J Mol Evol (1987) 26:47–58

Journal of
Molecular Evolution

DNA Turnover and the Molecular Clock

Gabriel A. Dover

Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, United Kingdom

Summary. Many detailed studies on the mechanisms by which different components of eukaryotic nuclear genomes have diverged reveal that the majority of sequences are seemingly not passively accumulating base substitutions in a clocklike manner solely determined by laws of diffusion at the population level. It appears that variation in the rates, units, biases, and gradients of several DNA turnover mechanisms are contributing to the course of DNA divergence. Turnover mechanisms have the potential to retard, maintain, or accelerate the rate of DNA differentiation between populations. Furthermore, examples are known of coding and noncoding DNA subject to the simultaneous operation of several turnover mechanisms leading to complex patterns of fine-scale restructuring and divergence, generally uninterpretable using selection and/or neutral drift arguments in isolation. Constancy in the rate of divergence, where observed over defined periods of time, could be a reflection of constancy in the rates and units of turnover. However, a consideration of the generally large disparity between rates of turnover and mutation reveals that DNA clocks, which would be independently driven by turnover in separate genomic components, would tend to be episodic. The utility of any given DNA sequence for measuring time and species relationships, like individual proteins, is proportional to the extent to which all contributing forces to the evolution of the sequence, internal and external, are understood.

Key words: Molecular clock — Molecular drive — DNA turnover — Genome evolution

Offprint requests to: G.A. Dover

Introduction

In this essay I explore the potential effects of widespread mechanisms of DNA turnover (unequal crossing-over, gene conversion, slippage, transposition, and RNA-mediated changes) on the DNA molecular clock. This is not a review of the evidence for the mechanisms themselves, but considers what is currently known about variation in their rates, units, biases, and locations, and illustrates, with examples, the extent to which they contribute to rates of DNA divergence.

Eukaryotic nuclear genomes, and some mitochondrial and chloroplast genomes, are subject to a variety of mechanisms of nonreciprocal exchanges of sequence (turnover), often operating simultaneously within the same DNA region, and which are potentially capable of accelerating, retarding, or maintaining constant the rate of DNA differentiation between populations. Furthermore, recent observations of nonrandom distributions of runs of polynucleotides (pure and cryptic DNA simplicity—see below) in most currently available coding and noncoding sequences indicate that it is unlikely that eukaryotic genomes consist of long stretches of kinetically complex "single-copy" DNA, which is the assumed substrate for the clocklike accumulation of mutations (Tautz et al. 1986).

The mechanisms of DNA turnover cause continual fluctuations in the copy-number of DNA motifs (which may vary in length from two to several thousand nucleotides) during the lifetime of an individual—a process that can lead to the dissemination (molecular drive) of mutations through a sexual population and affect observed levels of divergence (see below). Hence, neither conservation nor divergence of sequence should be understood solely in terms of selection and drift (Dover and Tautz 1986). Turnover mechanisms have been reviewed exten-

sively elsewhere (for example, see Davidson and Britten 1971; Smith 1973; Ohta 1980, 1983; Dover 1982, 1986a,b; Flavell 1982, 1986; Singer 1982; Arnheim 1983; several chapters in MacIntyre 1985).

Whose Hand on the Clock?

The debate about the existence of clocks and their rates of tick is enmeshed in the controversy over the relative contributions of different evolutionary processes involved with the spread of mutations through populations. So far as I am aware, there are three general processes for changing the genetic composition of a population: genetic drift, natural selection, and molecular drive. Genetic drift is generally considered to be responsible for the periodic diffusion of neutral or nearly neutral mutations through populations, giving rise to a molecular clock (Kimura 1983). Recently, attempts have been made to explain the clocklike accumulation of mutations and rate fluctuations by selection models incorporating changing environments and simple forms of epistasis (Gillespie 1986a,b,c). Both explanations assume that a mutation is a single event whose fate depends on the vagaries of drift or differences in fitness among individuals or gametes; although Gillespie briefly discussed at least one DNA-generated peculiarity such as the correlations in identification of neighboring bases (Blaisdell 1985) that would influence rate variation.

Molecular drive attempts to explain observed patterns and rates of DNA divergence starting from the peculiarities of DNA turnover, before resorting to considerations of drift and selection. It is a consequence of continual gains and losses of DNA that cause small but persistent patterns of non-Mendelian segregation, which, in turn, can promote the long-term spread of DNA variants through a sexual population (Dover 1982; Ohta and Dover 1984). Patterns of within-species homogeneity in multigene families and in internally repetitious genes testify to the effects of these internally driven processes, which can be both stochastic or unidirectional in activity. In addition, the fixation of alleles at nonrepetitive loci might be due in some measure to turnover between allelic sequences, for example, by gene conversion (see later) (Lamb and Helmi 1982).

Neither drift nor selection alone can explain the homogenization of a variant sequence among loci, or among repetitive units within a locus, throughout a population. In the case of nonrepetitive loci, however, criteria other than the distribution of variants need to be examined in order to assess the relative contributions of selection, genetic drift, or molecular drive. I do not believe that this can be done at present using the traditional assumptions that feed into mathematical population genetics. It requires an experimental investigation into the internal dynamics of the genome and the external dynamics of the ecology. For example, it has been shown mathematically that "weak" selection can overcome "strong" conversion bias (Walsh 1985, 1986). If this were the case in reality, homogeneity patterns would not be observed in most multigene families and internally repetitious genes. It would appear more useful to harness mathematics to reality and not vice versa. In general, it can be concluded with some confidence (for examples, see below) that some of the observed levels of accumulated DNA differences between taxa are a reflection of DNA turnover, in addition to the activities of natural selection and genetic drift.

A Testable Null Hypothesis

The difference between molecular drive and selection/drift is that the first is a consequence of the internal dynamics of irregular DNA behavior, while the latter two are based on the external dynamics of the population. This distinction provides a useful starting point for the sorting of their relative contributions to DNA evolution. One testable *null* hypothesis would be to assume initially that all observed levels of differentiation, whether erratic or clocklike, are due to the behavior of the genetic material. Failing this, then alternative explanations can be considered. This is not as difficult as it might seem because the evolutionary footprints of different DNA turnover mechanisms are directly observable in the genome, with appropriate analytical and comparative procedures (Ohno and Epplen 1983; Blaisdell 1985; Jeffreys et al. 1985; Karlin and Ghandour 1985; Karlin et al. 1985; Strachan et al. 1985; Eickbush and Burke 1986; Fitch 1986; Smithies and Powers 1986; Tautz et al. 1986, 1987). Some examples of what has been uncovered by such procedures are described below.

