 20 lysine codons among 575 *D. melanogaster* genes drawn from GenBank (Akashi, 1996a). The 'scaled' χ^2 (Shields et al., 1988) was used to measure the level of bias in a given gene. χ^2 values for deviations from an A+T content of 60%, the average base content of *D. melanogaster* introns (Shields et al., 1988; Moriyama and Hartl, 1993) were calculated for each synonymous family. The sum of the χ^2 values was divided by the total number of codons in a gene to give a measure of codon bias independent of gene length, referred to as 'scaled' χ^2 (AT60%). Among 575 *D. melanogaster* genes, this measure ranges from 0.04 to 1.9 (Akashi, 1996a).

Patterns of silent divergence were examined in *D. melanogaster* and *D. simulans* to test whether major codon usage has remained stable in these lineages (Table 1). In the eight genes examined, preferred and unpreferred mutations have substituted at approximately equal rates in *D. simulans*; the null hypothesis of steady-state codon bias is not rejected (*G*-test for goodness of fit with Williams' correction, $G=0.04$, $P=0.84$). In *D. melanogaster*, however, seven of the eight genes examined show an excess of unpreferred fixations. Summing over all loci, unpreferred fixations outnumber preferred fixations by over 10-fold ($G=51.1$, $P<10^{-5}$); major codon usage has undergone a dramatic, and apparently genome-wide, reduction in *D. melanogaster*. The following sections will employ population genetics theory to attempt to identify both the evolutionary forces maintaining steady-state codon bias in the

Table 1
Synonymous fixations in *D. melanogaster* and *D. simulans*

Symbol	Gene	*mel* ($m=6$)		*sim* ($m=5$)	
		Unpref	Pref	Unpref	Pref
Adh	Alcohol dehydrogenase	1	1	0	0
*Adh*r	Adh-related	5	1	1	1
boss	Bride of sevenless	7	0	1	1
Mlc1	Myosin alkali light chain 1	3	0	0	1
per	period	10	0	3	2
Pgi	Phosphoglucose isomerase	12	2	1	4
Rh3	Rhodopsin 3	6	0	3	0
Zw	Zwischenferment	10	0	4	4
	Total	54	4	14	13

GenBank accession numbers or references for these sequences are: *Adh* (*mel*: M17834-37, M19547, M22210, *sim*: X57362-64, M36591, X00607, *ere*: X54120, *ore*: M37837, *yak*-13: X54120, X57365-76, *tei*: X54118). *Adhr* [*mel*: (Kreitman and Hudson, 1991), *sim*: (Sumner, 1991), *ere*: X54116, *yak*: X54120, *tei*: X54118]. *boss* [*mel*, *sim*, *tei*-3, *yak*-4 (Ayala and Hartl, 1993)]. *Mlc*1 (*mel*: L37312017, *sim*: L49010-14, *tei*: L49008, *yak*: L49007). *per* (*mel*: L07817-19, L07821, L07823, L07825, *sim*: L07828-32, *yak*: X61127). *Pgi* (*mel*: L27544-46, L27553-55, *sim*: L27547-51, U20556-59, U20564-65, *yak*-13: L27673-85, *tei*-1: J. H. McDonald, pers. commun.). *Rh3* [*mel*, *sim*, *tei*-5, *yak*-5: (Ayala et al., 1993)]. *Zw* (*mel*: U43167, U43165, U42748, U42742, U42738–39, *sim*: L13876, L13878, L13881, L13883, L13891, *yak*: U42750). *mel*, *sim*, *yak*, *tei*, *ere* and *ore* refer to *D. melanogaster*, *D. simulans*, *D. yakuba*, *D. teisseri* and *D. orena*, respectively. *m* is the number of alleles examined for each gene in each species. All samples are unbiased with respect to allozyme polymorphism, except the sequences of *Adh* in *D. melanogaster*.

