PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

# Conservation and Divergence in Multigene Families: Alternatives to Selection and Drift

G. A. Dover and D. Tautz

*Phil. Trans. R. Soc. Lond. B* 1986 **312**, 275-289
doi: 10.1098/rstb.1986.0007

*hil. Trans. R. Soc. Lond.* B **312**, 275–289 (1986)
*rinted in Great Britain*



# Conservation and divergence in multigene families: alternatives to selection and drift

By G. A. Dover and D. Tautz

*Department of Genetics, University of Cambridge, Cambridge CB2 3EH, U.K.*

It is generally assumed that conservation and divergence of DNA signify function (selection) and no function (drift), respectively. This assumption is based on the view that a mutation is a unique event on a single chromosome, the fate of which depends on selection or drift. Knowledge of the rates, units and biases of widespread mechanisms of non-reciprocal DNA exchange, in particular within multigene families, provides alternative explanations for conservation and divergence, notwithstanding biological function. Such mechanisms of DNA turnover cause continual fluctuations in the copy-number of variant genes in an individual and, hence, promote the gradual and cohesive spread of a variant gene throughout a family (homogenization) and throughout a population (fixation). The dual processes (molecular drive) of homogenization and fixation are inextricably linked. Data are presented of the expected stages of transition in the spread of variant repeats by molecular drive in some non-genic families of DNA, seemingly not under the influence of selection. When a molecularly driven change in a given gene family is accompanied by the coevolution (mediated by selection) of other DNA, RNA or protein molecules that interact with the gene family then biological function is observed to be maintained despite sequence divergence. Conversely, the mechanics of DNA turnover and a turnover bias in favour of ancestral sequences can dramatically retard the rate of sequence change, in the absence of function. Examples of the maintenance of function by molecular coevolution and conservation of sequences in the absence of function, are drawn mainly from the rDNA multigene family.

## DNA turnover and classical evolutionary theory

t is generally assumed that observed levels of conservation and divergence of DNA sequences re a reflection of the presence and absence of biological function, respectively. Functionally onserved sequences are considered to be constrained in their evolution by natural selection nd, conversely, that non-functional diverged sequences have been free to accumulate nutations within a population by accidents of genetic drift. Both processes of evolutionary hange are based on the premise that a DNA mutation is a solitary event at a unique position n a single chromosome of one individual, obeying the laws of Mendelian inheritance. ccordingly, this restricted view of the behaviour of DNA makes of selection and drift the sole neans for increasing the frequency of a mutation in a population.

The solitary nature and rarity of mutation, coupled to an earlier generation's unavoidable gnorance of the dynamic behaviour of DNA, gave powerful theoretical support to the concept of atural selection during the formulation of the neo-Darwinian synthesis in the 1930s (Huxley 942). The realization that the Mendelian laws of inheritance, when transformed to whole opulations of infinite size and idealized free mating, cannot in themselves lead to the spread f one allelic variant in favour of another, made of selection an absolute prerequisite for adaptive

evolutionary change. R. A. Fisher's definition of natural selection as a means of achieving a highly improbable shift in the average genotype of a population (Fisher 1930), succinctly encapsulates the theoretical necessity for selection in a background of a continuous and stochastic recycling of chromosomes by sex. That infinite, sexually unconstrained, populations do not exist and that non-random sampling at the level of gametes and individuals (genetic drift) can increase the probability of gaining or losing an allele, out of proportion to that expected in Mendelian populations, has not altered the classical view that DNA and its mutations are simply the passive static providers of the grist for the evolutionary mill.

## DNA turnover and molecular drive

Today our knowledge of DNA and its evolution are radically different for ample studies on the molecular behaviour of eukaryote genomes over the past decade have revealed that, in addition to mutation, a vast proportion of the DNA is subject to a variety of non-reciprocal exchanges that can transfer mutational variants from one locus to another and from one chromosome to another. All such mechanisms of non-reciprocal exchange (gene conversion, unequal exchange, transposition, slippage replication and RNA-mediated transfers) induce rare but persistent non-Mendelian patterns of segregation which may promote the spread of mutations through a population over long periods of time. These mechanisms operate in both single-copy genes and multigene families, although their effects are more easily noticeable in multigene families. It is conceivable that strict Mendelian genes and stable Mendelian populations in long-term Hardy–Weinberg equilibria do not exist, except when generally observed over short periods of time and in small numbers of progeny. It is of significance that the mechanisms of DNA turnover proceed generally at rates of $10^{-2}$–$10^{-5}$ per generation (Dover 1982; Coen & Dover 1983) that lie between the mutation rate and the rate at which chromosomes are continually randomized between generations. Hence, at the level of the chromosome, the Mendelian laws of segregation and stable Mendelian populations are relevant and accurate; at the level of DNA they are not, except on a short-term observational basis. The degree to which the behaviour of the chromosomes and the behaviour of DNA are out of synchrony is of long-term evolutionary significance.

The process by which the genotypic composition of a population can be changed, as a consequence of the persistent non-Mendelian effects of DNA turnover, is called molecular drive (Dover 1982; Dover *et al.* 1982). It can promote, like selection, an improbable long-term shift in the mean genotype of a population. The shift takes place, however, not as a result of the selective sorting of allelic variants as a consequence of their effects on individual fitness, but as a consequence of the internal dynamics of DNA turnover.

## Concerted evolution and molecular drive: pattern and process

Direct and circumstantial evidence for mechanisms of DNA turnover in single-copy, multigenic and non-genic families is abundant and is not reviewed here. The accompanying papers by Engels; Flavell; Smithies & Powers; Bodmer (all this symposium) describe some of the evidence in the globin, immunoglobulin, histocompatibility and other multigene and non-genic families. Further information can be found in Dover (1982, 1986*b*), Coen *et al.* (1982*a*, *b*), Coen & Dover (1983), Ohta (1980, 1983).

One universal observation in the majority of repetitive families (or their subfamilies), is a




PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

much higher genetic similarity between two member genes from within a species than there is between genes from two different species. Relatively high levels of family (or subfamily) homogeneity within a species (a phenomenon known as concerted evolution) cannot be explained solely by drift or selection. This is because in the case of drift each locus of a gene family would be expected to become fixed for a different mutational variant, and not as observed for the same variant. Under drift alone levels of within and between species family variation are expected to be the same. Similarly, selection could fix a given variant gene at a given locus; however, it is generally inconceivable that selection would fix each locus at a time for the precise same variant, or indeed fix any locus at any time in non-genic families, often consisting of several thousand copies of seemingly junk repetitive DNA. The time required to achieve species diagnostic patterns of homogeneity would be far too long if selection were to wait for each locus to mutate to the same condition before fixation (Dover *et al.* 1982; Dover 1982; Brown & Dover 1981; Coen *et al.* 1982*b*; Strachan *et al.* 1982, 1985).