Molecular Coevolution: Starting at the Genome and Moving Outward

The genes coding for the eukaryotic 18S and 28S ribosomal RNAs are contained within a compound unit of spacers and genes that can be repeated several hundredfold, depending on species. There are several different turnover processes operating at different rates and periodicities within the unit (see later and Fig. 2). The net outcome of the continual gain and loss of units is that the rDNA family in each species has been homogenized for species-specific mutations. For reviews of rDNA and other DNA

families see Hood et al. (1975), Tartof (1975), Kedes (1979), Ohta (1980, 1983), Dover (1982), Flavell (1982), Arnheim (1983), Gerbi (1985), and Moss et al. (1985). Homogenization by molecular drive can account both for the gradual de novo amplification of a family (or its subfamilies) and for the subsequent spread of new variants through existing families (or subfamilies).

The rDNA family contains gene sequences that perform essential cellular functions. Additionally, there are spacer sequences required for the correct and efficient enhancement and promotion of transcription, and for rDNA replication and occasionally amplification (for reviews see Reeder 1984; Moss et al. 1985). What do we expect to happen to newly homogenized mutations within all such regions? There are, at least, two possibilities. The first is that at some critical threshold level in the replacement of an existing unit by a new variant, the effect on individual cellular functions leads to its selective elimination. However, the population dynamics of molecular drive is such that most individuals attain such a threshold within a short period of time relative to the long period of time it takes to homogenize the family throughout the population (Dover 1982; Ohta and Dover 1983, 1984). The second possibility is that such a gradual and cohesive spread of the new variant induces a compensatory change in the molecules that need to interact with the rDNA family or its products, in order to maintain efficient cellular functions. A clear example of molecular coevolution of this sort, interpreted as an interaction between molecular drive and natural selection, concerns the known biological incompatibilities between the rDNA promoters of one species and the polymerase I complex of another, when tested in in vitro transcription assays or in interspecific hybrids (Grummt et al. 1982; Arnheim 1983; Dover and Flavell 1984; Reeder 1984; Moss et al. 1985). The inception of this incompatibility cannot be explained by molecular drive alone. It requires the selection of available allelic variants of the polymerase I or cofactor genes (possibly "single-copy" genes), which are best able to interact with the changing spectrum of multiple promoters and enhancers (Dover and Flavell 1984). Recent detailed analysis of the promoters of four species of *Drosophila,* among which incompatibility occurs, shows that although the promoters have lower overall levels of divergence relative to the rest of the spacer, they are not refractory to mutation and to the spread of the mutations through the species (Dover and Tautz 1986; Tautz et al. 1987). Hence, they impose some compensatory coevolutionary change on the DNA-binding regulatory and transcriptional proteins. It is possible, also, to consider a more complex, and possibly a more realistic, involvement of selection along the lines developed by Zuckerkandl (1983) for the coevolution regulatory proteins and their receptor sequences.

In this particular example of molecular coevolution, our *null* hypothesis that all can be explained by the behavior of the genome has failed. Only by the additional and necessary consideration of the role played by selection can the observed phenomenon be explained. Although we have no *direct* evidence for natural selection, it becomes a logical necessity, pressurized as it is from forces within the genome and not only as a response to traditionally understood forces of the external ecology.

There are other examples of molecular coevolution (Dover and Flavell 1984; Gerbi 1985), not all of which are explicable in terms of an interaction between two independent and operationally distinct evolutionary processes. Nevertheless, it would seem that the first requisite steps to unraveling the functional significance of the mechanism and rate of DNA divergence in a given genomic component is the dissection of the DNA turnover mechanisms themselves. In the case of the rDNA, interspecific promoter divergence has not signified the accumulation of neutral mutations in a clocklike manner determined by laws of diffusion at the population level. Interestingly, the data reveal that promoter divergence is greater the longer two species have been separated (Moss et al. 1985; Tautz et al. 1987). This rough constancy in divergence rate is, however, a reflection of constancy in the rates of mutation and turnover within the region (see later).

Units, Biases, and Gradients of Turnover and DNA Divergence

In what ways can we expect the units, biases, and gradients of DNA turnover mechanisms to affect the degree and constancy of DNA differentiation? Some of these effects are obvious and easy to explain, in principle. The problems arise when a given DNA region is subject to several turnover mechanisms, all operating at different periodicities, rates, and biases. This is becoming increasingly the case as more regions are examined in detail.

Short and Long Units

DNA turnover mechanisms can operate at different periodicities, i.e., the unit of gain and loss may vary in length. Unequal crossing-over, transposition, gene conversion, and RNA-mediated exchanges may operate on long stretches of DNA up to several thousands of nucleotides. On the other hand, slippage-like mechanisms and telomeric-growth-like mechanisms generally involve short motifs of from 2 to 10 nucleotides in length. Gene conversion seems to be the most versatile in involving units ranging from

approximately 20 to 2000 or more nucleotides. The precise boundaries of conversion domains cannot be determined because the minimum and maximum lengths of the region involved with a nonreciprocal exchange between genes are fixed by the observed cluster of switched mutations and the positions of flanking molecular markers (for details see Baltimore 1981; Ollo and Rougeon 1983; Smithies and Powers 1986). Additionally, conversion can be polarized in that it may decrease along a gradient starting from some fixed site of initiation and stopping arbitrarily at any position (Rossignol et al. 1984). Conversion gradients have been observed directly in the human α-globin locus (Hess et al. 1984) and in the high-cysteine subfamily of the chorion gene superfamily in the silkmoth, *Bombyx mori* (Burke and Eickbush 1986; Eickbush and Burke 1986).

It is clear that the expected levels of divergence within any region of DNA would depend on the lengths of the units and gradients of turnover relative to the length of the region under comparison. If turnover units are smaller than the region under comparison, which may be a complete gene with its introns and exons, then high levels of variability would be observed between genes (Baltimore 1981) (Fig. 1a). In the case of gene conversion, the genes would be different mosaics of several smaller conversion domains, each of which would be, nevertheless, relatively homogeneous. Highly variable mosaic genes due to such disparities in unit lengths have been observed among, for example, the immunoglobulin (Bentley and Rabbitts 1983; Ollo and Rougeon 1983) and major histocompatibility genes (Steinmetz and Hood 1983), the globin genes (Goodman et al. 1984; Smithies and Powers 1986), the chorion genes (Burke and Eickbush 1986; Eickbush and Burke 1986), and others. The higher the rate of conversion the more variable are the genes. Hence, gene variability among the immunoglobulin genes does not signify a slow and evolutionarily insignificant rate of conversion, as has been suggested (Gojobori and Nei 1984). This would only be true if the unit of turnover had been larger than the genes under comparison (Fig. 1b).

Long turnover units embracing whole genes or substantial parts of genes and their flanking regions are also well documented. Indeed, the evidence for co-conversion and also conversion polarity was derived from classical tetrad analysis in fungi (Whitehouse 1982). Under such circumstances intraspecific heterogeneity of any region lying within the conversion domain is considerably reduced (Fig. 1b). Whether or not these processes also reduce interspecific heterogeneity (conservation) depends on the magnitude of any turnover bias that may be operating (see below). The resolution of seemingly paradoxical levels of within- and between-species variation, as exemplified by the *Drosophila Adh* genes (Kreitman and Aguadé 1986) and the major histocompatibility (H-2) locus in mice (Steinmetz et al. 1984) can be derived from such considerations (see below).