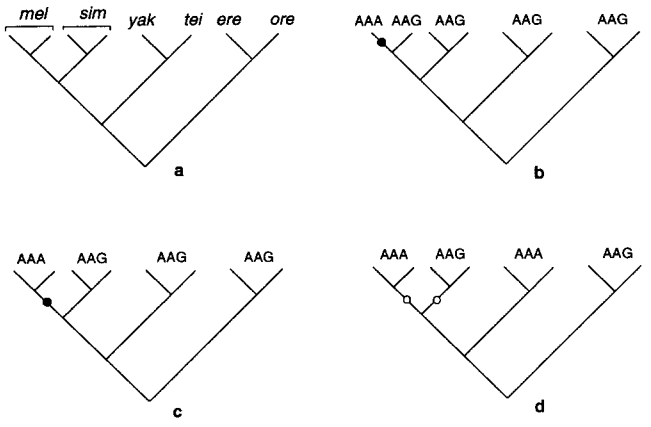

Fig. 2. Inferring the direction of synonymous changes. Tree (**a**) shows phylogenetic relationships within the *D. melanogaster* subgroup proposed by Lachaise et al. (1988) and Jeffs et al. (1994). Evolutionary trees connect a single codon that has changed within or between two alleles each of *D. melanogaster* and *D. simulans* and one allele each of *D. yakuba*, *D. teisseri*, *D. erecta* and *D. orena* – labeled *mel*, *sim*, *yak*, *tei*, *ere* and *ore*, respectively. The most parsimonious change for the codons of tree (**b**) is an unpreferred polymorphism, AAG→AAA, in the *D. melanogaster* lineage. The location of the mutation on the tree is marked with a dot. The most parsimonious change in tree (**c**) is an unpreferred fixation, AAG→AAA, in *D. melanogaster*. Tree (**d**) shows a case for which the direction of a synonymous mutation is ambiguous (multiple trees give the least number of changes). Such sites were not included in the analyses. Note that the relative positions of the *ere*/*ore* and *tei*/*yak* lineages do not affect inference of the direction of synonymous mutations. At three-, four- and six-fold redundant sites, mutations between major codons or between non-major codons were not included in the analyses because no predictions were made for their effects on fitness. Figure from Akashi (1995).

D. simulans lineage and the changes in these forces responsible for the decline of codon bias in *D. melanogaster*.

4. Frequency distributions and divergence under mutation–selection drift

Both major codon preference and mutational models of codon bias require weak evolutionary forces that remain relatively constant over time. This notion conforms with Kimura and Ohta's view of molecular evolution; gene frequency changes within species and the accumulation of fixed differences between species reflect a single process of mutation, genetic drift, and (for some mutations) unidirectional selection (Kimura and Ohta, 1971; Kimura, 1983). The process is governed by mutation rates and transition probabilities for changes in gene frequency. The transition probabilities depend on the product of effective population size and selection coefficient, $N_e s$ (Kimura, 1983). Fig. 3 depicts new mutations arising and evolving within a population of constant size. Most mutations are lost quickly, but a few survivors reach appreciable frequencies within the population and a subset of these go to fixation. Positive

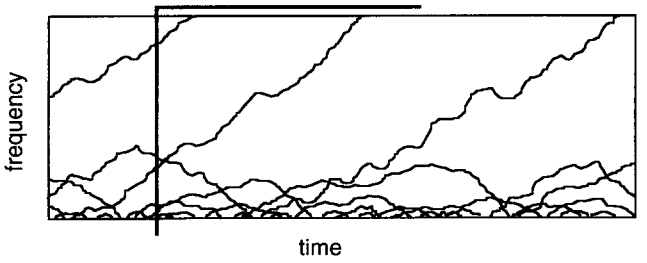

Fig. 3. Polymorphism as a transient phase of evolution. A schematic diagram of the evolutionary trajectories of mutations within a population. The thick vertical line represents within-population sampling. The number of trajectory lines intersecting this line represents the number of segregating sites, and the heights at which they intersect depict their frequencies in the population. Trajectory lines that reach the top of the diagram are evolutionary fixations or substitutions occurring between the populations. The number of such events that occur along the thick horizontal line reflects the number of fixed differences in the sample.

selection elevates the likelihood that a change will increase in frequency within a population, whereas negative selection decreases the likelihood for changes to spread (Fisher, 1930; Wright, 1938). Relative to neutral mutations, $s>0$ will increase the expected number of segregating sites in a sample of alleles, the frequencies with which the sites are segregating within the sample, and the number of fixed differences between alleles sampled from distantly related populations. Negative selection coefficients will have the opposite effect on these data. Under weak evolutionary forces, the timescale of this process is too great to observe small differences directly in the evolutionary trajectories of mutations in laboratory or in natural populations. Within- and between-species comparisons of DNA sequence data, however, may provide a means of identifying weak selection.

Fig. 4 illustrates the quantitative effects of selection on sequence variation. The histograms represent the expected proportions of newly arising mutations found at different frequencies in a sample of $m=5$ sequences. Note that 'frequency' here refers to the frequency of mutations within the sample of sequences rather than the frequency of major codons at a given locus. Under neutrality, most segregating mutations (frequency classes $r=1$ to $m-1$) are expected to be rare (i.e., found in a small number of the sampled alleles). The proportion of fixed mutations ($r=m$) depends on the amount of time examined on a given lineage. In the *D. simulans* lineage, the number of fixed neutral mutations is expected to be close to the number of singletons in the sample. Fig. 5b shows how natural selection can skew this distribution. Even very weak selection causes adaptive mutations to be in higher frequencies than negatively selected changes. The most noticeable skews occur for the frequency classes one and the fixed class, but the