In contrast to selection and drift, any internal mechanism of non-reciprocal transfer of information between loci, even in the absence of bias for a particular variant, dramatically raises the probability that a variant spreads to most loci of a family (or subfamily) in most individuals. The dual processes (molecular drive) of family homogeneity and population fixation are inextricably linked. This is in contrast to the 'double diffusion' process outlined by Kimura & Ohta (1979) in which homogenization and fixation are considered to be separate processes. It is, however, not satisfactory simply to consider a process of homogenization within a single chromosome lineage, followed by the accidental fixation of that lineage in the population by genetic drift. Concerted evolution is an observed pattern of homogeneity between genes that are often distributed on several different chromosomes. Hence it is important to take into account non-reciprocal exchanges both between homologous chromosomes and, where necessary, also between non-homologous chromosomes. Molecular drive is a population genetics *process*, emanating from non-Mendelian consequences of DNA turnover throughout the karyotype, that is invoked to explain the widespread *pattern* of variation known as concerted evolution.

The degree of intraspecific genetic identity between member genes, relative to that observed between related species, depends on the rates of family turnover, the size of the family, the number of chromosomes on which member genes are located and the size of the population (Dover 1982; Ohta 1980; Ohta & Dover 1983, 1984; Trick & Dover 1984). Each family is different in these respects and a blanket intraspecific total homogeneity is neither expected nor observed.

## Rates, units and biases of DNA turnover

To assess the rates, units and biases of DNA turnover and to evaluate their contribution to evolutionary change (relative to that contributed by selection and drift) it is necessary to investigate such features independently, in so far as this is possible. In the following sections we present and interpret the data on:

(i) the expected stages of transition during homogenization by molecular drive, in two abundant non-genic families, evolving in parallel in a group of sibling species of *Drosophila*;

(ii) the maintenance of the biological function of rDNA and other multigene families despite their continuous divergence through cycles of homogenization for new variants;

(iii) the facility to conserve sequences by mechanisms of turnover, in the absence of function and selection.




PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

### Transition stages during molecular drive

If molecular drive is a process for changing the genotypic composition of a population then families with a large number of repeats should contain all stages of transition during the homogenization of different variant repeats: like ripples in a pool. To investigate this expectation, two abundant non-genic families (of 3000–10000 members per individual, depending on species) have been isolated from the pooled DNA of several thousand individuals of each of seven sibling species of the *melanogaster* species subgroup of *Drosophila* (Strachan *et al.* 1985). The 360 family (average repeat length is 360 base pairs) is in blocks of tandemly arrayed repeats which are confined to the X chromosome (Brutlag 1981; Barnes *et al.* 1978; Strachan *et al.* 1982; Peacock *et al.* 1977); and the 500 family (average length is 500 base pairs) is distributed in blocks on all four pairs of chromosomes (Strachan *et al.* 1982). Up to 15 repeats of each family from each species have been cloned and sequenced. By this means an analysis can be made of the distribution of mutations, irrespective of the loci, chromosomes or individuals from which the repeats have been sampled.

The majority of mutations are single base substitutions distributed randomly through the lengths of the repeat units, with a transversion:transition ratio of approximately 2:1 in each family. These two features indicate that a base, at any nucleotide position in the repeats, can mutate with equal probability to any other base, and that each randomly produced variant can be driven through the population with equal probability. The analysis of the degree of variation within and between species reveals an order of magnitude greater variation between the consensus sequence of each species than exists between the individual repeats of each species. The 360 and 500 families have within- and between-species levels of variation of 2.7% and 32.4%, and 2.8% and 34.6%, respectively. Pairs of species that are known to be more closely related from a range of biological criteria (*D. yakuba* and *D. teissieri*, *D. simulans* and *D. mauritiana*) show a relatively smaller between-species difference. In the case of the latter pair of species, the within- and between-species variations are approximately the same. Hence, at the level of resolution of the comparison of species consensus sequences, concerted evolution is only observed between species of greater phylogenetic distance, leading to an apparent absence of homogenization in closely related species.

To circumvent this paradox, it is necessary to compare directly all available sequenced repeats of two species rather than the average sequence of the consensus. By this means the distribution of a given mutation at any given nucleotide position considered independently, can be monitored across all repeats shared between two species. Figure 1 depicts, schematically, the nature of this exercise. Each column is a nucleotide position from within a repeat, with open and closed circles depicting ancestral and derived bases, respectively. Ancestral bases are taken, parsimoniously, as those that are present in the majority of all pooled repeats of a pair of species. The six observed classes of mutational distribution represent all stages of transition, from class 1 (no homogenization of any degree in either species), through class 5 (species-specific homogeneity patterns: concerted evolution) to class 6 (the beginning of a second round of homogenization at a given nucleotide position in one or other species).

It is interesting to note that for the two pairs of species under comparison (*D. mauritiana* versus *D. simulans*; and *D. yakuba* versus *D. teissieri*) 96% of the variants in each family fall into these six classes and no others. This is to say that at the majority of nucleotide positions only two bases are observed at any one time, with one species remaining unchanged while considerable

homogenization for the new base is occurring in the other species. Additionally, the number of nucleotide positions that fall into transition stages 3, 4 and 5 (representing from 20–80% homogenization and fixation) are relatively few (table 1). Both these features indicate that the rate of spread by molecular drive is fast relative to the mutation rate, (see Strachan *et al.* (1985) for further details). Furthermore, from the data presented in table 1 it is apparent that: (i) there are instances of the complete spread of variant repeats in both families between the most recently diverged pair *D. mauritiana* and *D. simulans*, even though the total amount of difference is small; (ii) the higher number of fixed positions in each family between *D. yakuba* and

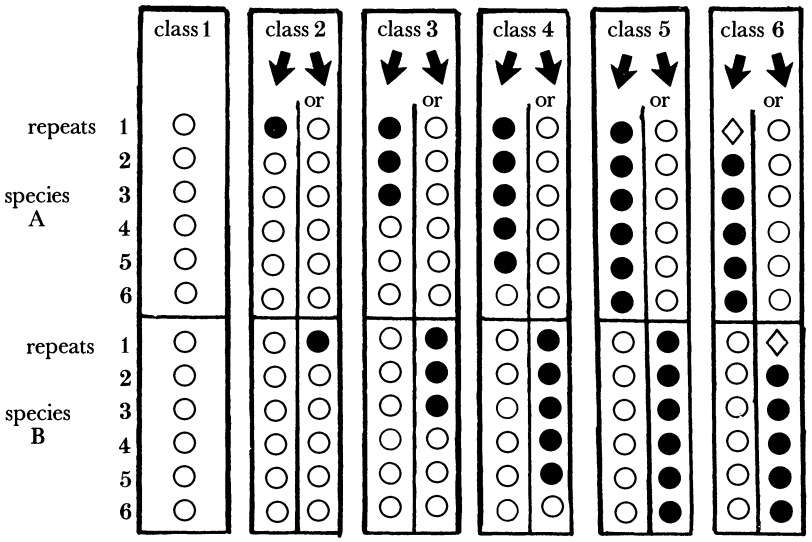


FIGURE 1. Graphic representation of transition stages during the spread of new mutations. Classes 1–6 represent the patterns of distribution of mutations at individual nucleotide positions across repeats 1–6 in two species A and B. Symbols ○, ● and ◇ represent nucleotide differences which can be A, T, G or C. Note that 96% of the positions that differ between the two pairs of species under comparison (see table 1) fall into these six classes. In classes 1–5 only two bases are found at a given position across all clones of a pair of species.

