Translational selection and molecular evolution

Hiroshi Akashi* and Adam Eyre-Walker†

An interplay among experimental studies of protein synthesis, evolutionary theory, and comparisons of DNA sequence data has shed light on the roles of natural selection and genetic drift in 'silent' DNA evolution.

Addresses
*Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas 66045-2106, USA;
e-mail: hiroshi@falcon.cc.ukans.edu
†Centre for the Study of Evolution, University of Sussex, Brighton, BN1 7ET, UK
Correspondence: Hiroshi Akashi

Current Opinion in Genetics & Development 1998, **8**:688–693

http://biomednet.com/elecref/0959437X00800688

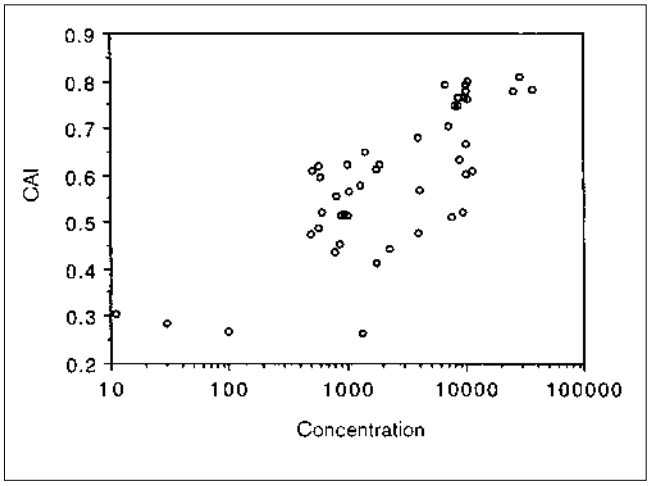
Introduction

Both functional and fitness consequences of synonymous codon usage have been identified in a number of taxa through a combination of biochemical studies of protein synthesis and analyses of DNA sequences. A balance among mutational processes, genetic drift, and selection to enhance the efficiency of protein synthesis, 'major codon preference', appears to explain patterns of codon usage at a large fraction of silent sites. In this review, we begin with a brief summary of patterns of codon usage that suggest translational selection for major codons. We then discuss the biochemical basis of adaptation at silent sites and present recent evidence both supporting and motivating modifications to the simple form of major codon preference. Finally, we discuss the role of translational selection in the evolution of both amino acid composition and protein size.

Patterns of codon usage: major codon preference

A number of patterns of base composition and DNA sequence variation suggest that natural selection discriminates among synonymous codons to enhance protein synthesis [1,2]. In *Escherichia coli* and *Saccharomyces cerevisiae*, genome-wide base composition at silent sites is biased toward a subset of 'major' codons for each amino acid. Among synonymous codons recognized by multiple tRNAs, major codons tend to be encoded by abundant tRNAs, and among synonymous codons recognized by a single tRNA, major codons generally show perfect Watson-Crick base pairing with the tRNA anticodon [3–8]. *In vivo* experiments in *E. coli* suggest that major codons can increase both the speed and the accuracy of translation [1,9]. Such codons could enhance fitness by increasing growth rates and/or by reducing the metabolic costs of protein synthesis. Correlations between tRNA pools and codon usage have also been noted in the genomes of *Salmonella typhimurium* [10], *Mycoplasma capricolum* [11], bacteriophage T7 [12], and, more recently, in the nuclear genome of *Drosophila melanogaster* [13•].