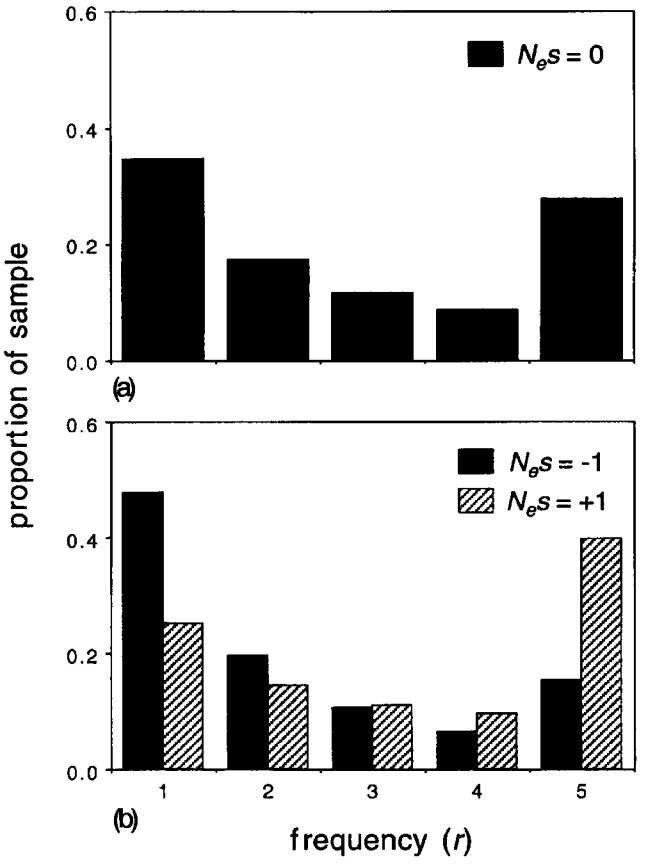

Fig. 4. Frequency distributions and divergence of nearly neutral mutations. The proportion of a sample of newly arising mutations at frequency classes $r=1$ to m in a sample of $m=5$ sequences [calculated using Sawyer and Hartl's sampling formulae (Sawyer and Hartl, 1992)]. In order to calculate the proportion of mutations in the fixed class, $t_{div}=0.6$, the time of divergence on the lineage (scaled to the effective population size), was estimated from *D. simulans* intron polymorphism and divergence data [see Akashi and Schaeffer (1997)]. (**a**) Expected frequency distribution and divergence under neutrality. (**b**) Effect of negative and positive selection on expected frequency distributions and divergence.

intermediate frequency classes also show some sensitivity to selection.

Differences in the average fitness effects of mutations can be identified by contrasting information from such histograms between functional classes of DNA changes. Sawyer et al. (1987) compared the frequency distributions of segregating mutations for samples of sequences from a single population (no divergence data). McDonald and Kreitman (1991) suggested comparing the numbers of polymorphic (pooled across frequency classes) and fixed differences among categories of mutations. Templeton (1996) attempted to combine these approaches by contrasting the numbers of singleton polymorphisms ($r=1$), polymorphisms at intermediate frequencies ($r=2$ to $m-1$), and fixed differences ($r=m$). Although each of these approaches is valid, none employs all the available information in the frequency distribution and divergence histogram. The Sawyer et al. approach does not consider numbers of fixed differences, the McDonald–Kreitman method does not include information about the frequencies with which polymorphic mutations segregate within a sample, and Templeton's test pools data for polymorphic mutations at frequencies greater than one. Because even very weak selection will affect each frequency class, contrasting the proportion of mutations across the histogram, referred to as the frequency distribution and divergence test, should be a statistically more powerful approach to detect fitness effects of DNA mutations.

Fig. 5 compares the statistical power of the four methods described above to detect mutation–selection drift. Under assumptions that the frequency distributions are stationary and that all sites evolve independently (no genetic linkage and independent effects on fitness), the numbers of sampled mutations in each frequency class are independent Poisson random variables (Sawyer and Hartl, 1992). Frequency distribution and divergence data were simulated under mutation–selection drift and the proportion of samples for which the statistical tests rejected the null hypothesis of fitness equivalence is plotted in Fig. 5. For all the tests, the power to detect selection increases initially with $N_e s$ but then falls off. This is because when $N_e s$ reaches approximately five, all sites become fixed for major codons and the number of preferred changes in the sample decreases to zero. However, the mutation–selection drift model, to explain codon bias in *Drosophila*, requires $N_e s$ in the range of roughly 1–3.

Each of the statistical methods shows some ability to detect weak selection (Fig. 5). For most parameter values tested, contrasts between the numbers of polymorphic and fixed differences are more sensitive to selection than comparisons of frequency distributions. Because the two approaches consider different aspects of the data, combining the two methods by contrasting the numbers of singletons, intermediate frequency polymorphisms, and fixed differences generally enhances the power of the statistical test. The frequency distribution and divergence test further extends this approach by employing more of the information in the sample. Under mutation–selection drift, the FDD test is either more powerful than, or indistinguishable from, the other approaches over all parameter values examined. The gain in power is greatest when the number of sampled alleles is large.

