CMLS, Cell. Mol. Life Sci. 58 (2001) 125–134
1420-682X/01/010125-10 $ 1.50 + 0.20/0

CMLS Cellular and Molecular Life Sciences

Review

Genes in sweeping competition

D. I. Nurminsky

Department of Anatomy and Cell Biology, Tufts University School of Medicine, 136 Harrison Ave., Boston (Massachusetts 02111, USA), Fax + 1 617 636 6536, e-mail: dnurmi01@granite.tufts.edu

Received 5 May 2000; received after revision 22 August 2000; accepted 24 August 2000

Abstract. Analysis of DNA variation is a powerful tool for detecting adaptation at the genomic level. The contribution of adaptive evolution is evident from examples of rapidly evolving genes, which represent the likely targets for strong selection. More subtle adaptation is also an integral component of routine maintenance of gene performance, continuously applied to every gene. Adaptive changes in the population are accomplished through selective sweeps, i.e. complete or partial fixation of beneficial alleles. The evidence is accumulating that selective sweeps are quite frequent events which, together with associated genetic hitchhiking, represent dominant forces that influence molecular evolution by shaping the variability pattern in the genome.

Key words. DNA polymorphism; adaptation; rapid evolution; codon bias; selective sweep.

Introduction

Positive selection at the nucleotide level constitutes the essence of Darwinian evolution. Favorable mutations are driven to fixation by natural selection, thus fitting the population to a perpetually changing environment. Although the Darwinian evolution model is more or less widely accepted, detecting adaptive selection at the molecular level has proven to be a complicated task, and the issue of the relative input of adaptation in molecular evolution is still being argued.

Fixation of a particular beneficial mutation is hard to detect because the fixation process may take a long time, and because it is not easy to identify the beneficial mutation in the pool of neutral or deleterious ones. Fortunately, adaptive evolution leaves a discernible footprint on the pattern of nucleotide polymorphism. A strongly selected beneficial allele rapidly heads to fixation through a selective sweep, displacing other less fit alleles from the population and thus decreasing nucleotide diversity at the selected locus. Polymorphic variants linked to the selected allele are also dragged to fixation, reducing nucleotide diversity in the adjacent regions as well. This process, known as genetic hitchhiking [1], creates a trough in the polymorphism pattern indicative of recent adaptive fixation in the region. Mathematical modeling suggests that hitchhiking of the neighboring neutral locus is efficient if $c < s$, and becomes negligible at $c \approx s$, where s is a selection coefficient for a favored allele, and c is a recombination fraction between the selected and the neutral loci [2, 3]. Rough estimations for *Drosophila* show that in a euchromatic region with a normal recombination rate of 2×10^{-8}/bp, for a weakly favored allele (such as a preferred synonymous codon with a selection coefficient of 10^{-5} [4]), hitchhiking will be effective within a range of just a few base pairs. In contrast, strong selection with a coefficient of 10^{-2} under the same circumstances would cause detectable hitchhiking at distances up to dozens of kilobase pairs away.

Other events may lead to reduction of polymorphism at a given locus similar to selective sweep, but the outcomes differ at the genomic and populational levels. One widely discussed mechanism, the population bottleneck, involves transient drastic decline of population

size, leading to a genome-wide reduction of polymorphism within a population. Bottleneck should have similar effect on all loci within a given population, unlike a selective sweep that affects only the region around a selected locus. Another possible mechanism, background selection, comprises the removal of recurring deleterious mutations from population. This results in elimination of polymorphic variants linked to the mutant loci and thus renders nucleotide variation low [5, 6]. Background selection should act similarly in different populations or in closely related species with similar chromosomal structure, since its action relies on the local genomic features such as gene density and recombination rate. In contrast, selective sweep is often a population-specific event. Hence, a locus-specific reduction in polymorphism could be explained by a recent selective sweep, or by another mechanism such as background selection. However, if the polymorphism deficit is both locus specific and population specific, it is unequivocally indicative of a selective sweep.

Here we shall review accumulating evidence for selective sweeps and associated genetic hitchhiking as a dominant force shaping the variability of the genome. Selective sweep is the mechanism by which adaptive evolution operates at the genomic level. The contribution of adaptive evolution will be emphasized by outlining examples of rapidly evolving genes that represent the major targets for selection. Positive selection is also a component of routine maintenance of gene performance, continuously applied to every gene. Numerous examples of individual selective sweeps associated with adaptation will be outlined. Moreover, genome-wide analysis of the correlation between codon bias and nucleotide polymorphism will be provided, implying the principal role of genetic hitchhiking in maintaining the level of genome variability. This model is consistent with an astonishingly high estimation of frequency of selective sweeps. Finally, the evidence will be summarized for selective sweep-driven differentiation between populations.

Detecting adaptive evolution

In *Drosophila*, cross-hybridization studies followed by DNA sequencing [7] led to the estimation that as many as one-third of genes are undergoing rapid evolution. Conserved genes are likely to provide housekeeping functions that are unchanged between distantly related organisms, and should be mostly subject to purifying selection. Rapidly evolving genes are likely to experience more frequent adaptive changes due to another type of selection that diversifies species and fits them to different ecological niches. Alternatively, fast evolution of a gene may be determined by relaxed selective pressure that would allow accumulation of excessive random mutations as a result of unconstrained neutral evolution.

Statistical tests have been devised to discriminate between adaptive and neutral evolution. One group of tests, introduced by McDonald and Kreitman, analyzes the long-term history of a gene by comparing the ratios of nonsynonymous to synonymous nucleotide replacements within and between the species. The expectation is that neutral evolution would result in similarity between the patterns of divergence and polymorphism, whereas adaptive changes would probably not—for example, excessive amino acid differences may occur between species, indicating diversifying selection [8–10]. A similar test developed by Hudson, Kreitman and Aguade is based on the expectation that neutral evolution should act similarly in generating both divergence and polymorphism in different genomic regions. The test includes comparison of the evolution of two regions, one ordinarily being a noncoding segment used as a control for expectation of neutrality, and the other a coding region of interest [11]. In these cases, significant deviation from the neutral model indicates possible selection.