Figure 1



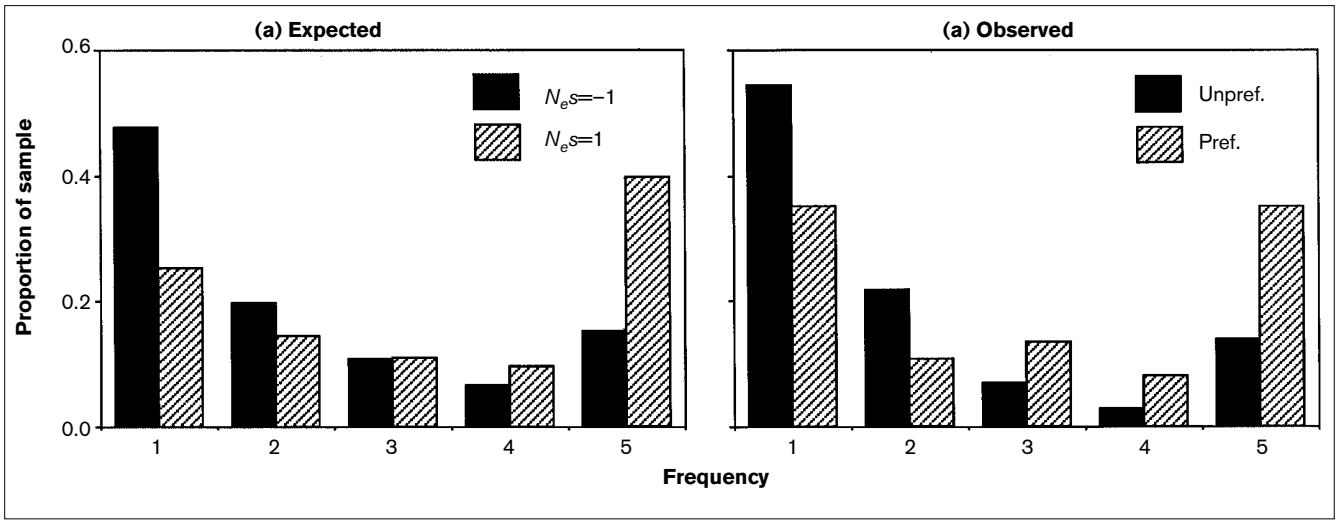
Codon bias and protein abundance in *E. coli*. Codon adaptation index (CAI) [72] values are plotted against protein concentration (molecules per cell) for 46 *E. coli* genes. CAI values are a positive function of codon bias. Figure from [14].

Although overall patterns of codon usage correlate with tRNA pools in many species, codon bias varies considerably among genes within their genomes. In *E. coli* and yeast, major codon usage shows a strong correlation with gene expression levels (Figure 1) [6,10,14–16]. Such patterns are consistent with major codon preference because the fitness benefit of a translationally superior codon should increase with the number of aminoacyl-tRNA selections it experiences. Similar relationships have been established in a number of other prokaryotes [17–20] and fungi [21–23], as well as in the nuclear genomes of *D. melanogaster* [24], *Arabidopsis thaliana* [25], *Dictyostelium discoideum* [26], *Caenorhabditis elegans* [27], and in the chloroplast genomes of a number of plant and algal lineages [28].

Population genetic tests of major codon preference

The simplest model of major codon preference is an evolutionary balance among natural selection favoring translationally superior major codons and mutation pressure and genetic drift allowing the presistence of slightly deleterious minor codons [29,30]. Population genetics theory predicts an inverse relationship between the efficacy of natural selection at a given nucleotide site and genetic linkage to other selected sites [29,31–35]. Under major codon preference, the effectiveness of selection, and thus levels of codon bias, should be a positive function of regional rates of

Figure 2



Population genetics of silent DNA mutations in *D. simulans*. **(a)** The expected proportions of preferred and unpreferred mutations found segregating within, or fixed between, populations were calculated according to the findings of Sawyer and Hartl [73] for a sample of five alleles and t_{div}, time of divergence scaled to N_e generations, of 0.6. The x-axis represents the number of alleles in which newly arisen mutations are present. The histogram shows the expected proportion of mutations in each frequency class under a two-state model of mutation-selection-drift with the scaled fitness difference, $N_e s$, between major and minor codons set to one. **(b)** Observed proportions of preferred and unpreferred silent mutations in a sample of five alleles of each of eight *D. simulans* genes. Data were pooled across loci. Adapted from [38•] (see reference for details of the method).

recombination. Kliman and Hey [36] showed that, in the *D. melanogaster* genome, genes in regions of reduced crossing-over show lower codon bias relative to genes located in regions of higher recombination.