TABLE 1. PARTITIONING OF THE DISTRIBUTION OF MUTATIONS AT INDIVIDUAL NUCLEOTIDE POSITIONS ACROSS ALL CLONES IN THE 360 AND 500 FAMILY SEQUENCES OF TWO CLOSELY RELATED PAIRS OF SPECIES (SEE ALSO FIGURE 1)

| | | | 360 family | | 500 family | |
|---|---|---|---|---|---|---|
| | | | *mauritiana* versus *simulans* | *yakuba* versus *teissieri* | *mauritiana* versus *simulans* | *yakuba* versus *teissieri* |
| classes of mutational distribution | | | | | | |
| class | species A | species B | | | | |
| 1 | $N_1$ only | $N_1$ only | 278 (76.8) | 113 (56.4) | 348 (72.9) | 424 (76.1) |
| 2 | $N_1$ only | $N_1 > N_2$ | 72 (19.9) | 21 (10.3) | 104 (21.8) | 74 (13.3) |
| 3 | $N_1$ only | $N_1 = N_2$ | — | 2 (1.0) | — | 3 (0.6) |
| 4 | $N_1$ only | $N_2 > N_1$ | 6 (1.7) | 4 (2.0) | 2 (0.4) | 10 (1.8) |
| 5 | $N_1$ only | $N_2$ only | 5 (1.4) | 55 (27.5) | 2 (0.4) | 38 (6.8) |
| 6 | $N_1$ only | $N_2 > N_3$ | 1 (0.3) | 9 (4.5) | 1 (0.2) | 9 (1.6) |

$N_1$ represents any nucleotide and species A and B are interchangeable. See text and figure 1 for an explanation of the classes of mutational distribution that represent transition stages. Figures refer to the absolute number of positions falling into each class. Figures in brackets represent percentages obtained by dividing the absolute numbers of positions within each class by the number of nucleotide positions available for comparison. Those mutational events that could not be assigned an unambiguous location were excluded from the analysis. For example in the cases of runs of a particular nucleotide, deletions and insertions representing the same nucleotide are topographically ambiguous.




PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

*D. teissieri* than exists between *D. mauritiana* and *D. simulans*, indicates that the spread of mutant repeats by molecular drive is constant over time; (iii) the lower than expected number of fixed mutations (classes 5 and 6) in the 500 family (relative to the 360 family) in both pairs of species is indicative of a slower absolute rate of spread in this family than in the 360 family. The slower rate of change in the 500 family (located on all chromosomes) might be a consequence of physical constraints on interchromosomal homogenization, which are not affecting the 360 family repeats confined to the X chromosomes.

## Molecular coevolution: DNA divergence and the maintenance of function

Species-specific patterns of multigene family homogeneity are indicative of continual rounds of homogenization for new variant repeats, leading inevitably to interspecific divergence. A key consideration concerns the maintenance of biological function during this process of divergence. What might be the role of selection? To understand the nature of the interaction between drive and selection, and the time at which it might take place, we need to consider first the pattern and dynamics of population change under molecular drive.

The large disparity in rates between mutation, turnover and the randomization of chromosomes by sex (see §2) leads to a specific prediction that at any given transition stage there will be a small population variance relative to the two extremes of no homogenization and full homogenization for a new variant (Dover 1982; Ohta & Dover 1984). This is to say, that with rates of turnover ranging from $10^{-2}$ to $10^{-5}$ per generation, it is improbable that one or a few individuals become homogenized in advance of the rest of the interbreeding population. Hence, molecular drive can effect a slow, cohesive shift in the mean composition of a gene family in a population, without the generation of large differences between individuals in the ratio of old to new variants, at any given generation.

The maintenance of a relatively high genetic similarity between individuals throughout a period of transformation is of significance to our understanding of the interaction with selection. The size of the variance in the ratio of old to new variants per individual depends on many parameters (Dover 1982; Ohta & Dover 1983, 1984). It is unlikely that in the case of a large multigene family of tens to hundreds of members, all producing the identical product, that one mutant member would have an immediate effect on phenotype of sufficient magnitude as to confer a selective differential. It is conceivable, however, that at some stage of transition in the spread of a variant within a functionally important region, that selection would recognize the accumulating effect on phenotype in a cohesively evolving group of individuals and interact with any subsequent spread.

The degree of interaction between selection and drive would depend on the nature of the essential molecular recognitions between a gene family (or its products) and other molecules involved with a given cellular function. A change in one part could lead to a readjustment in another: that is, a molecular coevolution of parts ensuring the maintenance of biological function (Dover & Flavell 1984).

To illustrate molecular coevolution we draw on data largely derived from the rDNA multigene family. The applicability of these modes of interaction to other gene families should be clear.





PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

### The maintenance of function by molecular coevolution in the rDNA family

Figure 2*a* depicts an rDNA unit of *Drosophila*, which is typical of many genera of plants and animals (reviewed in Long & Dawid 1980; Fedoroff 1979; Glover 1981; Reeder 1984; Dover & Flavell 1984). It consists mainly of two genes (responsible for the 18 *S* and 28 *S* ribosomal RNAs) separated by a spacer. The compound unit is repeated from tens to several thousand times in tandem arrays, often on several different chromosomes, depending on species. Spacers generally consist of a tandem array of subrepeats that contain duplications of promoters and enhancers required for efficient transcription starting in the spacer immediately preceding the 5′ end of the 18 *S* gene. A typical pattern of concerted evolution is revealed by the presence of species-specific mutations within each of the spacer subrepeats of each rDNA unit in all tandem arrays (Arnheim 1983; Coen *et al.* 1982*b*). Detailed investigation in several species clearly show that unequal exchanges at the two levels of subrepeats and of whole rDNA units are taking place, within and between chromosomes (Tartof 1984; Fedoroff 1979; Coen *et al.* 1982*a*, *b*; Coen & Dover 1983). Hence, the dynamics of change in the rDNA family is complicated by the existence of one set of repeats nested within another set and both subject to two interlocking levels of turnover.