Another group of tests, such as Tajima's D statistics [12, 13] are designed to detect recent selective events inconsistent with neutral evolution. In the equilibrial population without selection, the frequency of polymorphic variants forms a spectrum characteristic of balance between random mutation and genetic drift. If a selective event such as selective sweep renders nucleotide diversity in the locus low, then for some time most of the polymorphisms will be represented by newly accumulated mutations, thus skewing the frequency spectrum towards rare variants. On the contrary, balancing selection may maintain a set of few major alleles, resulting in overabundance of frequent polymorphisms.

Individual analyses of rapidly evolving genes, where applied, usually result in rejection of neutrality by at least one of the tests, implying adaptive changes. Moreover, applying these tests to more conserved genes, as discussed below, often reveals substitution patterns indicative of adaptive evolution. However, adaptive evolution may be even more frequent than detected by using the tests. Theoretical analysis showed that Tajima's D test is only able to detect departure from neutrality in a quite narrow time window after a selective sweep [14]. Other tests are more effective, but they often lack power due to insufficient sample size, especially in regions with low polymorphism.

Rapidly evolving genes

As mentioned, the size of the genomic segment affected by hitchhiking is proportional to the strength of selection driving the selective sweep. The larger the hitchhiked segment, the higher the probability of detecting a selective sweep in the region. This brings our attention to the rapidly evolving genes that represent potential targets for strong selection.

Male reproductive genes

A major part of rapidly evolving genes are sex related. High-resolution two-dimensional (2D) protein electrophoresis revealed that gonadal proteins are much more divergent between *Drosophila* species than nongonadal samples, and that gonadal proteins from males are more divergent than those from females [15]. Survey of the available gene sequences also indicated a high ratio of replacement to silent differences (K_a/K_s) for the sex-related genes in *Drosophila* and *Caenorhabditis* [16]. The examples presented below include mostly the genes directly involved in gamete production, mating and fertilization, although Civetta and Singh [17] reviewed the evidence that genes involved in a wider spectrum of sex-related activities also evolve at an elevated rate.

We recently reported a new gene for sperm axonemal dynein subunit, *Sdic*, which was created from duplicated copies of two other genes within approximately 3 million years [18]. During this evolutionarily short time period, a number of nucleotide changes followed to create a first exon coding for an N-terminus characteristic for axonemal dynein intermediate chains. Low polymorphism was detected in the *Sdic* region, consistent with a recent selective sweep at or around *Sdic*. Although an alternative scenario was put forward to justify the low *Sdic* polymorphism [19], additional studies [20, D. I. Nurminsky, D. D. DeAguiar and D. L. Hartl, unpublished] indicate that selective sweep remains the most likely explanation.

Ting and colleagues [21] described the rapid evolution of *Odysseus*, a *Drosophila* hybrid sterility gene. Homologs were found in other organisms, allowing authors to estimate a thousandfold increase in the *Odysseus* evolution rate associated with recent speciation. Newly accumulated divergence shows the high K_a/K_s ratio indicative of adaptive evolution. The authors implied that *Odysseus* probably acquired a new function in male reproduction after gene duplication. The old neural function is obviously carried out by another duplicated copy, *Dunc-4*, that remains very conserved [22]. Somewhat similar is the case of duplication of the male accessory gland protein gene *Acp70A* in *Drosophila subobscura* and *Drosophila madeirensis* [23]. Duplicated *Acp70A* copies actively diverge from each other, and one of the copies undergoes faster evolution than its counterpart. Gene duplication has been considered as a versatile tool used by evolution to isolate distinct functions of multifunctional gene products [24]. In light of this theory, it is intriguing that in the related species *D. melanogaster* (no duplication), Acp70A apparently is a multifunctional peptide, because it has an effect on (i) increase in the egg laying rate and (ii) decrease in female remating efficiency. The attractive explanation that in *D. subobscura*, distinct Acp70A functions were split between duplicated copies, requires further study. Other genes for accessory gland proteins, such as *Acp26Ab* [25], *Acp29AB* [26] and *Acp26Aa* [27], also demonstrate unusually high K_a/K_s ratios and an excess of fixed replacement differences between species, indicative of positive selection. The same trend was observed for the gene for Esterase-6, a component of seminal fluid [28].

Willett [29] provided evidence for directional selection acting on the pheromone-binding proteins (PBPs) in moths. Statistical tests indicated a drastic excess of substitutions in PBPs in some of the moth lineages. In one case this accelerated evolution may be associated with the acetate-to-aldehyde pheromone change. But this explanation may not be expanded to other cases, because pheromone changes in other lineages were not associated with adaptive evolution of PBP, and a comparable increase in the PBP evolution rate was observed in a lineage with no pheromone change.

The trend observed in insects, and in *Drosophila* in particular, is evident in other taxa. Positive selection has been implied in evolution of a gamete recognition protein of sea urchin [30], and of abalone sperm fertilization genes [31] which are involved in the sperm-egg interaction. In mice [32], a high K_a/K_s ratio associated with low polymorphism implies positive selection acting on the androgen-binding protein locus (*Abpa*), which is presumably involved in premating isolation between subspecies. Polymorphism for *Abpa* in the domestic mouse is much lower than for two other *X*-linked genes with the same recombination rate, indicative of a recent selective sweep. In primates, extensive survey shows generally higher than average K_a/K_s for the male reproduction-associated genes [33]. Protamines, which represent a major integral component of spermatozoa, have extremely high K_a/K_s values and low replacement polymorphism, indicative of rapid adaptive evolution.

Pesticide resistance

Treatment of populations with exterminating drugs poses probably one of the toughest selective pressures of all. Faced with the 'adapt or die' alternative, pests effectively follow the first option. Advances in molecular biology allowed the identification of the genes re-

sponsible for pesticide resistance, presenting populational biologists with a unique opportunity to look directly at the presumptive targets of strong selection.
Taylor and colleagues [34] reported evidence for selection on the pyrethroid (cypermethrin) resistance gene in populations of the budworm *Heliotis virescens*. Pyrethroids act on voltage-gated sodium channels coded by the *hscp* locus. The authors found that the *hscp* allele diversity was significantly reduced in more resistant populations, consistent with selective sweep(s) associated with selection at the *hscp* locus. Selection for certain alleles of the glutamate-gated chloride channel and P glycoprotein in the same species was associated with treatment with different pesticides [35, 36]. In the mosquito *Aedes aegypti*, 2 decades of organophosphate (OP) treatment resulted in OP resistance that is determined in part by amplification of the esterase genes. Accordingly, low nucleotide polymorphism was detected at the esterase gene, indicative of a selective sweep at the esterase locus [37].