Selection for codon bias also predicts differences in the evolutionary dynamics of putative fitness classes of silent mutations interspersed within a region of DNA [37]. Under major codon preference, silent mutations from non-major to major codons, 'preferred' mutations, should confer a small fitness benefit to the organism. Mutations in the reverse direction, 'unpreferred' mutations, should incur a fitness cost of the same magnitude. Comparisons of the within and between species evolutionary dynamics of preferred and unpreferred mutations have revealed patterns quite similar to those predicted by mutation-selection-drift [38•] (Figure 2).

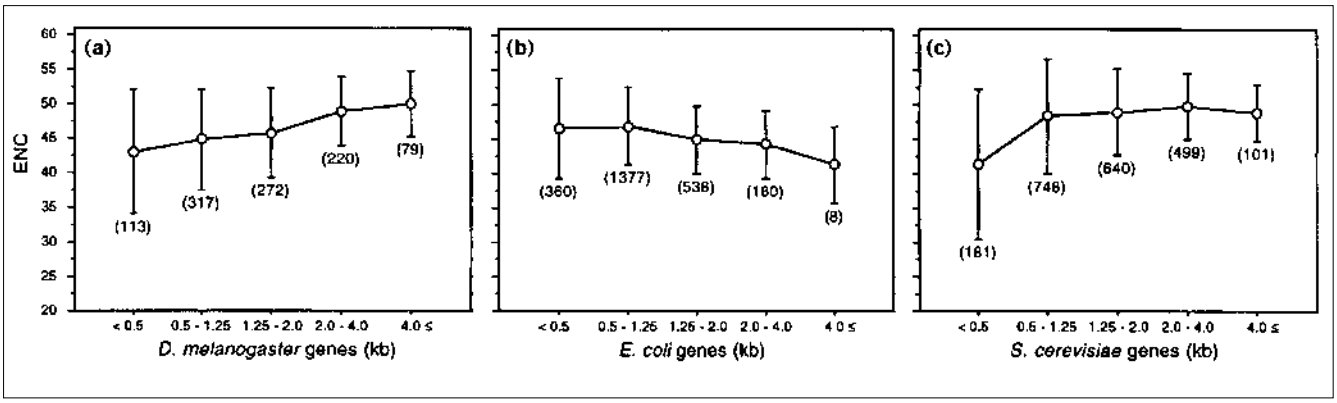
Finally, measures of silent DNA divergence are inversely related to synonymous codon usage bias between *E. coli* and *S. typhemurium* [39,40], *D. melanogaster* and *D. pseudoobscura* [41–43], *C. elegans* and *C. briggsae* [44,45], and among land plant chloroplast genes [46]. Although such patterns are qualitatively consistent with greater 'selective constraint' in highly expressed genes, the quantitative relationship between codon bias and silent divergence may not fit that predicted by major codon preference [47]. For some synonymous families in *E. coli*, the relationship between substitution rate and expression level does not appear to be related to codon bias.

Biochemical bases of selection at silent sites

Although the patterns discussed above strongly support the notion that major codons are generally favorable, the relationship between codon usage, the biochemistry of protein synthesis, and the fitness of organisms in populations is only beginning to emerge. At least three facets of translation could be affected by synonymous codon usage: the rate of elongation, the cost of proofreading, and the accuracy of translation (including rates of missense and processivity errors) [30,48]. During protein synthesis, a ribosome waits at a particular codon for the arrival of an aminoacyl-tRNA. This process is potentially costly: ribosomes are relatively expensive to synthesize, and the time for which they are idle should be minimized. Once the tRNA is bound to the RNA in the ribosome, elongation factor Tu is released with the hydrolysis of GTP. This reaction, known as 'substrate' or 'kinetic' proofreading, provides extra time to discern whether the correct tRNA is bound to the mRNA but the process is energetically costly: if the tRNA is rejected it must be recharged with elongation factor and GTP. In addition, despite proofreading, incorrect amino acids can be misincorporated into the growing peptide chain with the possibility that the protein produced will be either functionless or of reduced activity. Finally, the ribosome can undergo processivity errors while translocating, such as 'slipping' out of frame or 'dropping off' (premature termination) from the mRNA. Such errors are likely to result in dysfunctional proteins.