One recent and surprising observation in insects, mammals, amphibia and protozoa is an interspecific incompatibility between the RNA polymerase I complex of one species and the rDNA unit of another, when assayed in *in vitro*, and some *in vivo*, heterologous transcription systems (Grummt *et al.* 1982; Mishima *et al.* 1982; Miesfeld & Arnheim 1984; Skinner *et al.* 1984; reviewed in Reeder 1984; Dover & Flavell 1984). What might be the evolutionary pressure for the rapid divergence of the genes coding for *Pol* I and its cofactors, given the critically important role for ribosomal RNA in prokaryotes and eukaryotes?

A model of molecular coevolution envisages this pressure to emanate primarily from the continual homogenization of new variants in the multiple promoter–enhancer regions that are recognized by *Pol* I, its cofactors and other proteins involved with transcription regulation (Dover & Flavell 1984; Coen 1983). The slow and cohesive mode of population change (see §6) would provide the time and the relaxed conditions for selection to increase the frequency of available alleles of the *Pol* I and other genes, which are more efficient at transcribing the slowly increasing sets of new promoters–enhancers.

It is unlikely that the pressure for coevolution emanates first from the selection of new alleles of *Pol* I and the cofactor genes, to be followed by a change in the whole spectrum of enhancers and promoters. An individual faced suddenly with such new alleles might be at a selective disadvantage, rather than the converse, with respect to the several thousand promoters and enhancers that are no longer efficiently recognized. Furthermore, as outlined in §2, it is difficult to envisage how all loci of promoters and enhancers could achieve genetic identity via selection alone, unless we assume that the turnover mechanisms themselves are under selective control. A large body of experimental evidence in rDNA and other multigene families indicates that the mechanisms of turnover are continuous and proceed at their own prescribed rates, independently of selection.

Although we envisage molecular coevolution to arise primarily out of the spreading effects of non-reciprocal exchanges in gene families, there would clearly be situations in which selection itself might reinforce the spread of a variant to maintain an adequate number of compatible




PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

promoters and enhancers. Hence the extent of the interplay between natural selection and molecular drive could vary at different stages during the change in sequence composition of a multigene family. Furthermore, if there are multiple genes for polymerase I and its cofactors then there is the reciprocal opportunity emanating from these families to influence the rDNA family.

If the model of molecular coevolution between *Pol* I and its promoters is correct then new variants must be observed to be homogenized in the key promoter regions and not just the surrounding irrelevant sequences. In figure 2*b*, *c*, we present the data on the sequences of the true gene promoters (at the start of transcription) and the internal spacer promoters, of three *Drosophila* species. The region around transcription initiation (from about −55 to +30) is only partly conserved between all three species; although the similarities between *D. virilis* and *D. hydei* are much higher than each is to *D. melanogaster*, in keeping with the phylogenetic relationships of the three species. A second partly conserved region appears in the region −116 to −100. Complete divergence between all three species occurs from position +35 and throughout the external transcribed spacer (ETS), for at least 300 b.p.

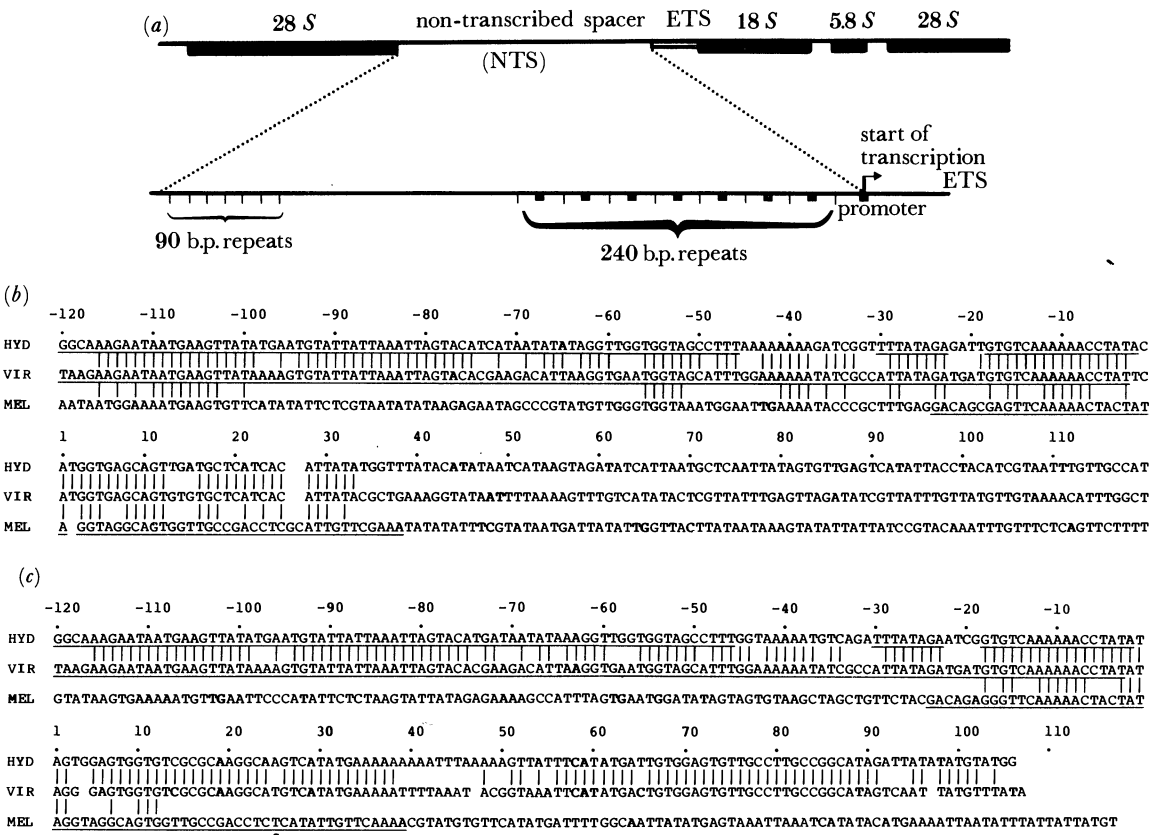


FIGURE 2. Comparison of the rDNA promoter regions and the 230 and 240 b.p. repeats from *D. hydei*, *D. virilis* and *D. melanogaster*.

(*a*) Structure of the *D. melanogaster* rDNA repeat unit. The spacer region has been enlarged. The promoter duplications are indicated with the black boxes. Abbreviations: ETS is the external transcribed spacer; 5.8S, 18 *S* and 28 *S* are the genes for the respective rRNAs.