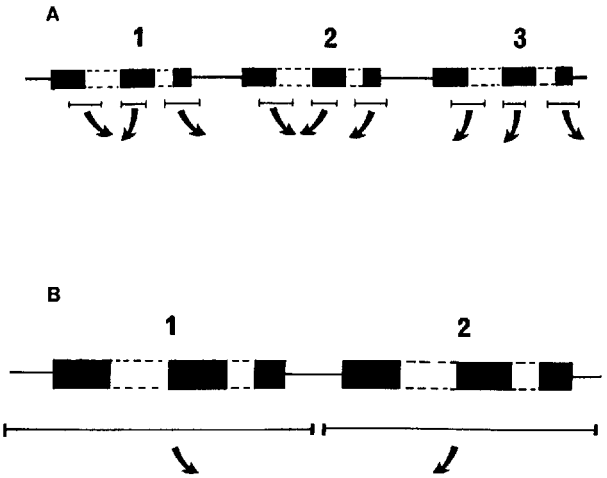

Fig. 1. Units of turnover and units of DNA under comparison. **a** Turnover unit smaller than comparative unit. Turnover mechanisms such as gene conversion and slippage may operate at periodicities below the level of the DNA region under comparison. At the level of the comparative unit, highly variable mosaic regions will ensue, although individual conversion or slippage units will be homogeneous. A gene triplication is shown (exons in black) with three small conversion domains operating between them. **b** Turnover unit larger than comparative unit. Turnover mechanisms such as gene conversion and unequal crossing-over may operate at periodicities above the level of the DNA region under comparison. All regions within the unit, for example exons, introns, nontranslated portions, and 5′ and 3′ flanking sequences, will show reduced levels of intraspecific variability relative to interspecific divergence. For examples, such as the *Drosophila Adh* locus and the mouse H-2 locus, see text. A gene duplication is shown, with each gene contained within a conversion domain. Thin lines indicate domains of gene conversion.

Simultaneously Operating Turnover Units

The simultaneous operation of several turnover mechanisms operating at different periodicities complicates the final outcome. Gene conversion, transposition, slippage, and unequal crossing-over are known to coexist within the chorion and rDNA multigene families (Coen et al. 1982a,b; Coen and Dover 1983; Goldsmith and Kafatos 1984; Eickbush and Burke 1986; Tautz et al. 1987). The evolutionary history of both these families, as determined by the internal mechanisms, can be monitored in great detail. For example, it is clear from the patterns of distribution of mutations among 28 of the 30 genes of the high-cysteine chorion subfamily of *B. mori* that slippage-like mechanisms are operating at different periodicities; first at a level of $(Cys\text{-}Gly)_2$, then at a higher level of a 30-bp unit of $(Cys\text{-}Gly)_2$ $(Cys\text{-}Gly\text{-}Gly)_2$. These tandemly arrayed

motifs have replaced other ancestral motifs of Gly-Tyr-Gly-Gly. Furthermore, gene conversion has been involved with the spread of the motifs to each of the carboxyl- and amino-terminal "arms" of the genes that surround a more conserved central domain, and to all of the 30 genes within the subfamily. Examination of the extent to which different mutations are shared by different subsets of the 30 genes reveals that mutations are at different stages of transition in the process of accumulation or elimination. Such transition stages during molecular drive have been revealed in two abundant noncoding families (the "360" and "500") in species of *Drosophila* (Strachan et al. 1985; see below). Hence, the levels of observed variation at any instant in time depend on the number of mutations expected to be in transition, which in turn depends on the rates of mutation, turnover, and the sizes of the family and population (Ohta 1980, 1983; Dover 1982; Ohta and Dover 1984).

Both the chorion gene subfamily and the *Drosophila* "360" and "500" families show that gene conversion or unequal crossing-over, respectively, operate freely within the families, in that the probability of occurrence between any two members is independent of their proximity or chromosomal position. Furthermore, in the case of the chorion genes there is a gene conversion gradient, starting at a region of repetitive simple-sequence motifs generated by a slippage-like mechanism. The initiation of the α-globin gradient also begins within a region of simple-sequence DNA (Hess et al. 1984). Hence, the products of one turnover mechanism influence the activities of another mechanism. The details of DNA evolution in the chorion family, as with many other gene families and single-copy genes, repay close attention (Goldsmith and Kafatos 1984; Eickbush and Burke 1986 and references therein). For a simplified and crude account of this system see Dover (1986b,c).

Complex Patterns in rDNA: A Case History

Complex patterns of DNA divergence are observed in the multiple compound rDNA unit because both slippage and unequal crossing-over at several different periodicities are operating simultaneously. A full description of these events is presented elsewhere (Tautz et al. 1987; Hancock and Dover 1987). Figure 2 illustrates some of the products of turnover as revealed in the dot-matrix comparisons in three species of *Drosophila*. It is clear that in *Drosophila melanogaster* there are three levels of subrepetition within the intergenic spacer (IGS formerly called NTS—see Coen and Dover 1983; Tautz and Dover 1986), at 95 bp, 330 bp, and 240 bp, respectively. Variation in copy-number indicates that unequal crossing-over occurs at the three levels and also at the level of the whole rDNA unit (Coen and Dover 1982, 1983; Coen et al. 1982a,b). This has led to the expected patterns of concerted evolution between the species. The 330-bp repeat is itself composed of a 240-bp unit plus 90 bp of the 95-bp unit, indicating that the periodicity of unequal crossing-over can vary and generate a longer unit that is composed of two different nonhomologous repetitive units. Fluctuations in the periodicities of unequal crossing-over are a common observation in many species of plants and animals (for original references see Davidson and Britten 1971; Southern 1975; Dover et al. 1982; Flavell 1982; Singer 1982; Miklos 1985; Cross and Dover 1987). The new unit can sometimes consist of a member of a family plus some adjacent single-copy DNA which in turn becomes repetitive on subsequent amplification (Flavell 1982, 1986). Repeated fluctuations in the unit length of turnover, whether by slippage, conversion, or unequal crossing-over, induce a continual scrambling of new and old DNA motifs. In the case of slippage this has been observed on a wide scale giving rise to cryptic DNA simplicity in many genic and nongenic sequences (Tautz et al. 1986).

Drosophila virilis and *Drosophila hydei* (Fig. 2) show two additional features not observed in *D. melanogaster*. The first is that a 220-bp unit has replaced the 240-bp unit, or vice versa, in the time separating the species. This is a further example of unequal crossing-over taking effect at a novel periodicity and spreading the structural novelty to all 500 spacers within all individuals of the species (see also Cross and Dover 1987). Secondly, the region complementary to the 95-bp repetition in *D. melanogaster* is less obviously repetitive. Instead, these two species show high levels of cryptic DNA simplicity due to the operation of slippage-like mechanisms on short DNA motifs below the level of the 95-bp repetition (Tautz et al. 1986, 1987; and see legend to Fig. 2). The data suggest, further, that the rate of slippage is faster than the rate at which unequal crossing-over at the 95-bp periodicity can homogenize new variants, leading to the gradual obliteration of the 95-bp repetition.