These simulations also show that increasing the number of sites sampled has a greater impact on statistical power than increasing the numbers of alleles. Although only five alleles were analyzed in *D. simulans*, close to 2000 silent sites were examined across the eight genes in Table 2. If codon bias is maintained under mutation–selection drift in this lineage, then comparing

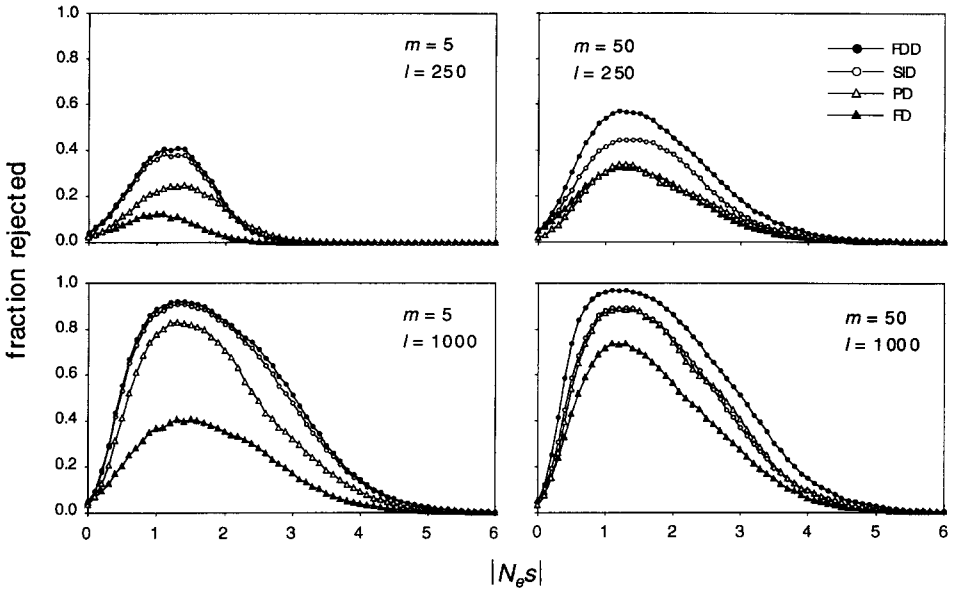

Fig. 5. Statistical power to detect weak selection. For a given absolute value of $N_e s$, the numbers of preferred and unpreferred mutations in each frequency class were sampled from Poisson distributions with means calculated from Sawyer and Hartl's formulae (Sawyer and Hartl, 1992). FD refers to frequency distributions, PD to polymorphism and divergence, SID to singletons, intermediate frequencies, and divergence, and FDD to frequency distributions and divergence (see text). The y-axis plots the proportion of tests which reject fitness equivalence, $P<0.05$. One-tailed Mann–Whitney U-tests were applied to FD, SID, and FDD data, and one-tailed Fisher's exact tests were applied to PD data. The four graphs show results for sample sizes, $m=5$, 50 sequences from a population and for sequence lengths, $l=250$, 1000 mutable sites. Each data point is based on 10 000 simulations of two-state mutation-selection drift (Eq. 3) with parameters $N_e=10^6$, $t_{div}=0.6$, u (per site)$=3\times10^{-8}$, and v (per site)$=2\times10^{-8}$. u/v of 1.5 gives an equilibrium mutational base composition of 60% A+T, the average base composition of *D. melanogaster* introns (Shields et al., 1988; Moriyama and Hartl, 1993).

Table 2
Frequency distribution and divergence of silent DNA polymorphism in *D. melanogaster* and *D. simulans*

Species	r	n_r	
		Unpref	Pref
D. simulans	1	55	12
	2	22	4
	3	7	5
	4	3	3
	5	14	13
D. melanogaster	1	26	2
	2	12	0
	3	14	0
	4	7	1
	5	10	0
	6	54	4

The number of non-ancestral mutations, n_r, segregating at frequency, r, in the samples of five *D. simulans* and six *D. melanogaster* sequences are shown for preferred (unpref) and preferred (pref) silent changes. Data were pooled across the eight genes listed in Table 1. See Akashi and Schaeffer (1997) for frequency distribution data for individual genes.

frequency distribution and divergence data should have a high probability of rejecting fitness equivalence between preferred and unpreferred mutations.

5. Evidence for major codon preference in D. simulans

Under major codon preference, advantageous preferred silent mutations will show frequency distributions and divergence skewed toward higher values than deleterious unpreferred mutations. Equivalent fitness effects is the null hypothesis in this comparison, and neutrality of both classes of mutations (the purely mutational model) is a subset of this null. Table 2 shows the numbers of silent mutations in each frequency class pooled across the eight *D. melanogaster* and *D. simulans* genes examined. The sample size of preferred changes is small in *D. melanogaster*, and none of the statistical tests shows any evidence of fitness differences between silent mutations. In *D. simulans*, however, the proportions of preferred and unpreferred mutations segregating at different frequencies and fixed in the sample are similar to that expected under weak selection (Fig. 6). The 37 preferred mutations are segregating at higher frequencies and are more likely to be fixed than the 101 unpreferred changes (Mann–Whitney U-test, $z=3.12$, $P=0.0009$, one-tailed). The other statistical tests were also significant at the 5% level: frequency distributions are skewed toward higher values (Mann–Whitney U-test, $z=1.71$, $P=0.044$), ratios of polymorphism to divergence are lower (Fisher's exact test, $P=0.007$), and the ratios of singleton, intermediate frequency, and fixed