Adaptive evolution of conserved genes

Analysis of rapidly evolving genes provides a fascinating picture of adaptive evolution at work. The examples above represent only a sample and exclude many other known genes undergoing rapid adaptive changes, such as the genes involved in the host-pathogen interactions (for review see [38–40] and references therein).
Adaptive evolution, however, is not at all limited to rapidly evolving genes. Statistical analysis of polymorphism and divergence indicated that a number of genes from a randomly selected sample show signs of adaptive changes ([41–43], reviewed in [44, 45]). Not all of these genes fall into categories for which rapid evolution may be expected, such as male reproductive genes, and many code for well-conserved proteins.
Hartl and Taubes [46] proposed that even without evolution, slightly advantageous substitutions have to be selectively fixed to compensate for random fixation of slightly deleterious substitutions accumulated by mutagenesis. According to this view, selective fixation is a routine feature of evolution rather than a dramatic change in adaptation. Selection also apparently acts on 'silent' substitutions, resulting in codon bias [4, 47]. Fine analysis of polymorphism and divergence pattern indicated that a significant proportion of silent sites and a majority of amino acid replacements affect fitness [10]. These findings imply that along with strong positive selection that drives adaptive changes at a limited number of loci, a weaker positive selection exists that represents an integral part of gene performance maintenance. The input of the latter could be immense, as the weakness of selection is compensated for by a wide spectrum of targets that includes apparently every gene and even multiple loci within single genes.
Weak selection may increase the frequency of a beneficial allele but not drive it to fixation, especially when selection at an adjacent locus favors polymorphic variants linked to another allele(s). Weakness of selection would also influence the size of the genomic region affected by a selective sweep, probably restricting it to a rather small intragenic segment. Nevertheless, linkage disequilibrium between polymorphisms is expected, meaning that certain neutral polymorphic variants would be dragged to higher frequencies due to linkage with selected site(s), thus forming a limited set of haplotypes. If the observed haplotype number is lower than expected from random assortment of polymorphisms, partial selective sweep of the prevalent haplotype may be inferred [42, 48].
Analysis of *Superoxide dismutase* (*Sod*) by Hudson and colleagues [42] revealed a lower than expected number of haplotypes, with one haplotype representing about 50% of population but containing very little polymorphism. The authors suggested that a selective mutation linked to or embedded within this haplotype is sweeping through the population. Kirby and Stephan [49, 50] used the sliding window approach, allowing them to detect a specific *white* gene region subjected to selection. Depaulis and colleagues [51] detected a strong reduction in the haplotype numbers in the *Suppressor of Hairless* locus, suggestive of a selective sweep. In the same population, reduction in haplotype number for another gene, *Fbp2*, was detected [52]. Remarkably, the major coding *Fbp2* haplotype was nearly absent from two other tested populations, suggesting a population- and locus-specific selective event (i.e. selective sweep).
Several mutually nonexclusive scenarios may lead to incomplete fixation/sweep. The simplest one is a selective sweep in progress. Another scenario evokes balancing selection for more than one polymorphic variant, leading to the accumulation of a relevant number of major haplotypes. It is interesting in this respect that analysis of the *Adh* gene, which provided data consistent with balancing selection at the Adh fast/slow polymorphic site [53], indicated a significant linkage between polymorphisms reminiscent of the pattern observed for the *white* gene [49, 50]. Yet another explanation for incomplete fixation/sweep is epistatic interaction between polymorphisms, which may result in selection for more than one combination of polymorphic variants (similar to the simple case of balancing selection as discussed above) [50]. Such interaction between amino acid replacements has been shown directly using the in vitro evolution approach [54]. Finding that nucleotide pairing in RNA may affect gene expression [55, 56] suggests the mechanism for epistatic interaction be-

tween silent substitutions as well. Finally, a trafficking hypothesis has been put forward [50] which suggests that weak selection for a certain allele may be counteracted by selection at the linked locus, thus stalling the fixation process until the two favored mutations are recombined together. Consistent with this explanation, a major haplotype of the 5′ region of *Fbp2* is highly polymorphic in the coding region, whereas the major coding haplotype is polymorphic in the 5′ region [52]. A similar pattern was observed for the sex-peptide *Acp70A* gene, where two different haplotype sets were detected in the 5′ region and in the coding region [57].

Low-recombination regions

The effect of a selective sweep on adjacent polymorphisms is predicted to increase in low-recombination regions. With a decrease in recombination rate, each polymorphic site ultimately becomes linked to a large number of loci, thus increasing the chance for it to hitchhike with positive selection at any of these. The same reasoning, however, is valid to explain a suggested increase of background selection pressure in the low-recombination regions: multiple loci linked to the polymorphic site represent a large target for purifying selection, thus increasing the frequency of removal of polymorphic variants along with negatively selected mutations.

Both selective sweeps and background selection are apparently responsible for the low overall polymorphism in low recombination regions and in species with reduced genome-wide recombination. The trend was observed in *Drosophila* [58–61] (reviewed in [45]), and in mammals [62–64]. In plants [65, 66] (reviewed in [67]), lower polymorphism was detected in genomic regions with lower recombination, and lower overall polymorphism was detected in self-fertilizing plants than in outcrossers.

Although the effects of selective sweeps and of background selection are difficult to distinguish from each other, their cumulative effect is quite evident in regions of very low recombination. However, such regions harbor only a small proportion of the genes, and the role of hitchhiking in controlling the pattern of polymorphism in the gene-rich regions with moderate to high recombination rates still remains unclear. The analysis of the relationship between codon bias and polymorphism presented below provides sound evidence for the genomewide influence of hitchhiking on polymorphism.