High levels of cryptic DNA simplicity, i.e., regions consisting of a finite number of different repetitive motifs scrambled among themselves (Tautz et al. 1986), are also observed in the so-called "expansion segments" (see Gerbi 1985) within the 28S genes of the rDNA unit (Hancock and Dover 1987). Interestingly, the sequence composition of the segments is uniform within a species, indicating that despite the activities of slippage, there is coevolution of segments. Whether this is due to (1) selection (as is probably the case concerning the molecular coevolution between diverging multiple promoters and the polymerase I complex—see above), (2) biased mutation, or (3) another turnover mechanism, such

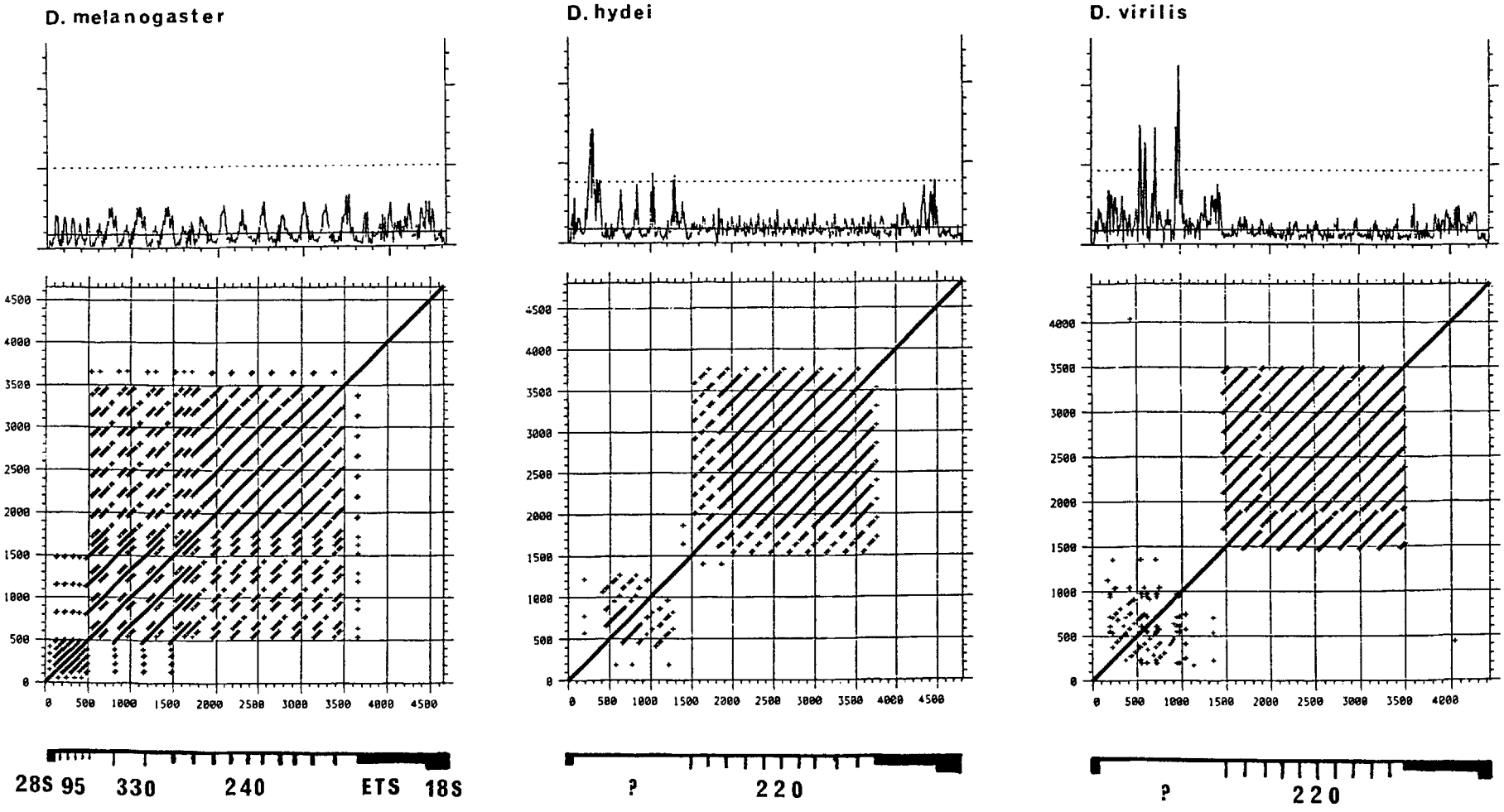

Fig. 2. Patterns of internal repetition and "simplicity" in the IGS spacer of the repetitive rDNA unit of three species of *Drosophila*. The figure shows the dot-matrix plots of DNA sequences which cover the region from the end of the 28S rRNA gene throughout the whole spacer until the start of the 18S rRNA gene. The sequences from each of the three species (*D. melanogaster, D. hydei,* and *D. virilis*) are compared with themselves such that the internal repetitive regions become apparent. The simplicity profile of each sequence is shown on top of the dot-matrix plots. See Tautz et al. (1986) for a description of the method used for obtaining the simplicity profile. The solid horizontal line represents the mean overall simplicity factor derived from 10 randomized sequences, each of 10 kb, with the same base composition as the natural sequence. The dotted horizontal line represents the position of the highest peak that any of the 10 × 10 kb of randomized sequence was able to muster. Some peaks from the natural sequence are above the dotted line. Most peaks of all three species are statistically significantly above the solid line. The complex pattern of internal repeats in *D. melanogaster* is caused by subrepeats (330 bp) which comprise various pieces of two other basic repeats, the 240-bp repeats in the center of the spacer and the 95-bp repeats at the 5′ end of the spacer. The 240-bp repeats are known to carry promoter duplications which might function as enhancers. A 220-bp repetition exists also in *D. virilis* and *D. hydei* and is clearly apparent in the dot-matrix plots. However, the region 5′ of the 220-bp array is not repetitive, but instead shows high simplicity. Slippage-generated simplicity has almost obliterated the arrays of repeats found in the equivalent regions in *D. melanogaster*. Slippage is operating on units that are smaller than the units of unequal crossing-over and at a seemingly faster rate (see text).