(*b*) Promoter regions; position 1 marks the transcriptional start in *D. melanogaster*. Small insertions and deletions have been introduced to optimize alignment. The regions that are duplicated in the 230 and 240 b.p. repeats in each species are underlined.

(*c*) 230 and 240 b.p. repeats. The sequences are aligned to optimize homology around the sites of the promoter duplication. These regions are also underlined for each species (compare (*b*)).

As in *D. melanogaster*, there are duplications of promoter sequences in the 230 base pair (b.p.) subrepeats of *D. virilis* and *D. hydei*, although only the region upstream of the initiation site is duplicated. In *D. virilis* the complete homology starts immediately upstream of the presumptive point of transcription initiation, thus making the real promoter part of the 230 b.p. repeats. This is in contrast to *D. melanogaster*, where an approximate 150 b.p. gap separates the promoter from the 240 b.p. repeats. In *D. hydei* this gap is present, but it is only 10–15 b.p. long (compare the underlined regions in figure 2*b*).

It is clear that the promoters of *D. virilis* and *D. melanogaster* (between which *Pol* I–rDNA incompatibility occurs, (Kohorn & Rae 1982)) are not refractory to mutational change and to the spread of such mutants. Furthermore, there are high levels of intraspecific identity between spacer promoters and gene promoters, indicating that all promoters throughout the chromosomal arrays are part of one homogenizing pool of repeats.

The comparison of the 230 and 240 b.p. repeats (figure 2*c*) reveals a surprisingly high conservation between *D. virilis* and *D. hydei*, while the homology with *D. melanogaster* is confined to the region of the promoter duplication.

We have also analysed the spacer region beyond the end of the 28 *S* gene and found complete divergence between all three species (not shown). However, the region contains strong S1-sensitive single-stranded DNA sites in all three species, potentially involved with transcription termination. (Franz *et al.* 1985).

The various different levels of divergence and conservation in the compound rDNA unit are shown in figure 3. Conservation generally decreases (see regions 1–4 in figure 3) as we move from the genes (4) to the promoters and enhancers (3) to the sequences flanking the promoters and enhancers in each subrepeat (2) and finally to the spacer sequences and ETS that flank the array of subrepeats (1). Are these different levels of divergence and conservation a consequence of different degrees of constraint on molecular coevolution in the various regions or are they a consequence of differential rates and biases of turnover at the two levels of subrepeats and repeats? The extent to which turnover dynamics can retard and accelerate the rate of sequence evolution is described in §8.

In figure 3 we illustrate schematically how constraints on molecular coevolution might be the basis for the observations. The 18 *S* and 28 *S* genes are relatively but not absolutely conserved; although it is known that rRNA secondary structure is well conserved across Kingdoms (see Gerbi *et al.* 1982; Noller 1984). It is clear that despite divergence at the level of DNA (albeit at a low level), the maintenance of rRNA structure and function has required the coevolution of compensatory changes in one part of the molecule as a response to changes in another. Indeed, the most divergent regions of the genes are those that correspond to the RNA regions involved with double-stranded stems of single-stranded RNA loops, where coevolution is expected to be maximal if structure is to be maintained. The single-strand loops, by contrast, are highly conserved in sequence, presumably as a consequence of severe constraints imposed by the molecular interactions between these regions and the tens of ribosomal proteins involved with ribosome biogenesis. Some such proteins, binding to both single- and double-stranded RNA are known to be very highly conserved. It is important to remember that mutations would arise in such regions and begin to spread through the family and the population, in that the gene regions are part of the whole rDNA unequal exchange turnover unit. At some point of transition, however, selection has not tolerated most mutations in the DNA regions corresponding to the single-strand RNA loops, due to an apparent impossibility for coadaptation between many diverse molecules.




PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

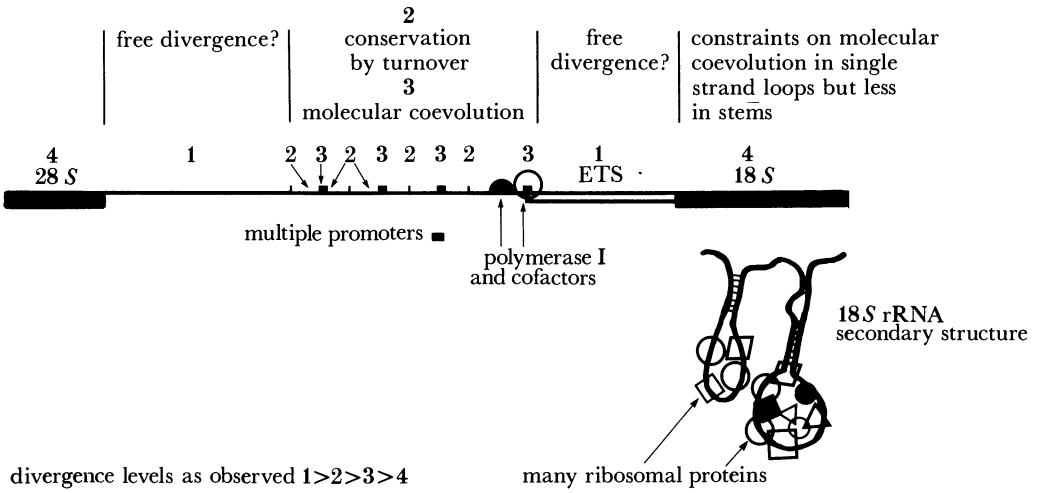


FIGURE 3. Regions of conservation and divergence in the rDNA. The drawing of the rDNA unit is simplified and not to scale. Abbreviations are the same as in figure 2. See text for further details.

In contrast to the genes, the observed higher levels of divergence between species promoters is a reflection of a greater tolerance to molecularly driven changes in that selection can promote the spread of appropriate coadapted allelic variants of the much fewer proteins involved with the biology of transcription. However, the observed differences between promoters and their immediate flanking sequences in each subrepeat are more difficult to explain. An explanation is not easily at hand in that the intrasubrepeat flanking sequences (see region 2 in figure 3) are less divergent than the remainder of the spacer sequences (1) and the ETS (1) that lie outside the subrepeats. It could be that such small differences in region 2 reflect, as before, levels of biological function and the ensuing constraints on coevolution imposed by the inflexibility of interaction with the products of other genes. This would imply that the 220 b.p. of each subrepeat are intimately interacting with proteins. It is unlikely, however, that such large regions of DNA are covered with regulatory proteins. Additionally, region 2 has diverged more than region 3 (which lies in the same subrepeat units) in the more distant comparisons between *D. melanogaster* and either *D. hydei* or *D. virilis*. It is possible therefore that the relatively high levels of sequence conservation in region 2 between *D. hydei* and *D. virilis* are a reflection of the dynamics of turnover *per se* (compare §8).