In vivo experiments in *E. coli* have shown that, during polypeptide chain elongation, the speed of aminoacyl-tRNA

Figure 3



Codon bias and gene length in **(a)** *D. melanogaster*, **(b)** *E. coli*, and **(c)** *S. cerevisae*. The means and standard deviations of a measure of codon bias, the effective number of codons (ENC) [74], are plotted for five length categories. The number of genes included in each category is shown in parentheses. ENC is inversely related to the degree of codon usage bias. Adapted from [58••].

selection at a given codon is proportional to the cognate tRNA's abundance [49]. Faster cognate tRNA recognition could confer a fitness advantage by enhancing any, or all, of the three facets of translation. But which, if any, of these processes determines the fitness effects of synonymous codon usage? The quantitative relationship between tRNA concentration and codon usage in *E. coli* grown under a variety of conditions [50] shows a remarkably good fit to the predictions of a model of growth optimization through selection for enhanced translational elongation rates [51••]. Berg and Silva [52••], however, have found that the correspondence between rates of elongation and patterns of synonymous codon bias in one synonymous family is not strong. The glutamic acid codon GAA is translated ~3.5 times faster than its synonym GAG, irrespective of context in *E. coli* [53]. Although selection to maximize the rate of elongation should favor GAA over GAG in all contexts, the frequency of GAA compared to GAG only increases with expression level among glutamic acid codons followed by G (i.e. GAA.G is increasingly favoured over GAG.G) [52••]. It is possible that mutational biases or other selection pressures, perhaps on local nucleotide composition, may act against GAG.nonG in *E. coli*.

Berg and Kurland's model of selection for translational elongation rates is consistent with the relationship between codon use and tRNA pools [51••] but similar models of selection on proofreading costs and translational accuracy have not been as thoroughly investigated. Two lines of evidence from DNA sequence comparisons suggest that the reduction of translational misincorporation rates also confers a fitness advantage to major codons. The fitness cost of errors in protein synthesis should be a function of the number and size of dysfunctional peptides produced. If natural selection biases codon usage to enhance translational accuracy, then, within a gene, selection should be stronger at codons encoding constrained amino acids — those at which a misincorporation would disrupt protein function — than at codons encoding less constrained amino acids. In *D. melanogaster*, higher major codon usage in DNA-binding motifs than elsewhere in transcription factor genes and in conserved than non-conserved amino acid positions in interspecific comparisons of proteins are consistent with codon selection for translational accuracy [54]. Such a pattern is not observed *in E. coli* [55]. Interestingly, although analyses of codon usage in mammals is complicated by large-scale patterns of base compositional bias [56], a similar correlation between silent and protein rates of evolution *within* loci suggests either that mutational biases vary within genes or that codon selection for translational accuracy also acts in mammals [57•].

Selection for translational accuracy also predicts a relationship between gene length and codon bias. Given the same level of expression, the energetic cost of producing dysfunctional peptides should be a function of their size. Eyre-Walker [14] showed that, for *E. coli* genes encoding proteins found in equimolar amounts, codon bias is indeed a function of gene length. Such a pattern has also been found among yeast ribosomal proteins [58••].

Moriyama and Powell [58••] have studied genome-wide relationships between codon bias and protein length in *E. coli*, *S. cerevisiae*, and *D. melanogaster*. Surprisingly, codon bias increases in larger proteins in *E. coli* but decreases in *S. cerevisiae and D. melanogaster* (Figure 3). The interpretation of such patterns is complicated by the multiplicity of factors that could influence the relationship between protein size and codon bias. Given the same level of expression and protein functional constraint, selection for translational accuracy predicts that longer genes should show greater codon bias. If expression levels differ among genes, however, then selection to decrease protein size — perhaps in the face of selection for protein stability or function — should be a positive function of expression levels. Finally, the efficacy of selection at a given nucleotide site is inversely related to the

mutation rate to deleterious mutations at linked sites [29,35]. Silent sites in longer genes may be linked to a greater number of deleterious mutations (at both silent and replacement sites) resulting in reduced codon bias relative to smaller genes ([29]; JM Comeron, M Kreitman, personal communication). As the three forces may act simultaneously, disentangling among their contributions may require analyses of more genomes to identify how recombination rates and gene expression levels influence the relationship between gene length and codon bias.