Codon bias, recombination and nucleotide diversity

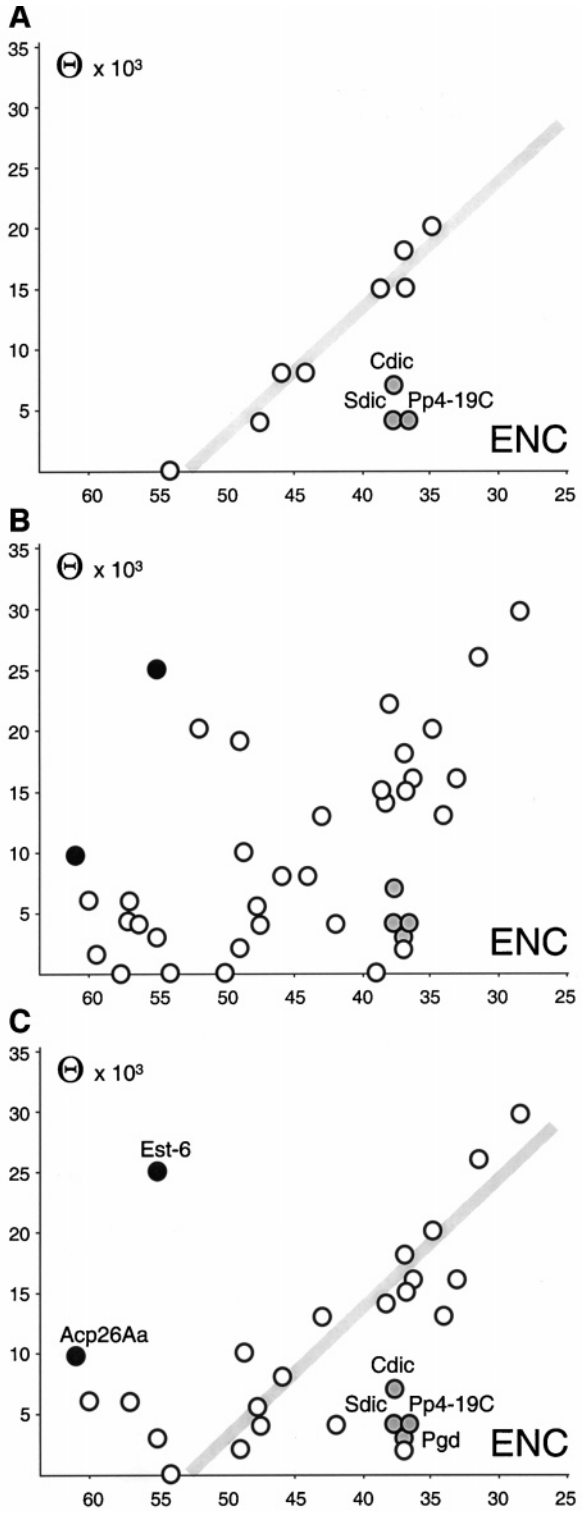
Recurrent selective sweeps and persistent background selection are both predicted to influence molecular evolution by (i) decreasing the rate of fixation of slightly beneficial mutations, (ii) increasing the rate of fixation of slightly deleterious mutations and (iii) increasing the overall rate of evolution, by way of reducing the effective population size [68–70]. Whereas no significant differences in the overall rate of evolution have been reported for low-recombination regions, the observed influence of recombination rate on codon bias is indicative of the other two predicted trends. As mentioned, synonymous codons are not completely neutral [4, 47], and mutations from preferred to rare codons may be considered as slightly deleterious, whereas backwards mutations are considered as slightly beneficial. As expected, in low-recombination regions codon bias is reduced.

Correlation between codon bias and recombination rate has been reported [43, 71, 72]. Comprehensive studies demonstrated that codon bias depends also on other factors intrinsic to the particular genes, such as the gene size [72, 73], and the level of gene expression [74]. Direct proof that recombination rate is indeed a major factor that determines codon bias was provided by analysis of the *yellow* gene in *D. melanogaster* and *D. subobscura* [75]. In *D. subobscura*, *yellow* is located in a high-recombination region and has higher codon bias than in *D. melanogaster*, where it is located in a subtelomeric low-recombination region.

The question arises whether the dominant influence of genetic hitchhiking on the levels of codon bias and nucleotide polymorphism extends beyond the low recombination regions. If this is true, then for a genomewide sample of genes we should expect a positive correlation between codon bias and polymorphism, because hitchhiking has a similar (negative) effect on both. Note that if selection on synonymous sites is dominant, a negative correlation between bias and polymorphism would be expected: less-biased genes would have more synonymous polymorphisms due to relaxed selective constrains. Positive correlation between nucleotide variation and codon bias has been observed by Moriyama and Powell [43]. These data, supported by the wider-scale analysis presented below, imply that the effect of hitchhiking is appreciable not only for low-recombination regions, but also for a genomewide selection of loci.

Genes located on the *X* chromosome of *D. melanogaster* at the eu-heterochromatic junction show a sharp transition from high codon bias characteristic of the euchromatic genes (high recombination) to the low codon bias typical for heterochromatic genes (low recombination) [20]. We measured synonymous polymorphism for these genes and observed a strong positive correlation between codon bias and nucleotide variation [D. I. Nurminsky et al., unpublished data] (fig. 1A). The outliers grouped below the diagonal are the genes that possess a

lower than expected polymorphism due to apparent selective sweep in the *Sdic* region [18], and they include the *Sdic* itself and two adjacent genes, *Cdic* and *Pp4-19C*.



Analysis of more information available from the literature generates a picture which is not as clear (fig. 1B). One of the possible reasons is that whereas the nucleotide polymorphism values for the genes in figure 1A were measured using the same set of stocks from around the world, measurements in figure 1B were performed using different stock sets ranging in origin from all-world collections to single-population samples. That there may be as significant as twofold difference between polymorphism values in different populations was shown by Begun and Aquadro [76] and others, and will be discussed in more detail below. On average, however, the error may not be that large, since our measurements of nucleotide polymorphism at *Zw* and *runt* were within 20% of the values reported earlier [41, 77], despite the difference in sample sets.

Another possible source of excessive 'noise' is the difference in gene size. Too short genes provide inaccurate estimates of codon bias. At the same time, codon bias selection is ineffective for large genes, for which the bias is always low [72]. Detailed study by Moriyama and Powell [73] revealed that codon bias measurements are inaccurate for genes shorter than 300 bp, and the effect of gene size on codon bias becomes apparent for genes longer than 2000 bp. Removal of too short (less than 100 amino acids) and too long (more than 700 amino acids) genes from the data set yields the results shown in figure 1C. Correlation between codon bias and nucleotide polymorphism is quite apparent. Outliers with higher than expected polymorphism (above the diagonal) include *Est-6* and *Acp26Aa*, both well known as hypervariable genes [28, 78]. Lower than expected polymorphism of *Cdic*, *Sdic*, *Pp4-19C* and *Pgd* has been attributed to recent selective sweeps [D. I. Nurminsky et al., unpublished data, 18, 79].