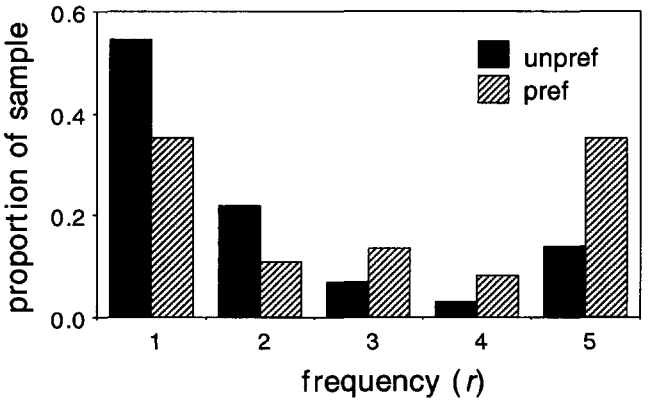

Fig. 6. Frequency distributions and divergence of synonymous DNA mutations in *D. simulans*. The proportion of 101 unpreferred (black) and 37 preferred (striped) mutations segregating at the given frequencies or fixed in the sample are shown. Pooled data from eight *D. simulans* genes from Table 2.

differences are skewed toward higher values for preferred than for unpreferred mutations (Mann–Whitney *U*-test, $z=2.8$, $P=0.003$).

These patterns are both consistent with mutation–selection drift and difficult to explain in the absence of weak selection. Although the power curves of Fig. 5 were generated under assumptions of stationarity and free recombination, inference of fitness differences do not depend on these assumptions (Sawyer et al., 1987). Genetic linkage and departures from equilibrium can skew the frequency distributions from that expected under stationarity, but, because the two classes of mutations are interspersed in DNA, such effects will have an equivalent impact on the two classes of mutations and will not cause their distributions to differ.

Comparisons of the frequency distributions and divergence of synonymous DNA changes also distinguish between major codon preference and mutational models of codon bias (Akashi, 1995). In *Drosophila*, higher mutation rates from A/T→C/G than in the reverse direction could explain codon usage bias (Eq. (2)). Differences in mutation rates between the two classes of mutations will affect the numbers of segregating and fixed mutations, but will not affect the proportion of mutations in each frequency class. Recent changes in mutation rates, however, could affect these comparisons (Eyre-Walker, submitted). If the ratio of mutation rates, u/v, has increased since the most recent common ancestor to the polymorphism segregating within a population, then the number of polymorphic unpreferred mutations will be higher, and the frequency spectra of such mutations will be skewed toward rares. However, differences in the frequency distributions of preferred and unpreferred mutations have been observed in *D. pseudoobscura* as well as in *D. simulans*; it is unlikely that the same recent change in mutation pressure has occurred independently in these species (Akashi and Schaeffer, 1997).

Neither departures from stationarity and linkage equilibrium nor mutational biases appear to explain the observed differences between the frequency distributions and divergence of preferred and unpreferred silent changes in *D. simulans*. Major codon preference appears to be the predominant explanation for the maintenance of codon bias in this lineage. However, the simple mutation–selection-drift model will probably require refinement. For example, evidence that selection to enhance translational fidelity plays a role in biasing codon usage (Akashi, 1994; Eyre-Walker, 1996) suggests within-gene variation in selection coefficients at silent sites. The findings presented above also do not exclude the possibility that selection pressures other than major codon preference contribute to patterns of codon bias. In *E. coli*, reduced codon bias at the start of genes (Bulmer, 1988; Eyre-Walker and Bulmer, 1993) and the evolutionary persistence of non-major codons (Maynard Smith and Smith, 1996) suggest that selection may favor non-major codons at some sites. The extent to which such forces act in *Drosophila* remains to be established.

6. Reduced codon bias in D. melanogaster

Patterns of synonymous DNA evolution suggest that codon bias has been maintained under a balance among mutation pressure, genetic drift and natural selection in the *D. simulans* lineage and in the common ancestor to *D. melanogaster* and *D. simulans*. Under such a model, the reduction in codon bias in *D. melanogaster* could be due to changes in either mutational biases, u/v, or the product of effective population size and selection coefficient, $N_e s$ (Eq. (3)). Changes in these parameters make distinguishable predictions for changes in the base composition of sites under different selection pressures (Akashi, 1996b). Fig. 7a shows the expected change in codon bias between species under different mutational biases. The change is largest for low codon bias genes, remains relatively flat as MCU increases, and decreases in more highly biased genes. Fig. 7b shows the expected change in codon bias between species differing in selection intensity at silent sites. Changes in $N_e s$ predict small differences in codon bias in low codon bias genes and greater differences as selection intensity for codon bias increases. The change in codon bias decreases in very highly biased genes, but few loci show such high major codon usage.