Conflicting selection pressures: evidence for antagonistic pleiotropy at silent sites

Although the benefits of major codon usage are well established, our understanding of the persistence of translationally suboptimal codons is less clear. Mutation pressures and genetic drift may account for many, if not most, minor codons, but both biochemical studies and sequence comparisons suggest that, at particular locations within genes, minor codons may be beneficial. For example, in *E. coli*, programmed frameshift events can require ribosomal pausing during elongation and slippage of the translational apparatus on the mRNA. Some such events depend on the use of minor codons [59,60•]. Furthermore, proper intracellular protein folding may require translational pauses, mediated by minor codon usage or mRNA structure, at specific locations within genes [61].

Several patterns of DNA sequence variation also support the notion that minor codons can be advantageous. In *E. coli*, reduced codon bias and silent divergence at the start of genes [62–65] may reflect constraints imposed by ribosome binding [64]. Similar patterns have been discovered recently in *D. melanogaster* [66•]. Major codon usage also decreases at the end of *E. coli* genes; apparently due, in part, to the fact that many genes overlap the Shine-Dalgarno, or coding sequence, of the next gene on the chromosome [65]. Finally, Maynard Smith and Smith [67] have shown that some sites in the middle of genes are occupied by a minor codon across a wide range of very diverged enteric bacteria. This suggests that natural selection may favor the persistence of apparently translationally inferior codons; however, the functional basis and relative frequency of minor codon preference remains to be established.

Conclusions: translation selection and protein evolution?

This review has focussed on translational selection at silent sites. The body of evidence supporting major codon preference has increased in the past several years through new methods and in a larger number of genomes. At the same time, some intriguing data have emerged that suggest that this model may require some refinement; minor codons may also be preferred at certain sites.

The contribution of translational selection in protein evolution has been the object of far less attention. As discussed above, protein length may be reduced to a greater extent in highly expressed genes in order to maximize the efficiency of translation. In addition, if tRNAs for some amino acids are translated more efficiently or accurately than others (or if some amino acids are less energetically costly to synthesize [68•]), then codons encoding such amino acids could also be favored at the level of translation. In *Mycoplasma capricolum* and *E. coli* [11], as well as in yeast [69•], tRNA pools match the amino acid usage of proteins, and in *E. coli*, the amino acid composition of proteins differs among highly and lowly expressed genes [70]. It is possible that tRNA pools are simply selected to match the amino acid composition of highly expressed proteins. Alternatively, tRNA pools and the amino acid composition of proteins may be co-adapted to enhance translational efficiency. Andersson and Kurland [71] have put forth convincing evidence that genomic evolution can drive the evolution of the translational system. It will be of great interest to determine the extent to which the converse holds; translational selection may play an important role in patterning both silent and replacement nucleotide composition as well as protein and genome size.

Acknowledgements

This work was supported by a grant from the National Science Foundation/Alfred P Sloan Foundation to Hiroshi Akashi. A Eyre-Walker is a Royal Society University Research Fellow.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Andersson SGE, Kurland CG: **Codon preferences in free-living microorganisms**. *Microbiol Rev* 1990, **54**:198-210.

2. Sharp PM, Stenico M, Peden JF, Lloyd AT**: Codon usage: mutational bias, translational selection, or both?** *Biochem Soc Trans* 1993, **21**:835-841.

3. Post LE, Nomura M: **DNA sequences from the str operon of *Escherichia coli***. *J Biol Chem* 1980, **255**:4660-4666.

4. 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.** *J Mol Biol* 1981, **151**:389-409.