The picture observed suggests that hitchhiking due to selective sweeps and background selection has a significant genomewide influence on nucleotide variation because it is the force responsible for the positive correlation between codon bias and nucleotide diver-

Figure 1. Scatterplot of nucleotide polymorphism (θ) versus codon bias in *D. melanogaster*. ENC (effective number of codons [96]) axis is inverted because codon bias increases with decrease of ENC. Datasets include (*A*) the genes *Zw*, *Bap*, *AnnA*, *Sdic*, *Cdic*, *Pp4-19C*, *run*, *ShakB*, *tty*, *slgA* and *su(f)*, located at the base of *X* chromosome; (*B*) genes from (*A*) plus the data summarized in [44], and reported in [15, 51–53, 61, 90]; (*C*) same as (*B*) with the genes coding for proteins shorter than 100 amino acids and longer than 700 amino acids removed. Linear regression shown in *A* and *C* does not include genes with reported hypervariability (*Acp26Aa* and *Est-6*, labeled black), and with unusually low polymorphism due to recent selective sweeps (*Sdic*, *Cdic*, *Pp4-19C* and *Pgd*, grey).

sity. Therefore, the level of diversity observed for most genes could be described as a result of quasi-equilibrium between recurrent selective sweeps that render nucleotide polymorphism low, and the mutation process that restores it. The lower the recombination rate (and larger the target for selection), the more frequent are sweeps and the lower is the overall level of polymorphism. Background selection obviously provides input by rendering polymorphism even lower, although its relative efficiency in high-recombination regions is not clear. Excessive reduction of polymorphism, such as in the *Sdic* and *Pgd* regions, may result from unusually strong selection for beneficial alleles.

How frequent are selective sweeps?

Analysis of microsatellite polymorphism patterns represents a novel and powerful tool for detection of selective sweeps. Since the mutation rate of microsatellites is high, polymorphism would be restored between rounds of selection, thus resolving the quasi-equilibrial low-polymorphism level into a chain of consecutive selective sweeps. Mutation rates for the *Drosophila* microsatellites are 10^{-5}–10^{-6} mutations/locus/generation [80–82], which is several orders of magnitude higher than the nucleotide mutation rate. Theory predicts that in this situation the dependence of polymorphism on recombination rate should be diminished [83].

Indeed, no significant correlation between heterozygocity at the microsatellite loci and recombination rate was detected in *Drosophila* [84, 85]. When a different measure, the variance in repeat size, was used to evaluate the microsatellite variability, reports regarding dependence on recombination rates ranged from absence of correlation [86] to significant correlation [85]. The discrepancy between the results obtained with two different measures may originate from the nature of the parameters considered. Heterozygocity is a measure for locus variation in the population that reflects the number and frequencies of alleles. Variance in repeat size accounts for the same factors, and in addition for the molecular structure of particular alleles. For example, a rare allele that has a minor influence on heterozygosity may provide a significant input in the variance if its length drastically differs from the length of the major allele. Therefore, correlation between the variance in repeat size and the recombination rate observed by Schug and colleagues [85] could result from peculiarities in the mutation pattern in the set of microsatellites used by the authors. In any case, even if the correlation between microsatellite polymorphism and recombination rate does exist, it is much more concealed than that for nucleotide polymorphism.

About 2 N generations are needed for a population to effectively recover from a selective sweep by nucleotide mutagenesis [87] (ca. 80,000 years for *D. melanogaster*). This time is proportional to the mutation rate. Considering the microsatellite mutation rates that are two or more orders of magnitude higher, the time frame for selective sweep detection using microsatellites is less than 1000 years. This period fits well with the hypothesized history of world colonization by *D. melanogaster* within last 10,000 years, thus emphasizing the value of microsatellites for the detection of population-specific selective events.

Schlotterer and colleagues [86] surveyed seven natural *D. melanogaster* populations for polymorphism at 10 microsatellite loci and detected six locus- and population-specific polymorphism reductions (i.e. selective sweeps). Using the estimated time frame of 1000 years for the detection of sweeps, these data result in a figure of 4×10^{-6} sweep per locus per generation in each population. Each given locus, therefore, would on average undergo a sweep once in 10,000 years.

Mathematical models for mutational accumulation after selective sweep have been developed, and their application to the polymorphism data from *Drosophila* led to the estimation of 0.1 N generations since the last sweep [87, 88], which for *D. melanogaster* transforms into ca. 4000 years. The analysis, which mostly included data on low-recombination regions, was based on the expectation that all mutations accumulated after selective sweep are neutral. This does not account for the possible input of background selection that would render polymorphism lower than expected under neutrality. Hence, sweeps at a frequency lower than calculated would suffice for the observed level of polymorphism, making a figure of 0.1 N an underestimate. Thus, the sweeps are probably more rare than 1 in 4000 years, which is in general agreement with the estimate of 1 in 10,000 years from above.

These evaluations of selective sweep frequency suggest that a given locus experiences a hitchhiking due to a selective sweep once in 10,000 years, and has enough time to recover microsatellite polymorphism (1000 years required), but not nucleotide polymorphism (80,000 years required). These estimates, although crude, explain absent or diminished correlation between microsatellite polymorphism and recombination rate, and provide support to the model that describes quasi-equilibrium between mutation and recurrent sweeps as a major factor in the control of nucleotide polymorphism pattern in the genome.

We could also estimate that, with about 10^4 genes in *Drosophila*, each population would experience a selective sweep on average once a year. There should be therefore a reasonable number of locus-specific differences between populations due to selective sweeps.

Selective sweeps and differentiation of populations

Theoretical analysis [89] suggests that a strong selective sweep may result in differentiation of populations at the hitchhiked locus if the gene flow between populations is low enough. Partial sexual isolation between *Drosophila* populations has been described [90, 91] that, along with geographical separation, could provide a significant barrier to gene flow. The differences between populations would eventually fade away if a beneficial allele that caused the selective sweep in one population is advantageous in others as well. Such an allele will expand over the species, but interpopulational differences could still be detected until the process is complete. Some alleles, however, may confer adaptation to the local environment specific for particular populations and therefore remain confined to these populations.