Regions under the weakest selection pressure for base composition show the greatest sensitivity to changes in mutational biases. Since all major codons end in either G or C, the mutational hypothesis predicts a greater decrease in G+C content at neutrally evolving sites than at silent sites in coding regions in *D. melanogaster*.

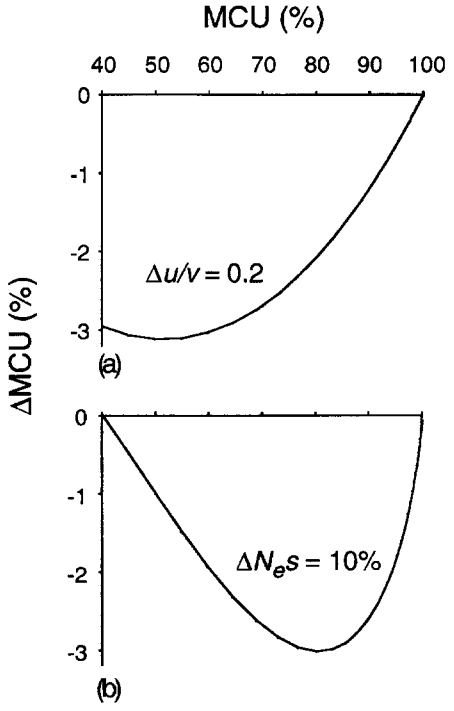

Fig. 7. Sensitivity of major codon usage to changes in mutational biases and selection intensity. The x-axis plots an initial state of major codon usage [MCU is equivalent to q in Eqs. (2) and (3)], and the y-axis shows the expected percentage decline in codon bias for a given parameter change. (**a**) ΔMCU is shown for an increase of u/v from 1.5 to 1.7. (**b**) ΔCU is plotted for a 10% decrease in N_eS. The relationships were determined from Eq. (3) and assume that codon bias evolution has reached steady-state and that changes in mutational bias or selection intensity are uniform across loci. From Akashi (1996b).

Changes in N_eS predict the opposite pattern, a larger excess of G/C→A/T changes in coding regions. Because introns were difficult to align with outgroup sequences, comparisons were made at sites fixed in each species but differing between species. Under the null hypothesis of no difference in base composition, the number of sites at which *D. melanogaster* encodes G or C and *D. simulans* encodes A or T ($mel_{GC}sim_{AT}$) should equal the number of sites in the opposite configuration ($mel_{AT}sim_{GC}$). The 16 introns examined (from the eight genes of Table 1) show little evidence for a change in base composition: 30 intron sites are $mel_{AT}sim_{GC}$, and 26 are $mel_{GC}sim_{AT}$. Silent sites within exons, however, show significant differences in base compositions between these species: 83 sites are $mel_{AT}sim_{GC}$, whereas 25 are $mel_{GC}sim_{AT}$ (Fisher's exact test, $P=0.004$). Changes in N_eS, rather than in mutational biases, appear to explain the reduction of codon bias in the *D. melanogaster* lineage.

These analyses assume that some intron sites evolve neutrally, but do not assume that intron sequences are completely free of functional constraint. A number of findings suggest that intron sequences affect mRNA splicing and gene expression levels (Mount et al., 1992; Schaeffer and Miller, 1993; Stephan and Kirby, 1993; Laurie and Stam, 1994; Leicht et al., 1995). However, relatively high rates of nucleotide substitution (Hudson et al., 1994) and insertion/deletion evolution (Akashi, 1996b) suggest that introns contain a large fraction of unconstrained sites.

Two additional findings further support a relaxation of selection in *D. melanogaster*. The ratio of preferred to unpreferred substitutions in this lineage is consistent with a five-fold or greater reduction of N_eS since its split from *D. simulans*. In addition, unpreferred substitutions appear to be distributed uniformly among different synonymous families (Akashi, 1996b). Although data for a larger number of genes are desirable, these findings suggest that synonymous DNA evolution may be close to neutral in the *D. melanogaster* lineage. It is important to note, however, that these analyses do not establish the cause(s) of the smaller N_eS; the effective population size may have been smaller in *D. melanogaster* (allowing a greater proportion of deleterious mutations to drift to fixation) or the fitness effects of mutations affecting translational efficiency may have been smaller in this lineage.