5. 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.** *J Mol Biol* 1982, **158**:573-597.

6. Bennetzen JL, Hall BD: **Codon selection in yeast.** *J Biol Chem* 1982, **257**:3026-3031.

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

8. Hani J, Feldman H: **tRNA genes and retroelements in the yeast genome**. *Nucleic Acids Res* 1998, **26**:689-696.

9. Parker J: **Errors and alternatives in reading the universal genetic code.** *Microbiol Rev* 1989 **53**:273-298.

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

11. Yamao F, Andachi Y, Muto A, Ikemura T, Osawa S: **Levels of tRNAs in bacterial cells as affected by amino acid usage in proteins.** *Nucleic Acids Res* 1991, **19**:6119-6122.

12. Sharp PM, Rogers MS, McConnell, DJ: **Selection pressures on codon usage in the complete genome of bacteriophage T7.** *J Mol Evol* 1984, **21**:150-160.

13. • Moriyama EN, Powell J: **Codon usage bias and tRNA abundance in *Drosophila*.** *J Mol Evol* 1997, **45**:514-523.
tRNA abundances correlate with synonymous codon usage in *D. melanogaster*. Developmental-stage-specific tRNA concentrations may also affect patterns of codon usage.

14. Eyre-Walker A: **Synonymous codon bias is related to gene length in *Escherichia coli*: selection for translational accuracy?** *Mol Biol Evol* 1996, **13**:864-872.

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

16. Sharp PM, Cowe E: **Synonymous codon usage in *Saccharomyces cerevisiae*.** *Yeast* 1991, **7**:657-678.

17. Shields DC, Sharp PM: **Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational biases.** *Nucleic Acids Res* 1987, **15**:8023-8040.

18. Andersson GE, Sharp PM: **Codon usage in the *Mycobacterium tuberculosis* complex.** *Microbiology* 1996, **142**:915-925.

19. Pouwels PH, Leunissen JA: **Divergence in codon usage of Lactobacillus species.** *Nucleic Acids Res* 1994, **22**:929-936.

20. Malumbres M, Gil JA, Martin JF: **Codon preference in corynebacteria.** *Gene* 993, **134**:15-24.

21. Lloyd AT, Sharp PM: **Codon usage in *Aspergillus nidulans*.** *Mol Gen Genet* 1991, **230**:288-294.

22. Lloyd AT, Sharp PM: **Synonymous codon usage in *Kluyveromyces lactis*.** *Yeast* 1991, **9**:1219-1228.

23. Lloyd AT, Sharp PM: **Evolution of codon usage patterns: the extent and nature of divergence between *Candida albicans* and *Saccharomyces cerevisiae*.** *Nucleic Acids Res* 1992, **20**:5289-5295.

24. Shields DC, Sharp PM, Higgins DG, Wright F: **'Silent' sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons.** *Mol Biol Evol* 1988, **5**:704-716.

25. Chiapello H, Fisacek F, Caboche M, Henaut A: **Codon usage and gene function are related in sequences of *Arabidopsis thaliana*.** *Gene* 1998, **209**:GC1-GC38.

26. Sharp PM, Devine KM: **Codon usage and gene expression level in *Dictyostelium discoideum*: highly expressed genes do 'prefer' optimal codons.** *Nucleic Acids Res* 1989, **17**:5029-5039.

27. Stenico M, Lloyd AT, Sharp PM: **Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases.** *Nucleic Acids Res* 1994, **22**:2437-2446.

28. Morton BR: **Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages.** *J Mol Evol* 1998, **46**:449-459.

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

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

31. Muller HJ: **The relation of recombination to mutational advance.** *Mutat Res* 1964, **1**:2-9.

32. Hill WG, Robertson A: **The effect of linkage on limits to artificial selection.** *Genet Res* 1966, **8**:269-294.

33. Felsenstein J: **The evolutionary advantage of recombination.** *Genetics* 1974, **78**:737-756.

34. Barton NH: **Linkage and the limits to natural selection.** *Genetics* 1995, **140**:821-841.

35. Charlesworth B: **The effect of background selection against deleterious mutations on weakly selected, linked variants.** *Genet Res Camb* 1994, **63**:213-27.