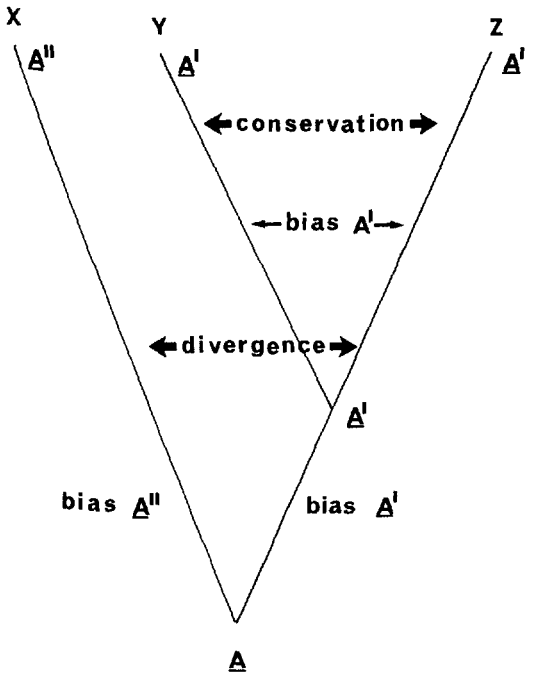

Fig. 3. Turnover bias and the conservation/divergence of sequences. An initial bias (for example, conversion bias) in favor of A′ in the lineage leading to species Y and Z and of A″ in the lineage leading to X accelerates the rate of divergence between lineages. The same conversion bias, operating subsequently in the bifurcating lineages leading separately to species Y and Z, retards the rate of divergence between the two lineages. The divergent or conservation consequences of bias depend on the phylogenetic point of comparison. A can be a "single-copy" gene (two alleles) or a multigene family $(A)_n$.

as gene conversion, in addition to slippage, is open to discussion. For the moment, it is mentally easier to accept that gene conversion has been transferring the cryptically simple products of slippage from one segment to another. This would be analogous to the way in which gene conversion has spread the new high-cysteine slippage-generated motifs to all members of the chorion subfamily in *B. mori* (see above).

Turnover Bias: Conservation and Divergence Are Two Sides of the Same Coin

Two of the turnover mechanisms can be biased in terms of the increase in probability that a particular variant sequence is amplified over the rest. These are transposition and gene conversion. The first can give rise to new families in its duplicative mode. Small but persistent biases in gene conversion have been observed by tetrad analysis in most examined genes in fungi (for reviews see Lamb and Helmi 1982; Whitehouse 1982), although the molecular basis for favoring one sequence over another is not known (see Rossignol et al. 1984). Duplicative transposition and biased gene conversion have received more attention than other turnover mechanisms because the former became the basis for the selfish DNA concept (Doolittle and Sapienza 1980; Orgel and Crick 1980), and because the earliest calculations on the effects of non-Mendelian behavior on the genetic structure of populations were based on biased conversion (Gutz and Leslie 1976; see Ohta 1980, 1983 and Dover 1982 for reviews). The latter also became wrongly identified as exclusively synonymous with molecular drive (see Dover 1986b). The other turnover mechanisms are not known to be biased although it can be assumed that if slippage and unequal crossing-over are homology-dependent mechanisms of recombination, then variant sequences might be less likely to be involved. In such circumstances a turnover mechanism would tend to feedback and conserve the first array of repeats that are arbitrarily amplified (Dover and Tautz 1986).

The first obvious point about turnover bias is that it can dramatically affect the rate of spread of a variant through a family and a population (Ohta 1980, 1983; Nagylaki and Petes 1982). The second, less obvious, point is that the same bias in favor of a given variant can be implicated in both divergence and conservation of sequences between species. They represent two sides of the same coin. This is illustrated in Fig. 3. An initial bias in favor of A′ over A in the lineage leading to species Y and Z can lead to rapid divergence between this lineage and that leading to species X. The continual operation of the same bias in subsequent bifurcations of this lineage would retard the rate of sequence divergence between species Y and Z because many new variants would be converted back to the favored A′ sequence. This would be true for both multiple copy and "single-copy" genes. The distinction between divergence and conservation is simply a reflection of the point in phylogenetic time at which our observations are being made.

Adh Divergence in Drosophila: A Reexamination

At this juncture we can reexamine the *Drosophila Adh* data (Kreitman and Aguadé 1986). In *D. melanogaster* there is an overall similar level of polymorphism between a 4.0-kb region of seemingly meaningless DNA flanking the 5′ end of the *Adh* gene and the nontranslated regions, introns, and exons of the transcription unit itself. If just the "effectively" silent codon positions are taken into account then there is a 5.7 higher level of silent polymorphism in the exons than in the 4.0-kb region. Comparison between *D. melanogaster* and the closely related sibling species *Drosophila simulans* shows, however, that the 4.0-kb region is more than twice as divergent as the transcription unit, and that

the levels of "effective" silent divergences are now about the same. As pointed out by Kreitman and Aguadé (1986), differences in levels of intraspecific divergence between adjacent regions are unexpected on a neutral model of DNA evolution, given the interspecific comparisons. One explanation, offered tentatively by Kreitman and Aguadé, is that the level of intraspecific polymorphism in the 4.0-kb region is seemingly low only by comparison with the excessively high and possibly selectively balanced polymorphism in the gene. An alternative explanation is that the two-fold higher level of interspecific variability in the 4.0-kb region is due to slippage-like mechanisms, which are responsible for the observed high (70%) A+T richness, frequent runs of homonucleotides, and frequent small deletions and additions in this region. The additional operation of gene conversion, embracing the 4.0-kb region, would then tend to reduce intraspecific variability, making this region roughly equivalent in variation to the structural locus. Combinations of slippage and conversion are known to be operating in other genomic components (see above).

It is important to note that mechanisms that reduce variability within a species (homogeneity) do not reduce interspecific variability. Homogenization and conservation are not the same thing, as evidenced by the concerted evolution patterns in the case of multiple copy sequences. New variants can be arising continuously and be homogenized separately in each species. In general, it appears that a neutral clocklike basis of DNA differentiation is inadequate to explain the *Adh* observations; the resolution of the problem, however, depends on whether we wish to start with the dynamics of selection or with the dynamics of the genome.

The application of genomic dynamics to other seemingly paradoxical cases of widely differing levels of divergence in adjacent DNA regions can also offer testable alternative explanations. For example, Steinmetz et al. (1984) have uncovered a relatively monomorphic region of 170 kb at the H-2 locus in mice which is known to contain introns, exons, and a lot of intergenic nonsense DNA, probably of varying degrees of repetition. In contrast, the adjacent region contains the classical Class I and II genes which are highly polymorphic. Steinmetz and colleagues concluded that the monomorphic region arose because it is basically refractory to mutation for reasons linked to chromatin structure. Although we should keep an open mind regarding this possibility (the genome has not yet revealed all its secrets), nevertheless, it is probably more realistic to suggest that gene conversion is reducing heterogeneity levels in this region. This would entail a long conversion domain starting at the E_β boundary between the two regions which is known to be a recombinational hotspot (Steinmetz et al. 1984). Small gene conversion domains are known to be responsible for the high levels of polymorphism in the Class I and II genes (Strachan et al. 1984; Flavell et al. 1985; for review see Dover and Strachan 1987), and longer gene conversion domains could possibly be involved with the reduced levels of variation in the adjacent region. Differences in levels of heterozygosity are tied to the relationship between the unit lengths of conversion and the lengths of the regions under comparison, as described above.

Rates of Turnover, DNA Clocks, and Phylogenies

Rates of turnover can affect levels of DNA divergence in a number of ways.