7. Conclusions

The maintenance of codon bias under major codon preference is a dynamic equilibrium involving both the fixation of deleterious unpreferred mutations through the action of mutation pressure and genetic drift, and the 'compensatory' substitution of equal numbers of adaptive preferred mutations. Population genetic analyses suggest that codon bias in *D. simulans* has been maintained under such a scenario. Major codon preference is very sensitive to changes in selection intensity or mutational biases. In *D. melanogaster*, a reduction in N_eS appears to have allowed a genome-wide decline in codon bias. These findings suggest that micro-evolutionary analyses can reveal adaptation in genome-wide base composition when the effects of single nucleotide changes may be too subtle to detect through functional assays.

Acknowledgement

I am indebted to Marty Kreitman, John Gillespie, and Stanley Sawyer for their many contributions to this work. I am also grateful to Brian Charlesworth, David Cutler, Adam Eyre-Walker, Thomas Nagylaki, Eli Stahl, Chung-I Wu, and an anonymous reviewer for their criticism and suggestions. H.A. is supported by an NSF/Sloan Foundation Postdoctoral Fellowship in Molecular Evolution.

References

Akashi, H., 1994. Synonymous codon usage in *Drosophila melanogaster*: Natural selection and translational accuracy. Genetics 136, 927–935.

Akashi, H., 1995. Inferring weak selection from patterns of polymorphism and divergence at 'silent' sites in *Drosophila* DNA. Genetics 139, 1067–1076.

Akashi, H., 1996a. Natural selection and the population genetics of synonymous DNA changes in *Drosophila*. Ph.D. thesis, University of Chicago.

Akashi, H., 1996b. Molecular evolution between *Drosophila melanogaster* and *D. simulans*: Reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster*. Genetics 144, 1297–1307.

Akashi, H., Schaeffer, S.W., 1997. Natural selection and the frequency distributions of 'silent' DNA polymorphism in *Drosophila*. Genetics 146, 295–307.

Andersson, S.G.E., Kurland, C.G., 1990. Codon preferences in free-living microorganisms. Microbiol. Rev. 54, 198–210.

Ayala, F.J., Chang, B.S.W., Hartl, D.L., 1993. Molecular evolution of the *Rh3* gene in *Drosophila*. Genetica 92, 23–32.

Ayala, F.J., Hartl, D.L., 1993. Molecular drift of the *bride of sevenless* (*boss*) gene in *Drosophila*. Mol. Biol. Evol. 10, 1030–1040.

Bennetzen, J.L., Hall, B.D., 1982. Codon selection in yeast. J. Biol. Chem. 257, 3026–3031.

Berg, O.G., Martelius, M., 1995. Synonymous substitution-rate constants in *Escherichia coli* and *Salmonella typhimurium* and their relationship to gene expression and selection pressure. J. Mol. Evol. 41, 449–456.

Bulmer, M., 1988. Are codon usage patterns in unicellular organisms determined by selection–mutation balance. J. Evol. Biol. 1, 15–26.

Bulmer, M., 1991. The selection–mutation-drift theory of synonymous codon usage. Genetics 129, 897–907.

Carulli, J.P., Krane, D.E., Hartl, D.L., Ochman, H., 1993. Compositional heterogeneity and patterns of molecular evolution in the *Drosophila* genome. Genetics 134, 837–845.

Curran, J.F., Yarus, M., 1989. Rates of aminoacyl-tRNA selection at 29 sense codons in vivo. J. Mol. Biol. 209, 65–77.

Eyre-Walker, A., Bulmer, M., 1993. Reduced synonymous substitution rate at the start of enterobacteria genes. Nucleic Acids Res. 21, 4594–4603.

Eyre-Walker, A., Bulmer, M., 1995. Synonymous substitution rates in enterobacteria. Genetics 140, 1407–1412.

Eyre-Walker, A., 1996. Synonymous codon bias is related to gene length in *Escherichia coli*: selection for translational accuracy? Mol. Biol. Evol. 6, 864–872.

Fisher, R.A., 1930. The Genetical Theory of Natural Selection. Clarendon Press, Oxford.

Freese, E., 1962. On the evolution of base composition of DNA. J. Theor. Biol. 3, 82–101.

Gouy, M., Gautier, C., 1982. Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. 10, 7055–7064.

Grosjean, H., Fiers, W., 1982. Preferential codon usage in prokaryotic genes: the optimal codon-anti-codon interaction energy and the selective codon usage in efficiently expressed genes. Gene 18, 199–209.

Hudson, R.R., Bailey, K., Skarecky, D., Kwiatowski, J., Ayala, F.J., 1994. Evidence for positive selection in the superoxide dismutase (Sod) region of *Drosophila melanogaster*. Genetics 136, 1329–1340.

Ikemura, T., Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translation system. 1981. J. Mol. Biol. 151, 389–409.

Ikemura, T., Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes: Differences in synonymous codon choice patterns of yeast and *Escherichia coli* with reference to the abundance of isoaccepting transfer RNAs. 1982. J. Mol. Biol. 158, 573–597.

Ikemura, T., 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. 2, 13–34.

Jeffs, P.S., Holmes, E.C., Ashburner, M., 1994. The molecular evolution of the alcohol dehydrogenase and alcohol dehydrogenase-related genes in the *Drosophila melanogaster* species subgroup. Mol. Biol. Evol. 11, 287–304.