Differentiation has been observed between distinct *Drosophila* populations [26, 50, 52, 92–94], [95]. Comparison of *D. melanogaster* populations from different continents revealed a special status of African populations, consistent with partial sexual isolation of African flies from others [90]. Based on analysis of seven loci, Begun and Aquadro [75] observed that the DNA polymorphism level in the Zimbabwe population was generally higher than in North America. More detailed study on the *vermillion* gene demonstrated that the African population differs significantly from the American and Asian populations. Linkage disequilibrium between *vermillion* polymorphisms was detected in non-African populations, suggestive of selective sweeps [92]. Sequence analysis of the *glucose dehydrogenase* (*Gld*) locus in different populations of *D. melanogaster* also demonstrated drastic difference between the populations of Asia, Africa and America, with lowered polymorphism in the Zimbabwe population [93]. The effects are locus and population specific (lower polymorphism in Zimbabwe for *Gld*, but higher for *vermillion*), and thus are attributable to selective sweeps.

A significant difference between haplotype structures in the *y-ac-sc* region was observed between European and American populations [94, 95]. European populations showed a skew towards rare variants, which could be attributed to either selective sweep(s) or bottlenecking [95]. However, studies of other loci, including *white* [50] and *Sod* [42], indicated no polymorphism deficit in the European population, arguing against bottleneck. Similar evidence was reported for differentiation between two African populations, associated with a locus- and population-specific reduction of polymorphism (i.e. selective sweep) at the *Fbp2* locus [51, 52].

In conclusion, compelling evidence is becoming available for the pervasive influence of positive selection, consistent with the perception of selective sweeps and associated genetic hitchhiking as major factors in molecular evolution. Further studies, facilitated by recent theoretical advances and novel DNA sequencing technology, are required to estimate the input of selective sweeps relative to other factors, such as genetic drift and background selection.

1 Maynard Smith J. and Haigh J. (1974) The hitch-hiking effect of a favorable gene. Genet. Res. **23:** 23–35
2 Ohta T. and Kimura M. (1975) The effects of selected linked locus on heterozygosity of neutral alleles (the hitch-hiking effect). Genet. Res. **25:** 313–326
3 Stephan W., Wiehe T. H. E. and Lenz M. W. (1992) The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. Theor. Popul. Biol. **41:** 237–254
4 Akashi H. and Schaeffer S. W. (1997) Natural selection and the frequency distribution of 'silent' DNA polymorphism in *Drosophila*. Genetics **146:** 295–307
5 Charlesworth B., Morgan M. T. and Charlesworth D. (1993) The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303
6 Hudson R. R. and Kaplan N. L. (1995) Deleterious background selection and recombination. Genetics **141:** 1605–1617
7 Schmid K. J. and Tautz D. (1997) A screen for fast evolving genes from *Drosophila*. Proc. Natl. Acad. Sci. USA **94:** 9746–9750
8 McDonald J. H. and Kreitman M. (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. Nature **351:** 652–654
9 Sawyer S. A. and Hartl D. L. (1992) Population genetics of polymorphism and divergence. Genetics **132:** 1161–1176
10 Akashi H. (1999) Inferring the fitness effects of DNA mutations from polymorphism and divergence data: statistical power to detect directional selection under stationarity and free recombination. Genetics **151:** 221–238
11 Hudson R. R., Kreitman M. and Aguade M. (1987) A test of neural evolution based on nucleotide data. Genetics **116:** 153–159
12 Tajima F. (1989) Statistical method for testing the neutral mutation hypothesis. Genetics **123:** 585–595
13 Fu Y.-X. and Li W.-H. (1993) Statistical test of neutrality of mutations. Genetics **133:** 1289–1303
14 Simonsen K. L., Churchill G. A. and Aquadro C. F. (1995) Properties of statistical tests of neutrality for DNA polymorphism data. Genetics **141:** 413–429
15 Civetta A. and Singh R. S. (1995) High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in *Drosophila melanogaster* and *Drosophila virilis* group species. J. Mol. Evol. **41:** 1085–1095
16 Civetta A. and Singh R. S. (1998) Sex-related genes, directional sexual selection and speciation. Mol. Biol. Evol. **15:** 901–909
17 Civetta A. and Singh R. S. (1999) Broad-sense sexual selection, sex gene pool evolution and speciation. Genome **42:** 1033–1041
18 Nurminsky D. I., Nurminskaya M. V., De Aguiar D. and Hartl D. L. (1998) Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. Nature **396:** 572–575
19 Charlesworth B. and Charlesworth D. (1999) Scientific Correspondence. Nature **400:** 519–520
20 Nurminsky D. I. and Hartl D. L. (1999) Scientific Correspondence. Nature **400:** 520
21 Ting C. T., Tsaur S. C., Wu M.-L. and Wu C.-I. (1998) A rapidly evolving homeobox at the site of a hybrid sterility gene. Science **282:** 1501–1504
22 Ting C. T., Tsaur S. C. and Wu C.-I. (1999) Molecular evolution of a hybrid male sterility gene, *Odysseus*, and its twin brother, *Dunc-4*. Ann. Dros. Res. Conf. **40:** 23
23 Cirera S. and Aguade M. (1998) Molecular evolution of a duplication: the sex-peptide *(Acp70A)* gene region of