36. Kliman RM, Hey J: **Reduced natural selection associated with low recombination in *Drosophila melanogaster*.** *Mol Biol Evol* 1993, **10**:1239-1258.

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

38. • Akashi H: **Codon bias evolution in *Drosophila*: population genetics of mutation-selection-drift.** *Gene* 1997, **205**:269-278.
Comparisons of polymorphic and fixed nucleotide differences from naturally occurring populations are shown to be a powerful method for identifying even very weak selection for codon bias. Such comparisons show patterns very similar to those expected under a model of mutation-selection-drift at silent sites in *Drosophila simulans*.

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

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

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

42. Moriyama EN, Gojobori T: **Rates of synonymous substitution and base composition of nuclear genes in *Drosophila*.** *Genetics* 1992, **130**:855-864.

43. Carulli JP, Krane DE, Hartl DL, Ochman H: **Compositional heterogenaity and patterns of molecular evolution in the *Drosophila* genome.** *Genetics* 1993, **134**:837-845.

44. Kennedy BP, Aamodt EJ, Allen FL, Chung MA, Heschl MRP, McGhee JD: **The gut esterase gene (ges-1) from the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*.** *J Mol Biol* 1993, **229**:890-908.

45. Stenico M, Lloyd AT, Sharp PM: **Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases.** *Nucleic Acids Res* 1994, **22**:2437-2446.

46. Morton BR: **Codon use and the rate of divergence of land plant chloroplast genes.** *Mol Biol Evol* 1994, **11**:231-238.

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

48. Kurland CG: **Translational accuracy and the fitness of bacteria.** *Annu Rev Genet* 1992, **26**:29-50.

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

50. Dong H, Nilsson L, Kurland CG: **Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates.** *J Mol Biol* 1996, **260**:649-663.

51. •• Berg O, Kurland CG: **Growth rate-optimised tRNA abundance and codon usage.** *J Mol Biol* 1997, **270**:544-550.
A model in which tRNA concentrations and codon usage are optimized to maximize growth rates by enhancing translational elongation rates shows a good fit to data from *E. coli*. Both the degeneracies and association rate constants for different tRNA–codon combinations appear to be important in such models.

52. •• Berg OG, Silva PJN: **Codon bias in *Escherichia coli*: the influence of codon context on mutation and selection.** *Nucleic Acids Res* 1997, **25**:1397-1404.
The context of the bases in the nearest surrounding codon is shown to influence synonymous codon use. Comparisons of context dependence across expression levels show that both mutational and selective effects determine such patterns.

53. Sorensen MA, Pedersen S: **Absolute *in vivo* translation rates of individual codons in *Escherichia coli*. The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate.** *J Mol Biol* 1991, **222**:265-280.

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

55. Hartl DL, Moriyama EN, Sawyer S: **Selection intensity for codon bias.** *Genetics* 1994, **138**:227-234.

56. Bernardi G, Olofsson B, Filipski J, Zerial M, Salinas J, Cuny G, Meunier-Rotival M, Rodier F: **The mosaic genome of warm-blooded vertebrates**. *Science* 1985, **228**:953-958.

57. • Alvarez-Valin F, Jabbari K, Bernardi G: **Synonymous and nonsynonymous substitutions in mammalian genes: intragenic correlations.** *J Mol Evol* 1998, **46**:37-44.
Rates of silent and protein evolution are correlated within mammalian genes and both are correlated with G+C content. Intragenic correlations in mutation rates or selection for translational accuracy at constrained amino acid positions may explain such patterns.

58. Moriyama EN, Powell JR: **Gene length and codon usage bias in •• *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*.** *Nucleic Acids Res* 1998, **26**:3188-3193.
Genome-wide relationships between synonymous codon usage bias and protein lengths are investigated. Surprisingly, in *E. coli*, the relationship is positive, whereas in *D. melanogaster* and yeast, codon bias is negatively related to gene length. Although translational selection may play a role in such relationships, the cause(s) of these patterns remains to be determined.