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

Kimura, M., Ohta, T., 1971. Protein polymorphism as a phase of molecular evolution. Nature 229, 467–469.

Kliman, R.M., Hey, J., 1993. Reduced natural selection associated with low recombination in *Drosophila melanogaster*. Mol. Biol. Evol. 10, 1239–1258.

Kliman, R.M., Hey, J., 1994. The effects of mutation and natural selection on codon bias in the genes of *Drosophila*. Genetics 137, 1049–1056.

Kreitman, M., Hudson, R.R., 1991. Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogater* from patterns of polymorphism and divergence. Genetics 127, 565–582.

Lachaise, D., Cariou, M.L., David, J.R., Lemeunier, F., Tsacas, L., Ashburner, M., 1988. Historical biogeography of the *Drosophila melanogaster* species subgroup. Evol. Biol. 22, 159–225.

Laurie, C.C., Stam, L.F., 1994. The effect of an intron polymorphism on alcohol dehydrogenase expression in *Drosophila melanogaster*. Genetics 138, 379–385.

Leicht, B.G., Muse, S.V., Hanczyc, M., Clark, A.G., 1995. Constraints on intron evolution in the gene encoding the myosin alkali light chain in *Drosophila*. Genetics 139, 299–308.

Li, W.-H., 1987. Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. J. Mol. Evol. 24, 337–345.

McDonald, J.H., Kreitman, M., 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature 351, 652–654.

Maynard Smith, J., Smith, N.H., 1996. Site-specific codon bias in bacteria. Genetics 142, 1037–1043.

Moriyama, E.N., Hartl, D.L., 1993. Codon usage bias and base composition of nuclear genes in *Drosophila*. Genetics 134, 847–858.

Mount, S.M., Burks, C., Hertz, G., Stormo, G.D., White, O., Fields, C., 1992. Splicing signals in *Drosophila*: intron size, information content, and consensus sequences. Nucleic Acids Res. 20, 4255–4262.

Sawyer, S.A., Dykhuizen, D.E., Hartl, D.L., 1987. Confidence interval for the number of selectively neutral amino acid polymorphisms. Proc. Natl. Acad. Sci. USA 84, 6225–6228.

Sawyer, S.A., Hartl, D.L., 1992. Population genetics of polymorphism and divergence. Genetics 132, 1161–1176.

Schaeffer, S.W., Miller, E.L., 1993. Estimates of linkage disequilibrium and the recombination parameter determined from segregating nucleotide sites in the alcohol dehydrogenase region of *Drosophila pseudoobscura*. Genetics 135, 541–552.

Sharp, P.M., Li, W.-H., 1986. An evolutionary perspective on synonymous codon usage in unicellular organisms. J. Mol. Evol. 24, 28–38.

Sharp, P.M., Li, W.-H., 1987. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol. Biol. Evol. 4, 222–230.

Sharp, P.M., Li, W.-H., 1989. On the rate of DNA sequence evolution in *Drosophila*. J. Mol. Biol. 28, 398–402.

Sharp, P.M., Cowe, E., 1991. Synonymous codon usage in *Saccharomyces cerevisiae*. Yeast 7, 657–678.

Sharp, P.M., Stenico, M., Peden, J.F., Lloyd, A.T., 1993. Codon usage: mutational bias, translational selection, or both? Biochem. Soc. Trans. 21, 835–841.

Shields, D.C., Sharp, P.M., Higgins, D.G., Wright, F., 1988. 'Silent'

sites in *Drosophila* genes are not neutral: Evidence of selection among synonymous codons. Mol. Biol. Evol. 5, 704–716.

Stephan, W., Kirby, D.A., 1993. RNA folding in *Drosophila* shows a distance effect for compensatory fitness interactions. Genetics 135, 97–103.

Sueoka, N., 1962. On the genetic basis of variation and heterogeneity of DNA base composition. Proc. Natl. Acad. Sci. USA 48, 582–592.

Sueoka, N., 1988. Directional mutation pressure and neutral molecular evolution. Proc. Natl. Acad. Sci. USA 85, 2653–2657.

Sumner, C., 1991. Nucleotide polymorphism in alcohol dehydrogenase duplicate of *Drosophila simulans*: Implications for the neutral theory. Undergraduate thesis, Princeton University.

Templeton, A.R., 1996. Contingency tests of neutrality using intra/interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochroms oxidase II gene in the Hominoid primates. Genetics 144, 1263–1270.

Varenne, S., Buc, J., Lloubes, R., Lazdunski, C., 1984. Translation is a non-uniform process: effect of tRNA availability on the rate of elongation of nascent polypeptide chains. J. Biol. Chem. 180, 549–576.

Wright, S., 1938. The distribution of gene frequencies under irreversible mutation. Proc. Natl. Acad. Sci. USA 24, 253–259.