## DNA TURNOVER AND DNA CONSERVATION

Conservation of DNA sequences between species is usually ascribed to the action of selection preventing the spread of new variants. Furthermore, unexpected conservation of regions of DNA that might begin in the middle of a gene and extend out for some considerable distance into the 5′ or 3′ flanking sequences (and hence are not easily explicable in terms of selection on function), have been interpreted as regions of low mutation *per se* (for example, in the *Adh* single-copy gene in *D. melanogaster* (Kreitman 1983; Bodmer & Ashburner 1984)). It is more likely, however, that the contiguous conservation of arbitrary bits of genes and their flanking sequences is a reflection of the dynamics of turnover, which, in the case of single-copy genes, could occur between the two allelic states. In such cases, and in contrast to what is observed




PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

in the 360 and 500 family (see §5), the unit of turnover and the gene as a unit of function need not to be equivalent in length.

We envisage two ways in which turnover might contribute to the conservation of sequences, independently of selection.

### (*a*) *Turnover bias and conservation of sequence*

Mechanisms of turnover can cause either stochastic fluctuations in the copy-number of variant genes or a persistent biased increase in favour of one. Circumstantial evidence for both stochastic and directional modes is available from gene families (Dover 1982), although the relative contribution of each to the overall process of molecular drive is difficult to ascertain in many cases.

A bias in gene conversion is well documented in single-copy genes (for reviews see Whitehouse 1983; Fogel *et al.* 1981; Lamb & Helmi 1982); and the dramatic effects on the rate and probability of fixation of a favoured variant in multigene families has been studied (for reviews see Ohta 1980, 1983; Nagylaki & Petes 1982; Ohta & Dover 1984). Although the precise molecular basis of conversion bias is not understood, nevertheless it must be of sufficient persistence to allow for the complete spread of a variant in large gene families. Such persistence could be the basis for sequence conservation, in that after the favoured variant has become homogenized and fixed, it might retain this favoured position in subsequent conversions with newly arising variants (Dover 1982; R. Britten, personal communication). Hence, the persistence of the same bias for the same progenitor variant in two diverging species would greatly retard the rate of change in the favoured sequence. It is known that conversion bias is not an absolute property of a sequence but that a given favoured sequence can become disfavoured in conversions with other variants, depending on the genotypic composition of an organism (Lamb & Helmi 1982; Whitehouse 1983). The extent of conservation and the rate of retardation in a given sequence due to bias would depend, therefore, on the frequencies and magnitude of changes in the direction of bias.

In the case of the conservation of units of DNA that either (i) begin in a gene and end outside of it (such as the *Adh* gene in *Drosophila* and the γ-globin genes in humans: see Smithies & Powers, this symposium) or (ii) span many kilobases of DNA that embrace several genes and their intergenic sequences, (such as the β-globin cluster in primates or sections of the mouse H-2 histocompatibility locus (Barrie *et al.* 1981; Steinmetz *et al.* 1984)), it is possible that a low but persistent bias in favour of the extant sequence would convert out most new allelic variants. The length of the conserved domain would depend on the signals involved with initiation and termination of conversion. These are largely unknown, although there are pointers in several gene families that simple sequence motifs may be involved (see Smithies & Powers, this symposium). Whatever determines the start and end of conversion, it is clear that such signals exist irrespective, and in ignorance, of the boundaries of a gene as a unit of function. It is essential, therefore, when comparing sequences to know whether the unit of comparison is equivalent to the unit of turnover. Many small conversion domains of 20–30 b.p. can coexist within a gene (for example the variable and constant genes of the immunoglobulin family and genes of the histocompatibility complex; see review in Dover & Strachan (1985)) which would generate mosaic genes of high overall variability, according to the rates and biases of turnover in the separate domains (Baltimore 1981; Bregegere 1983). Hence, no meaningful statements can be made about turnover rates and biases and divergence levels if a gene, compounded of




PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

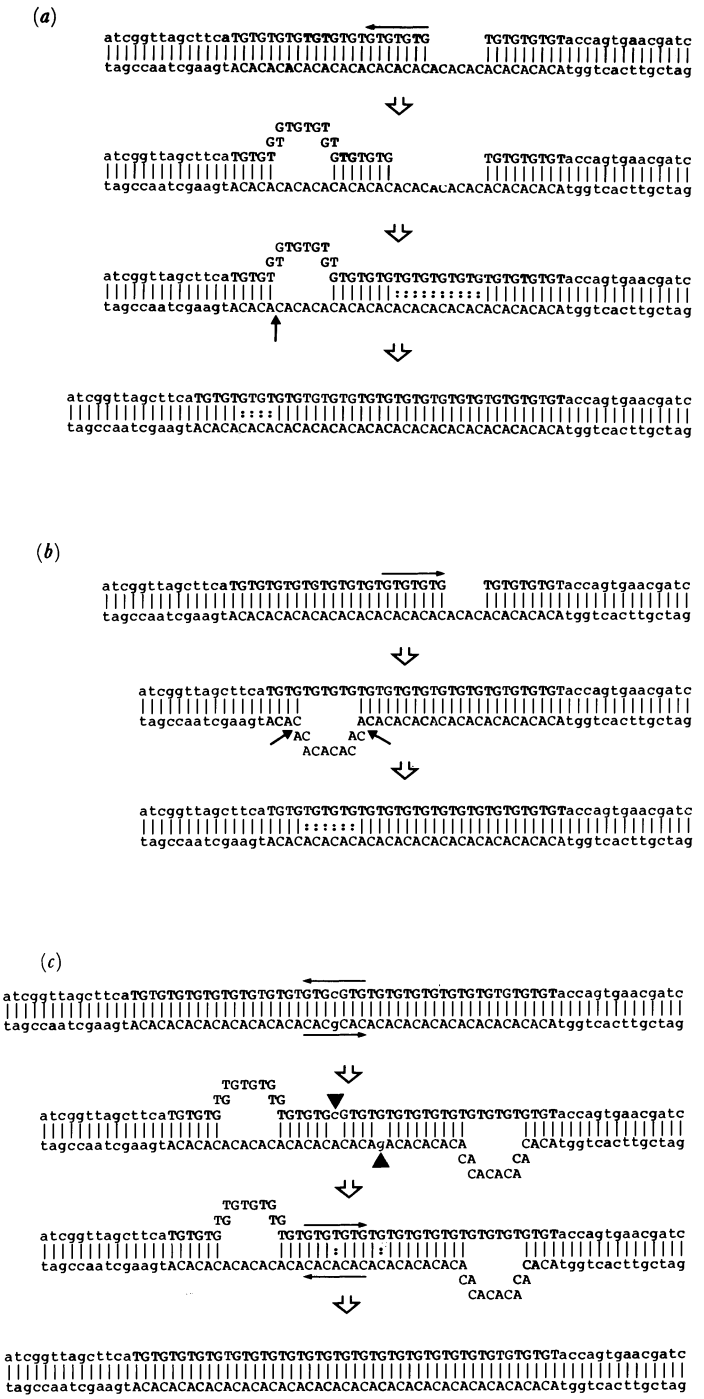


FIGURE 4. Model for the expansion (*a*), contraction (*b*), and self correction (*c*) of simple sequences by slippage.