59. Farabaugh PJ: **Programmed translational frameshifting.** *Annu Rev Genet* 1996, **30**:507-528.

60. Schwartz R, Curran JF: **Analyses of frameshifting at UUU- • pyrimidine sites.** *Nucleic Acids Res* 1997, **25**:2005-2011.
Minor codon usage and the nucleotide 3′ to the 'slippery' site can have substantial effects on the frequency of frameshifting. The UUU-phenlyalanine codon is particularly prone to frameshifting when followed by Ynn. Such dicodons are strongly avoided in highly expressed genes in *E. coli*.

61. Purvis IJ, Bettany AJE, Santiago TC, Coggins JR, Duncan K, Eason R, Brown AJP: **The efficiency of folding of some proteins is increased by controlled rates of translation *in vivo*: a hypothesis.** *J Mol Biol* 1987, **193**:413-417.

62. Bulmer M: **Codon usage and intergenic position.** *J Theor Biol* 1988, **133**:67-71.

63. Chen GF, Inouye M: **Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes.** *Nucleic Acids Res* 1990, **18**:1465-1473.

64. Eyre-Walker A, Bulmer M: **Reduced synonymous substitution rate at the start of enterobacteria genes.** *Genetics* 1993, **140**:1407-1412.

65. Eyre-Walker A: **The close proximity of *Escherichia coli* genes: consequences for stop codon and synonymous codon use.** *J Mol Evol* 1996, **42**:73-78.

66. Kliman RR, Eyre-Walker A: **Patterns of base composition within the • genes *of Drosophila melanogaster*.** *J Mol Evol* 1998, **46**:534-541.
Unusual patterns of intragenic base composition are found in *D. melanogaster*. At the start of genes (the first few hundred base pairs), third position G+C content increases. On a larger scale, however, G+C content declines across genes (from 5′ to 3′) in both exons and introns. The authors attribute the former pattern to selection against major codons at the start of genes but the cause(s) of the latter pattern is unclear.

67. Maynard Smith J, Smith NH: **Site-specific codon bias in bacteria.** *Genetics* 1996, **142**:1037-1043.

68. Craig CL, Weber RS: **Selection costs of amino acid substitutions • in ColE1 and Colla gene clusters harbored by *Escherichia coli*.** *Mol Biol Evol* 1998, **15**:774-776.
The authors calculate the bioenergetic costs and biosynthetic complexity (number of enzymes required) to synthesize amino acids in *E. coli*. Selection to reduce such costs and complexity may impose a constraint on the amino acid composition of proteins.

69. Xia X: **How optimized is the translational machinery in *Escherichia* • *coli*, *Salmonella typhimurium* and *Saccharomyces cerevisiae*.** *Genetics* 1998, **149**:37-44.
Under a model of translational selection where ribosomal elongation rates limit the rate of translation, tRNA concentrations are optimally adapted to codon usage when they match the square-root of the frequency of their corresponding synonymous codons. The model predicts a similar relationship between tRNA abundances and amino acid usage and data from *E. coli*, *S. typhimurium*, and *S. cerevisiae* appear to be in agreement with this prediction.

70. Lobry JR, Gautier C: **Hydrophobicity, expressivity, and aromaticity are the major trends of amino-acid usage in 999 *Escherichia coli* chromosome-encoded genes.** *Nucleic Acids Res* 1994, **22**:3174-3180.

71. Andersson SG, Kurland CG: **Genomic evolution drives the evolution of the translation system.** *Biochem Cell Biol* 1995, **73**:775-787.

72. Sharp PM, Li W-H: **The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications.** *Nucleic Acids Res* 1987, **15**:1281-1295.

73. Sawyer SA, Hartl DL: **Population genetics of polymorphism and divergence.** *Genetics* 1992, **132**:1161-1176.

74. Wright F: **The 'effective number of codons' used in a gene.** *Gene* 1990, **87**:23-29.