Sequence Homogeneity but Copy-Number Heterogeneity

Most rate measurements have been made on unequal crossing-over, gene conversion, and transposition. Although they range from as slow as 10^{-7} per generation to several percentages per generation, the mode is around 10^{-4}. Unusually fast rates of transposition are only found in permissive conditions of hybrid dysgenesis in insects; and gene conversion can occur at every cell generation during yeast mating-type switching. The more usual rate of unequal crossing-over of 10^{-4} is found in the rDNA family in *Drosophila* and other species (Arnheim 1983; Coen and Dover 1983), and in the human hypervariable minisatellite DNA family (Jeffreys et al. 1985). In the latter example, it is occurring at a rate of 10^{-4}/kb/generation; although some extremely variable loci are known with rates up to 30×10^{-4} in humans and mice (Wong et al. 1986; Jeffreys et al. 1987).

Ohta has demonstrated (1980, 1983) that the faster the rate of turnover the more pronounced are the levels of homogeneity. However, in the case of unequal crossing-over, the faster the turnover rate the higher is the variation in copy-number of family members. Hence, there are two levels of heterogeneity to be considered, which have been confounded on occasion. One is the sequence identity between member repeats and the other is variability in copy-number. Individual DNA "fingerprints," based on the human minisatellite DNA (Jeffreys et al. 1985) and on the *Drosophila* rDNA (Coen et al. 1982a; Coen and Dover 1983), are due, in the main, to copy-number variation only. Detailed analysis of sequence identity among repeats taken from within particular arrays of minisatellites underline, on a fine scale, the expected relationship between turn-

over rates and homogeneity levels (Jeffreys et al. 1985). Arrays which are generally invariant in copy-number show high levels of interrepeat sequence variability; whereas arrays that have high copy-number variation are more invariant at the sequence level. This is as predicted (Ohta 1980; Dover 1982). It represents a sort of miniaturized intraspecific concerted evolution.

Variation in Large DNA Families

At rates of 10^{-4} per generation it can be expected that large gene families of several hundreds to thousands of members would reveal mutations at different stages of homogenization, like ripples in a pool. The sequences of two abundant noncoding families (the "360" and "500"), shared between up to seven of the sibling species of the *melanogaster* species subgroup, reveal all the stages of transition in the spread, by unequal crossing-over, of variant repeats through the family and population (Strachan et al. 1985). Partially shared mutations by only a subset of repeats have been observed in other satellite DNA families (for review see Miklos 1985); and in the hypervariable minisatellite family (Jeffreys et al. 1985; Wong et al. 1986).

It is interesting to note that in the case of the "360" and "500" families, an approximately 2:1 ratio of transversions to transitions indicates that all nucleotide positions in all member repeats (of which there are 3000–10,000 depending on species) are seemingly free to mutate to any other nucleotide, and that the variant repeat can gain or lose in copy-number in a purely stochastic fashion. Similar random distributions of partially shared mutations among the member genes of the high-cysteine chorion subfamily in *B. mori* (described above) indicate that conversion-swapping is occurring on a free basis. The stochastic nature and freedom of operation of turnover in some instances has important implications for any clocklike rate of DNA divergence.

Constancy in Turnover Rates and the Molecular Clock

If the rates of unequal crossing-over or gene conversion remain constant, and if all member genes are equally likely to participate in nonreciprocal exchange, then a clocklike accumulation of mutations will probably ensue. The clock, however, will be "episodic" rather than regular, analogous to the potential "episodic" clock generated by selection (Gillespie 1986a,b). That is, bursts of substitutions are followed by long periods of stasis. The molecularly driven "episodic" clock emerges from the approximately average two orders of magnitude discrepancy between the rates of base substitution and turnover (Dover 1982). Turnover continues unabated whether there are variant repeats or not within a family. This is analogous to the turnover of new for old banknotes. Only rarely will a new mutation arise, the probability of which depends on the size of the family. At the initial stages of low variant frequency there is a high probability that the variant will be lost rather than increase in copy-number at the expense of the others. If, in the unlikely event that it does increase in frequency, then its further progression will improve rapidly, relative to the time it would take for another mutation to follow suit. The small number of variant repeats in transition in the "360" and "500" DNA families, at any given moment among approximately 10,000 flies from which the repeats were sampled, supports the above episodic dynamics of spread of variant repeats. A new banknote design introduced into an existing process of turnover rapidly spreads at the expense of the old; although in this case we are dealing with a biased system.

Neither the rate of spread nor the number of mutations observed to be in transition would affect the expectation that the number of fixed differences between any two species (i.e., fully homogenized variants) should be proportional to the amount of time elapsed between the species. This too is observed among the seven *Drosophila* species (Strachan et al. 1985). A phylogenetic reconstruction of interspecific distances based on levels of fixed differences in each family produces a phylogeny which is concordant with phylogenies based on several other diverse biological characters. Concordance between phylogenetic trees is observed despite the fact that the two families are seen to be evolving at different rates. Concordance emerges from constancy in rates within each family in all lineages and is not disturbed by the difference in rates between the families.

Conclusions

The examples given above illustrate that DNA turnover mechanisms can give rise to complex patterns of divergence in different genomic components. In general, it would appear that no cosmic clock with a uniform rate of divergence can be predicted from the behavior of the majority of genomic sequences. The same point has been reached by different investigators starting from different perspectives (Li et al. 1985; Palmer 1985; Britten 1986; Fitch 1986; Vawter and Brown 1986). Different regions of nuclear, mitochondrial, and chloroplast genomes are evolving at different rates within a species, and the same region can evolve at different rates in different species lineages. Ensuing variation in tick rates and

concordance or not of phylogenies based on separate components depend on the internal dynamics of the genome, overlaid by the more traditionally understood population and generation time parameters. If the genomic parameters are relatively constant over a given evolutionary span then phylogenetic relationships can be established. This does not mean, however, that the genome consists predominantly of long stretches of single-copy DNA accumulating substitutions in a clocklike manner only in accordance with laws of diffusion external to the genome.

Measurements of the melting temperature of reannealed DNA (heteroduplex mismatch) between supposedly isolated "single-copy" DNA from different species (Sibley and Ahlquist 1984) are not necessarily measuring the accumulation of point mutational differences. First, the measurements themselves are technically sensitive to the amount of residual percentage homology between two species, which reflects the comings and goings of sequences by DNA turnover. Secondly, and importantly, the so-called single-copy DNA can be expected to be composed of innumerable scrambled permutations of repetitive motifs from two to tens of bases in length, which are not separable using standard "single-copy" isolation techniques (Moyzis et al. 1981; Flavell 1982; Tautz et al. 1986). Such continual fine-scale restructuring of the genome by slippage-like mechanisms and RNA-mediated intermingling of repeats (Jagadeeswaran et al. 1981) represents a greater source of variation (the extent of which depends on the rates and modes of turnover) than that introduced by point mutations. On these grounds, it is not surprizing that there is considerable controversy over crude "single-copy" DNA–DNA hybridization studies and the support they are assumed to give to concepts of universal time-dependent clocks (Wilson et al. 1977; Sibley and Ahlquist 1984; Templeton 1985; Britten 1986; Ruvolo and Smith 1986).