The model in (*a*) and (*b*) starts during replication, after removal of the RNA primer from Okazaki fragment. We assume that a small gap (about ten nucleotides) exists in the double helix with two free ends which are not locked into their position by any enzyme and would thus be free for slippage.

The horizontal thin arrows above the sequences indicate the direction of slippage. Repair synthesis and ligation are indicated by colons and a dot respectively. Other arrows indicate the points of strand scission for repair synthesis.

The model in (*c*) assumes no free ends, but that slippage occurs of a long simple sequence stretch within itself and subsequent repair of mispaired base pairs as indicated by arrowheads.




PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

several conversion domains, is taken as the unit of comparison. Failure to appreciate this point has led to erroneous calculations on the contribution of DNA turnover to gene divergence and conservation (Gojobori & Nei 1984).

### (*b*) *Turnover feedback and conservation of sequence*

Turnover feedback and the conservation of sequences relates specifically to unequal exchanges and slippage replication. It is clear from the immense variety of repeated families, both in sequence and unit length, that any sequence motif can undergo *de novo* amplification. For tandem arrays in particular the mechanisms of amplification are likely to be unequal exchange and slippage replication. After amplification (which itself can be gradual and cohesive through a population rather than saltatory), the arrays are continually subject to the same mechanisms responsible for their amplification. Both unequal exchanges and slippage replication will tend to operate on unit lengths of sequence that are equivalent to the amplified repeat lengths. It is possible also, that new variant repeats do not participate as frequently in either of the two mechanisms, which probably rely on an initial degree of sequence homology. Hence, both mechanisms tend to feedback on themselves and continually promote and conserve their own sequence and length.

In the case of slippage replication it has been observed that all types of simple sequences can be generated by these processes (Tautz & Renz 1984). These repeats are so short (repeat unit length: one to a few nucleotides) that the same variants can repeatedly arise *de novo* in the different regions of the genome. Such ongoing mechanisms of *de novo* amplification and deletion of the same sequences, would question previous speculations that they have evolved as a consequence of function (Hamada *et al.* 1982). Figure 4 shows a model of slippage replication which is the most likely mechanism by which these sequences arise, and the manner in which the sequences feed back on themselves during amplification or deletion.

Finally, in figure 4*c*, we illustrate how self-correction of existing simple sequences, mediated by non-replicative slippage, can take place. This would also allow for the maintenance of sequences to a greater extent than that occurring in equivalent non-functional DNA elsewhere in the genome.

## Concluding remarks

There are tens of well characterized multigene families with specific effects on the metabolism, structure and behaviour of organisms. The developmental and evolutionary biology of these families are affected clearly by the activities of several mechanisms of DNA turnover. The relative contributions of natural selection, genetic drift and molecular drive to the shaping of multigene families have to be studied case by case. The detailed molecular investigations on the rDNA family in many organisms specifically underline the nature of the synergism between drive and selection and their involvement in the molecular coevolution of interacting molecules.

Molecular coevolution is a means of maintaining biological function despite the continual rounds of homogenization and fixation of new variants in functionally important regions of multigene families. Conversely, DNA regions of relative conservation exist, not necessarily through the action of stabilizing selection on function, but as the inevitable consequences of the biases, rates and feedback features of the DNA turnover mechanisms themselves.

The molecular biology and population dynamics of widespread misdemeanours of DNA




PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

behaviour are telling us that observed levels of DNA divergence and conservation are not explicable solely in terms of the selectionist–neutralist debate, reflecting function and no function. Given the triad of forces responsible for evolution, it is premature to categorize all functions and their contribution to survival as Darwinian adaptations (for specific examples see Dover (1986*a*) and Dover & Strachan (1986)).

We greatly appreciate the thoughtful comments made by Antonio Garcia-Bellido. We also thank June Hunt for typing the paper.

## References


Arnheim, N. 1983 Concerted evolution of multigene families. In *Evolution of genes and proteins* (ed. M. Nei & R. K. Koehn), pp. 38–61. Sunderland: Sinauer Ass. Inc.

Baltimore, D. 1981 Gene conversion: Some implications for immunoglobulin genes. *Cell* **24**, 592–594.

Barnes, S., Webb, D. A. & Dover, G. A. 1978 The distribution of satellite and main-band DNA in the *melanogaster* species subgroup of *Drosophila*. *Chromosoma* **67**, 341–363.

Barrie, P. A., Jeffreys, A. J. & Scott, A. F. 1981 Evolution of the β-globin gene cluster in man and primates. *J. molec. Biol.* **149**, 319–336.

Bodmer, M. & Ashburner, M. 1984 Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of *Drosophila*. *Nature, Lond.* **309**, 425–430.

Bregegere, F. 1983 A directional process of gene conversion is expected to yield dynamic polymorphism associated with stability of alternative alleles in class I histocompatibility antigens gene family. *Biochimie* **65**, 229–237.

Brown, S. D. M., & Dover, G. A. 1981 Organisation and evolutionary progress of a dispersed family in widely separated rodent genomes. *J. molec. Biol.* **150**, 441–446.

Brutlag, D. L. 1981 Molecular arrangements and evolution of heterochromatic DNA. *A. Rev. Genet.* **14**, 121–144.

Clark, C. G., Tague, B. W., Ware, V. C. & Gerbi, S. A. 1984 *Xenopus laevis* 28 rRNA: a secondary structure model and its evolutionary implications. *Nucl. Acids Res.* **12**, 6197–6220.

Coen, E. S. 1983 The dynamics of multigene family evolution in *Drosophila*. Ph.D. Thesis, University of Cambridge.

Coen, E. S. & Dover, G. A. 1983 Unequal exchanges and the coevolution of X and Y rDNA arrays in *Drosophila melanogaster*. *Cell* **33**, 849–855.

Coen, E. S., Thoday, J. M. & Dover, G. A. 1982*a* Rate of turnover of structural variants in the rDNA gene family of *Drosphila melanogaster*. *Nature, Lond.* **295**, 564–568.

Coen, E. S., Strachan, T. & Dover, G. A. 1982*b* Dynamics of concerted evolution of ribosomal DNA and histone gene families in the *melanogaster* species subgroup of *Drosophila*. *J. molec. Biol.* **158**, 17–35.