Drosophila subobscura and *Drosophila madeirensis*. Mol. Biol. Evol. **15:** 988–996
24 Hughes A. L. (1994) The evolution of functionally novel proteins after gene duplication. Proc. R. Soc. Lond. B **256:** 119–124
25 Aguade M. (1998) Diferent forces drive the evolution of the *Acp26Aa* and *Acp26Ab* accessory gland genes in the *Drosophila melanogaster* species complex. Genetics **150:** 1079–1089
26 Aguade M. (1999) Positive selection drives the evolution of the Acp29AB accessory gland protein in *Drosophila*. Genetics **152:** 543–551
27 Tsaur S. C., Ting C. T. and Wu C.-I. (1998) Positive selection driving the evolution of a gene of male reproduction, *Acp26Aa*, of *Drosophila*: II. Divergence versus polymorphism. Mol. Biol. Evol. **15:** 1040–1046
28 Karotam J., Boyce T. M. and Oakeshott J. G. (1995) Nucleotide variation at the hypervariable *Esterase 6* isozyme locus of *Drosophila simulans*. Mol. Biol. Evol. **12:** 113–122
29 Willett C. S. (2000) Evidence for directional selection acting on pheromone-binding proteins in the genus *Choristoneura*. Mol. Biol. Evol. **17:** 553–562
30 Metz E. C. and Palumbi S. R. (1996) Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein binding. Mol. Biol. Evol. **13:** 397–406
31 Metz E. C., Robles-Sikisaka R. and Vacquier V. D. (1998) Nonsynonymous substitutions in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. Proc. Natl. Sci. USA **95:** 10676–10681
32 Karn R. C. and Nachman M. W. (1999) Reduced nucleotide variability at an androgen-binding protein locus (*Abpa*) in house mice: evidence for positive natural selection. Mol. Biol. Evol. **16:** 1192–1197
33 Wyckoff G. J., Wang W. and Wu C.-I. (2000) Rapid evolution of male reproductive genes in the descent of man. Nature **403:** 304–309
34 Taylor M. F., Shen Y. and Kreitman M. E. (1995) A population genetic test of selection at the molecular level. Science **270:** 1497–1499
35 Blackhall W. J., Pouliot J. F., Prichard R. K. and Beech R. N. (1998) *Haemonchus contortus*: selection at a glutamate-gated chloride channel gene in ivermectin- and moxidectin-selected strains. Exp. Parasitol. **90:** 42–48
36 Blackhall W. J., Liu H. Y., Xu M., Prichard R. K. and Beech R. N. (1998) Selection at a P-glycoprotein gene in ivermectin- and moxidectin-selected strains of *Haemonchus contortus*. Mol. Biochem. Parasitol. **95:** 193–201
37 Yan G., Chadee D. D. and Severson D. W. (1998) Evidence for genetic hitchhiking effect asociated with insecticide resistance in *Aedes aegypti*. Genetics **148:** 793–800
38 Steinert M., Hentschel U. and Hacker J. (2000) Symbiosis and pathogenesis: evolution of the microbe-host interaction. Naturwissenschaften **87:** 1–11
39 Martinsohn J. T., Sousa A. B., Guethlein L. A. and Howard J. C. (1999) The gene conversion hypothesis of MHC evolution: a review. Immunogenetics **50:** 168–200
40 Yeager M. and Hughes A. L. (1999) Evolution of the mammalian MHC: natural selection, recombination and convergent evolution. Immunol. Rev. **167:** 45–58
41 Eanes W. F., Kirchner M. and Yoon J. (1993) Evidence for adaptive evolution of the *G6pd* gene in the *Drosophila melanogaster* and *Drosophila simulans* lineage. Proc. Natl. Acad. Sci. USA **90:** 7475–7579
42 Hudson R. R., Bailey K., Skaresky D., Kwiatowski J. and Ayala F. J. (1994) Evidence for a positive selection in the *Superoxide Dismutase* (*Sod*) region of *Drosophila melanogaster*. Genetics **136:** 1329–1340
43 Moriyama E. N. and Powell J. R. (1996) Intraspecific nuclear DNA variation in *Drosophila*. Mol. Biol. Evol. **13:** 261–277
44 Brookfield J. F. Y. and Sharp P. M. (1994) Neutralism and selectionism face up to DNA data. Trends Genet. **10:** 109–111
45 Aquadro C. F. (1997) Insights into the evolutionary process from patterns of DNA sequence variability. Curr. Opin. Genet. Dev. **7:** 835–840
46 Hartl D. L. and Taubes C. H. (1996) Compensatory near neutral mutations: selection without adaptation. J. Theor. Biol. **182:** 303–309
47 Shields D. C., Sharp P. M., Higgins D. G. and Wright F. (1988) 'Silent' sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. Mol. Biol. Evol. **5:** 704–716
48 Depaulis F. and Veulie M. (1998) Neutrality tests based on the distribution of haplotypes under an infinite-site model. Mol. Biol. Evol. **15:** 1788–1790
49 Kirby D. A. and Stephan W. (1995) Haplotype test reveals departure from neutrality in a segment of the *white* gene of *Drosophila melanogaster*. Genetics **141:** 1483–1490
50 Kirby D. A. and Stephan W. (1996) Multi-locus selection and the structure of variation at the *white* gene of *Drosophila melanogaster*. Genetics **144:** 635–645
51 Depaulis F., Brazier L. and Veuille M. (1999) Selective sweep at the *Drosophila melanogaster Suppressor of Hairless* locus and its association with the *In(2L)t* inverstion polymorphism. Genetics **152:** 1017–1024
52 Benassi V., Depaulis F., Meghlaoui G. K. and Veuille M. (1999) Partial sweeping of variation at the *Fbp2* locus in a West African population of *Drosophila melanogaster*. Mol. Biol. Evol. **16:** 347–353
53 Kreitman M. and Hudson R. R. (1991) Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics **127:** 565–582
54 Moore J. C., Jin H.-M., Kuchner O. and Arnold F. H. (1997) Strategies for the in vitro evolution of protein function: enzyme evolution by random recombination of improved sequences. J. Mol. Biol. **272:** 336–347
55 Parsch J., Tanda S. and Stephan W. (1997) Site-directed mutations reveal long-range compensatory interactions in the *Adh* gene of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **94:** 928–933
56 Parsch J., Braverman J. M. and Stephan W. (2000) Comparative sequence analysis and patterns of covariation in RNA secondary structures. Genetics **154:** 909–921
57 Cirera S. and Aguade M. (1997) Evolutionary history of the sex-peptide (*Acp70A*) gene region in *Drosophila melanogaster*. Genetics **147:** 189–197
58 Begun G. J. and Aquadro C. F. (1992) Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature **356:** 519–520
59 Langley C. H., McDonald J., Myashite N. and Aguade M. (1993) Lack of correlation between interspecific divergence and intraspecific polymorphism at the *suppressor of forked* region in *Drosophila melanogaster* and *Drosophila simulans*. Proc. Natl. Acad. Sci. USA **90:** 1800–1803
60 Zurovcova M. and Eanes W. F. (1999) Lack of nucleotide polymorphism in the *Y*-linked sperm flagellar dynein gene *Dhc-Yh3* of *Drosophila melanogaster* and *D. simulans*. Genetics **153:** 1709–1715
61 Stephan W. and Mitchell S. J. (1992) Reduced levels of DNA polymorphism and fixed between-population differences in the centromeric region of *Drosophila ananassae*. Genetics **132:** 1039–1045
62 Nachman M. W. (1997) Patterns of DNA variability at *X*-linked loci in *Mus* domesticus. Genetics **147:** 1303–1316
63 Nachman M. W. (1998) *Y* chromosome variation of mice and men. Mol. Biol. Evol. **15:** 1744–1750
64 Nachman M. W., Bauer V. L., Crowell S. L. and Aquadro C. F. (1998) DNA variability and recombination rates at *X*-linked loci in humans. Genetics **150:** 1133–1141
65 Liu F., Charlesworth D. and Kreitman M. (1999) The effect of mating system differences on nucleotide diversity at the *phosphoglucose isomerase* locus in the plant genus *Leavenworthia*. Genetics **151:** 343–357