DNA clocks are occasionally observable but, as I have tried to show, they are probably a reflection of the dynamics of DNA turnover mechanisms in defined components, as well as reflecting the diffusion dynamics of neutral mutations. A clock based on neutral drift alone would need to be operating within a region of DNA, perhaps silent codon changes in genes (Kimura 1977) or pseudogenes (Li et al. 1981), which can be shown exhaustively, by the right analytical and comparative procedures, not to have experienced any turnover from slippage to conversion during the time interval under investigation. Both natural selection and molecular drive, in their multifarious ways, can disturb the ticking of any clock generated by diffusion at the population level. To understand the full range of internal and external forces influencing DNA divergence, each sequence needs to be examined case by case. For the moment we can only guess at what lies beneath the tip of the iceberg.

References

Arnheim N (1983) Concerted evolution of multigene families. In: Nei M, Koehn RK (eds) Evolution of genes and proteins. Sinauer, Sunderland MA, pp 38–61

Baltimore D (1981) Gene conversion: some implications for immunoglobulin genes. Cell 24:592–594

Bentley DL, Rabbitts TH (1983) Evolution of immunoglobulin V_K genes: evidence indicating that recently duplicated human V_K sequences have diverged by gene conversion. Cell 32:181–190

Blaisdell BE (1985) Markov chain analysis finds a significant influence of neighbouring bases on the occurrence of a base in eukaryotic nuclear DNA sequences both protein-coding and noncoding. J Mol Evol 21:278–288

Britten RJ (1986) Rates of DNA sequence evolution differ between taxonomic groups. Science 231:1393–1398

Burke WD, Eickbush TH (1986) The silkmoth late chorion locus. I. Variation within two paired multigene families. J Mol Biol 190:343–356

Coen ES, Dover GA (1982) Multiple Pol I initiation sequences in the rDNA of spacers of *Drosophila melanogaster*. Nucleic Acids Res 10:7017–7026

Coen ES, Dover GA (1983) Unequal exchanges and the coevolution of X and Y rDNA arrays in *D. melanogaster*. Cell 33:849–855

Coen ES, Strachan T, Dover GA (1982a) The dynamics of concerted evolution of rDNA and histone gene families in the *melanogaster* species subgroup of *Drosophila*. J Mol Biol 158: 17–35

Coen ES, Thoday JM, Dover GA (1982b) Rate of turnover of structural variants in the rDNA gene family of *Drosophila melanogaster*. Nature 295:564–568

Cross NCP, Dover GA (1987) A novel arrangement of sequence elements surrounding the rDNA promoter and its spacer duplications in tsetse species. J Mol Biol 195:63–74

Davidson EH, Britten RJ (1971) Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. Q Rev Biol 66:111–138

Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601–603

Dover GA (1982) Molecular drive: a cohesive mode of species evolution. Nature 299:111–117

Dover GA (1986a) The spread and success of non-Darwinian novelties. In: Karlin S, Nevo E (eds) Evolutionary processes and theory. Academic Press, New York, pp 199–237

Dover GA (1986b) Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. Trends Genet 2:159–165

Dover GA (1986c) How to drive an egg. Trends Genet 2:300–301

Dover GA, Flavell RB (1984) Molecular coevolution: DNA divergence and the maintenance of function. Cell 38:623–624

Dover GA, Strachan T (1987) Molecular drive in the evolution of the immune superfamily of genes: the initiation and spread of novelty. In: Kelsoe G, Schulze DH (eds) Evolution and vertebrate immunity. University of Texas Press, Austin, pp 15–34

Dover GA, Tautz D (1986) Conservation and divergence in multigene families: alternative to selection and drift. In: Clarke BC, Robertson A, Jeffreys AJ (eds) Evolution of DNA. Phil Trans R Soc Lond [Biol] 312:275–289

Dover GA, Brown SDM, Coen ES, Dallas J, Strachan T, Trick

M (1982) The dynamics of genome evolution and species differentiation. In: Dover GA, Flavell RB (eds) Genome evolution. Academic Press, London, pp 343–374

Eickbush TH, Burke WD (1986) The silkmoth late chorion locus. II. Gradients of gene conversion in two paired multigene families. J Mol Biol 190:357–366

Fitch WM (1986) The estimate of total nucleotide substitutions from pairwise differences is biased. Phil Trans R Soc Lond [Biol] 312:317–324

Flavell RA, Allen H, Huber B, Wake C, Widera G (1985) Organisation and expression of the MHC of the C57 Black/10 mouse. Immunol Rev 84:29–40

Flavell RB (1982) Sequence amplification, deletion and rearrangement: major sources of variation during species divergence. In: Dover GA, Flavell RB (eds) Genome evolution. Academic Press, London, pp 301–324

Flavell RB (1986) Repetitive DNA and chromosome evolution in plants. Phil Trans R Soc Lond [Biol] 312:227–249

Gerbi SA (1985) Evolution of ribosomal DNA. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum, New York, pp 419–517

Gillespie JH (1986a) Natural selection and the molecular clock. Mol Biol Evol 3:138–155

Gillespie JH (1986b) Statistical aspects of the molecular clock. In: Karlin S, Nevo E (eds) Evolutionary processes and theory. Academic Press, New York, pp 255–272

Gillespie JH (1986c) Variability of evolutionary rates of DNA. Genetics 113:1077–1091

Gojobori T, Nei M (1984) Concerted evolution of the immunoglobulin V_H gene family. Mol Biol Evol 1:195–212

Goldsmith MR, Kafatos FC (1984) Developmentally regulated genes in silkmoths. Annu Rev Genet 18:443–487

Goodman M, Koop BF, Czelusniak J, Weiss ML, Slightom JL (1984) The eta-globin gene: its long evolutionary history in the beta-globin gene family of mammals. J Mol Biol 180:803–823

Grummt I, Roth E, Paule MR (1982) Ribosomal RNA transcription *in vitro* is species specific. Nature 296:173–176

Gutz H, Leslie JF (1976) Gene conversion: a hitherto overlooked parameter in population genetics. Genetics 83:861–866

Hancock J, Dover GA (1987) Molecular coevolution amongst 'expansion segments' of eukaryotic 26S/28S RNA gene in *Drosophila* and *Xenopus*. J Mol Evol (in press)

Hess JF, Schmid CW, Shen CKJ (1984) A gradient of sequence divergence in the human adult alpha-globin duplication units. Science 226:67–70

Hood L, Campbell JH, Elgin SCR (1975) The organization, expression and evolution of antibody genes and other multigene families. Annu Rev Genet 9:305–353

Jagadeeswaran P, Iorget BG, Weissmann SM (1981) Short interspersed repetitive DNA elements in eukaryotes: transposable DNA elements generated by reverse transcription of RNA Pol III transcripts? Cell 26:141–142

Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable 'minisatellite' regions in human DNA. Nature 314:67–73

Jeffreys AJ, Wilson V, Kelly R, Taylor BA, Bulfield G (1987) Mouse DNA 'fingerprints': analysis of chromosome localization and germ-line stability of hypervariable loci in recombinant mouse strains. Nucleic Acids Res 15:2823–2836

Karlin S, Ghandour G (1985) Comparative statistics for DNA and protein sequences: single sequence analysis. Proc Natl Acad Sci USA 82:5800–5804

Karlin S, Ghandour G, Foulser DE (1985) DNA sequence comparisons of the human, mouse and rabbit immunoglobulin kappa gene. Mol Biol Evol 2:35–45

Kedes LH (1979) Histone genes and histone messengers. Annu Rev Biochem 48:837–870

Kimura M (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature 267:275–276

Kimura M (1983) The neutral theory of evolution. Cambridge University Press, Cambridge, pp 1–367

Kreitman ME, Aguadé M (1986) Excess polymorphism at the *Adh* locus in *Drosophila melanogaster*. Genetics 114:93–110

Lamb BC, Helmi S (1982) The extent to which gene conversion can change allelic frequencies in populations. Genet Res 39: 199–217

Li WH, Gojobori T, Nei M (1981) Pseudogenes as a paradigm of neutral evolution. Nature 292:237–239

Li WH, Luo CC, Wu CI (1985) Evolution of DNA sequences. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum, New York, pp 1–84

MacIntyre R (ed) (1985) Molecular evolutionary genetics. Plenum, New York, p 641

Miklos GLG (1985) Localized highly repetitive DNA sequences in vertebrate and invertebrate genomes. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum, New York, pp 241–313

Moss T, Mitchelson K, de Winter R (1985) The promotion of ribosomal transcription in eukaryotes. Oxford Survey of Eukaryotic Genes 2:207–250

Moyzis RK, Bonnet J, Li DW, Ts'o POP (1981) An alternative view of mammalian sequence organization: short repetitive sequences organized into scrambled tandem clusters. J Mol Biol 153:871–896

Nagylaki T, Petes TD (1982) Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. Genetics 100:315–337

Ohno S, Epplen JT (1983) The primitive code and repeats of base oligomers as the primordial protein-encoding sequence. Proc Natl Acad Sci USA 80:3391–3395

Ohta T (1980) Evolution and variation in multigene families. Springer, Berlin

Ohta T (1983) On the evolution of multigene families. Theor Popul Biol 23:216–240

Ohta T, Dover GA (1983) Population genetics of multigene families that are dispersed into two or more chromosomes. Proc Natl Acad Sci USA 89:4079–4083

Ohta T, Dover GA (1984) The cohesive population genetics of molecular drive. Genetics 108:501–521

Ollo R, Rougeon F (1983) Gene conversion and polymorphism: generation of mouse immunoglobulin gamma 2a chain alleles by differential gene conversion by gamma 2b chain gene. Cell 32:515–523

Orgel LE, Crick FHC (1980) Selfish DNA: the ultimate parasite. Nature 284:604–606

Palmer JD (1985) Evolution of chloroplast and mitochondrial DNA in plants and algae. In: MacIntyre R (ed) Molecular evolutionary genetics. Plenum, New York, pp 131–216

Reeder RH (1984) Enhancers and ribosomal gene spacers. Cell 38:349–351

Rossignol JL, Nicolas A, Hamza H, Langin T (1984) Origins of gene conversion and reciprocal exchange in *Ascobolus*. Cold Spring Harbor Symp Quant Biol 49:13–21

Ruvolo M, Smith TF (1986) Phylogeny and DNA–DNA hybridization. Mol Biol Evol 3:285–289

Sibley CG, Ahlquist JE (1984) The phylogeny of the hominoid primates, as indicated by DNA–DNA hybridisation. J Mol Evol 20:2–15

Singer MF (1982) Highly repeated sequences in mammalian genomes. Int Rev Cytol 76:67–112

Smith GP (1973) Unequal crossover and the evolution of multigene families. Cold Spring Harbor Symp Quant Biol 38: 507–513

Smithies O, Powers PA (1986) Gene conversions and their

relation to homologous chromosome pairing. Phil Trans R Soc Lond [Biol] 312:291–302

Southern EM (1975) Long range periodicities in mouse satellite DNA. J Mol Biol 94:51–69

Steinmetz M, Hood L (1983) Genes of the major histocompatibility complex in mouse and man. Science 222:727–734

Steinmetz M, Malissen M, Hood L, Orn A, Maki RA, Dastournikoo GR, Stephan D, Bibb E, Romaniuk R (1984) Tracts of high and low sequence divergence in the mouse major histocompatibility complex. EMBO J 3:2995–3003

Strachan T, Sodoyer R, Damotte M, Jordan BR (1984) Complete nucleotide sequence of a functional Class I HLA gene, HLA-3: implications for the evolution of HLA genes. EMBO J 3:887–893

Strachan T, Webb DA, Dover GA (1985) Transition stages during molecular drive in multiple copy DNA families in *Drosophila*. EMBO J 4:1701–1708

Tartof KD (1975) Redundant genes. Annu Rev Genet 9:355–387

Tautz D, Dover GA (1986) Transcription of the tandem array of ribosomal DNA in *Drosophila* does not terminate at any fixed point. EMBO J 5:1267–1273

Tautz D, Trick M, Dover GA (1986) Cryptic simplicity in DNA is a major source of genetic variation. Nature 322:652–656

Tautz D, Tautz C, Webb DA, Dover GA (1987) Evolutionary divergence of promoters and spacers of the rDNA family of four *Drosophila* species: implications for molecular coevolution in multigene families. J Mol Biol 195:525–542

Templeton AR (1985) The phylogeny of the hominoid primates: a statistical analysis of the DNA–DNA hybridization data. Mol Biol Evol 2:420–433

Vawter L, Brown WM (1986) Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. Science 234:194–196

Walsh JB (1985) Interaction of selection and biased gene conversion in a multigene family. Proc Natl Acad Sci USA 82: 153–157

Walsh JB (1986) Selection and biased gene conversion in a multigene family: consequences of interallelic bias and threshold selection. Genetics 112:699–716

Whitehouse HLK (1982) Genetic recombination. Understanding the mechanisms. Wiley, New York

Wilson AC, Carlson SS, White TJ (1977) Biochemical evolution. Annu Rev Biochem 46:573–639

Wong Z, Wilson V, Jeffreys AJ, Thein SL (1986) Cloning a selected fragment from a human DNA 'fingerprint': isolation of an extremely polymorphic minisatellite. Nucleic Acids Res 14:4605–4616

Zuckerkandl E (1983) Topological and quantitative relationships in evolving genomes. In: Hélène C (ed) Structure, dynamics, interactions and evolution of biological macromolecules. Reidel, Dordrecht, pp 395–412

Received April 28, 1987