66 Stephan W. and Langley C. H. (1998) DNA polymorphism in Lycopersicon and crossing-over per physical length. Genetics **150:** 1585–1593
67 Charlesworth D. and Charlesworth B. (1998) Sequence variation: looking for effects of genetic linkage. Curr. Biol. **8:** R658–661
68 Birky C. W. and Walsh J. B. (1988) Effect of linkage on rates of molecular evolution. Proc. Natl. Acad. Sci. USA **85:** 6414–6418
69 Charlesworth B. (1994) The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet. Res. **63:** 213–227
70 Barton N. H. (1995) Linkage and the limits to natural selection. Genetics **140:** 821–841
71 Kliman R. M. and Hey J. (1993) Reduced natural selection associated with low recombination in *Drosophila melanogaster*. Mol. Biol. Evol. **10:** 1239–1258
72 Comeron J. M., Kreitman M. and Aguade M. (1999) Natural selection on synonymous sites is correlated with gene length and recombination in *Drosophila*. Genetics **151:** 239–249
73 Moriyama E. N. and Powell J. R. (1998) Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. Nucleic Acids Res. **26:** 3188–3193
74 Sharp P. M. and Li W. H. (1986) Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. Nucleic Acids Res. **14:** 7737–7749
75 Munte A., Aguade M. and Segarra C. (1997) Divergence of the *yellow* gene between *Drosophila melanogaster* and *D. subobscura*: recombination rate, codon bias and synonymous substitutions. Genetics **147:** 165–175
76 Begun G. J. and Aquadro C. F. (1993) African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature **365:** 548–550
77 Labate J. A., Biermann C. H. and Eanes W. F. (1999) Nucleotide variation at the *runt* locus in *Drosophila melanogaster* and *Drosophila simulans*. Mol. Biol. Evol. **16:** 724–731
78 Tsaur S. C. and Wu C.-I. (1997) Positive selection and the molecular evolution of a gene for male reproduction, *Acp26Aa* of *Drosophila*. Mol. Biol. Evol. **14:** 544–549
79 Begun G. L. and Aquadro C. F. (1994) Evolutionary inferences from DNA variation at the *6-phosphogluconate dehydrogenase* locus in natural populations of *Drosophila*: selection and geographic differentiation. Genetics **136:** 155–171
80 Schug M. D., MacKay T. F. C. and Aquadro C. F. (1997) Low mutation rates of microsatellite loci in *Drosophila melanogaster*. Nature Genet. **15:** 99–102
81 Schlotterer C., Ritter R., Harr B. and Brem G. (1998) High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. Mol. Biol. Evol. **15:** 1269–1274
82 Schug M. D., Hutter C. M., Wetterstrand K. A., Gaudette M. S., Mackay T. F. C. and Aquadro C. F. (1998) The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. Mol. Biol. Evol. **15:** 1751–1760
83 Wiehe T. (1998) The effect of selective sweeps on the variance of the allele distribution of a linked multiallele locus: hitchhiking of microsatellites. Theor. Popul. Biol. **53:** 272–283
84 Michalakis Y. and Veuille M. (1996) Length variation of CAG/CAA trinucleotide repeats in natural population of *Drosophila melanogaster* and its relation to the recombination rate. Genetics **143:** 1713–1725
85 Schug M. D., Hutter C. M., Noor M. A. F. and Aquadro C. F. (1998) Mutatuon and evolution of microsatellites in *Drosophila melanogaster*. Genetica **102/103:** 359–367
86 Schloetterer C., Vogl C. and Tautz D. (1997) Polymorphism and locus-specific effects on polymorphism at microsatellite loci in natural *Drosophila melanogaster* populations. Genetics **146:** 309–320
87 Perlitz M. and Stephan W. (1997) The mean and variance of the number of segregating sites since the last hitchhiking event. J. Math. Biol. **36:** 1–23
88 Wiehe T. and Stephan W. (1993) Analysis of a genetic hitchhiking model, and its application to DNA polymorphism data from *Drosophila melanogaster*. Mol. Biol. Evol. **10:** 842–854
89 Slatkin M. and Wiehe T. (1998) Genetic hitchhiking in a subdivided population. Genet. Res. **71:** 155–160
90 Wu C.-I., Hollocher H., Begun D. J., Aquadro C. F., Xu Y. and Wu M. L. (1995) Sexual isolation in *Drosophila melanogaster*. A possible case of incipient speciation. Proc. Natl. Acad. Sci. USA **92:** 2519–2523
91 Capy P., Veuille M., Paillette M., Jallon J. M., Vouidibio J. and David J. R. (1999) Sexual isolation between genetically differentiated sympatric populations of *Drosophila melanogaster* in Brazzaville, Congo: the first steps of speciation? Ann. Dros. Res. Conf. **40:** 25
92 Begun G. J. and Aquadro C. F. (1995) Molecular variation at the *vermillion* locus in geographically diverse populations of *Drosophila melanogaster* and *D. simulans*. Genetics **140:** 1019–1032
93 Hamblin M. T. and Aquadro C. F. (1997) Contrasting patterns of nucleotide sequence variation at the *glucose dehydrogenase* (*Gld*) locus in different populations of *Drosophila melanogaster*. Genetics **145:** 1053–1062
94 Beech R. N. and Leigh-Brown A. J. (1989) Insertion-deletion variation at the *yellow-achaete-scute* region in two natural populations of *Drosophila melanogaster*. Genet. Res. **53:** 7–15
95 Martin-Campos J. M., Comeron J. P., Miyashita N. and Aguade M. (1992) Intraspecific and interspecific variation at the *y-ac-sc* region of *Drosophila simulans* and *Drosophila melanogaster*. Genetics **130:** 805–816
96 Wright F. (1990) The 'effective number of codons' used in a gene. Gene **87:** 23–39