Coen, E. S., Strachan, T., Brown, S. & Dover, G. A. 1983 On the limited independence of chromosome evolution. In *Kew chromosome conference II* (ed. K. Jones), pp. 295–303. London: Allen & Unwin.

Dover, G. A. 1982 Molecular drive: a cohesive mode of species evolution. *Nature, Lond.* **299**, 111–117.

Dover, G. A. 1986*a* The spread and success of non-Darwinian novelties. In *Evolutionary processes and theory* (ed. S. Karlin & E. Nevo). New York and London: Academic Press. (In the press.)

Dover, G. A. 1986*b* *Evolution by means of molecular drive*. Cambridge University Press. (In preparation.)

Dover, G. A., Brown, S., Coen, E., Dallas, J., Strachan, T. & Trick, M. 1982 The dynamics of genome evolution and species differentiation. In *Genome evolution* (ed. G. A. Dover & R. B. Flavell), pp. 343–372. London: Academic Press.

Dover, G. A. & Flavell, R. B. 1984 Molecular coevolution: DNA divergence and the maintenance of function. *Cell* **38**, 622–623.

Dover, G. A. & Strachan, T. 1986 Molecular drive in the immune superfamily of genes: initiation and spread of novelty. In *Evolution and vertebrate immunity: the antigen receptor and MHC gene families* (ed. G. Kelsoe). University of Texas Press. (In the press.)

Fedoroff, N. V. 1979 On spacers. *Cell* **16**, 697–710.

Fisher, R. A. 1930 *The genetical theory of natural selection*. Oxford: Clarendon Press.

Fogel, S., Mortimer, R. K. & Lusnak, K. 1981 Mechanisms of meiotic gene conversion, or 'wanderings of a foreign strand'. In *Molecular biology of the yeast saccharomyces* pp. 289–339. New York: Cold Spring Harbor.

Franz, G., Tautz, D. & Dover, G. A. 1985 Occurrence of major S1 sensitive sites in the non-conserved spacer region of rDNA in *Drosophila* species. *J. molec. Biol.* **183**, 519–527.

Gerbi, S. A., Gourse, R. L. & Clark, C. K. 1982 Conserved regions within ribosomal DNA: locations and some possible functions. In *The cell nucleus rDNA* part A (ed. H. Busch & L. Rothblum), vol. 10, pp. 351–386. New York: Academic Press.

Glover, D. M. 1981 The rDNA of *Drosophila melanogaster*. *Cell* **26**, 297–298.

Gojobori, T. & Nei, M. 1984 Concerted evolution of the immunoglobulin $V_H$ gene family. *Molec. Biol. Evol.* **1**, 195–212.






PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Grummt, I., Roth, E. & Paule, M. R. 1982 Ribosomal RNA transcription *in vitro* is species-specific. *Nature, Lond.* **296**, 173–176.
Hamada, H., Petrino, M. G. & Kakunaga, T 1982 A novel repeated element with Z-DNA forming potential is widely found in evolutionary diverse eukaryotic genomes. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6465–6469.
Huxley, J. 1942 *Evolution: the modern synthesis.* London: Allen & Unwin.
Kimura, M. & Ohta, T. 1979 Population genetics of multigene family with special reference to decrease of genetic correlation with distance between gene members on a chromosome. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4001–4005.
Kohorn, B. D. & Rae, P. M. M. 1982 Accurate transcription of truncated rDNA templates in a *Drosophila* cell-free system. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1501–1505.
Kreitman, M. 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature, Lond.* **304**, 412–417.
Lamb, B. C. & Helmi, S. 1982 The extent to which gene conversion can change allele frequencies in populations. *Genet. Res.* **39**, 199–217.
Long, E. H. & Dawid, I. B. 1980 Repeated genes in eukaryotes. *A. Rev. Biochem.* **49**, 727–764.
Miesfeld, R. & Arnheim, N. 1984 Species-specific rDNA transcription is due to promoter-specific binding factors. *Molec. Cell. Biol.* **4**, 222–227.
Mishima, Y., Financsek, I., Kominami, R. & Muramatsu, M. 1982 Factors required for accurate transcription of mammalian rDNA: identification of species-dependent initiation factor. *Nucl. Acids Res.* **10**, 6659–6669.
Nagylaki, T. & Petes, T. D. 1982 Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. *Genetics* **100**, 315–337.
Noller, H. F. 1984 Structure of ribosomal RNA. *A. Rev. Biochem.* **53**, 119–162.
Ohta, T. 1980 Evolution and variation of multigene families. *Lecture Notes Biomath.* **37**. New York: Springer-Verlag.
Ohta, T. 1983 On the evolution of multigene families. *Theor. Pop. Biol.* **23**, 216–240.
Ohta, T. & Dover G. A. 1983 Population genetics of multigene families that are dispersed into two or more chromosomes. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4079–4083.
Ohta, T. & Dover, G. A. 1984 The cohesive population genetics of molecular drive. *Genetics* **108**, 501–521.
Peacock, W. J., Lohe, A. R., Gerlach, W. L., Dunsmuir, P., Dennis, E. S. & Appels, R. 1977 Fine structure and evolution of DNA in heterochromatin. *Cold Spring Harb. Symp. quant. Biol.* **43**, 1121–1135.
Reeder, R. H. 1984 Enhancers and ribosomal gene spacers. *Cell* **38**, 349–351.
Skinner, J. A., Öhrlein, A. & Grummt, I. 1984 *In vitro* mutagenesis and transcriptional analysis of a mouse ribosomal promoter element. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2137–2141.
Steinmetz, M., Malissen, M., Hood, L., Orn, A., Maki, R. A., Dastournikoo, G. R., Stephan, D., Gibb, E. & Romaniuk, R. 1984 Tracts of high and low sequence divergence in the mouse major histocompatibility complex. *EMBO J.* **3**, 2995–3003.
Strachan, T., Coen, E. S., Webb, D. A. & Dover, G. A. 1982 Modes and rates of change of complex DNA families of *Drosophila*. *J. molec. Biol.* **158**, 37–54.
Strachan, T., Webb, D. A. & Dover, G. A. 1985 Transition stages of molecular drive in multiple-copy DNA families in *Drosophila*. *EMBO J.* **4**, 1701–1708.
Tartof, K. 1974 Unequal mitotic sister chromatid exchange and disproportionate replication as mechanisms regulating ribosomal RNA gene redundancies. *Cold Spring Harb. Symp. quant. Biol.* **38**, 491–500.
Tautz, D. & Renz, M. 1984 Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucl. Acids Res.* **12**, 4127–4138.
Trick, M. & Dover, G. A. 1984 Unexpectedly slow homogenisation within a repetitive DNA family shared between two subspecies of tsetse fly. *J. molec. Evol.* **20**, 322–329.
Whitehouse, H. L. K. 1983 *Genetic recombination – understanding the mechanisms.* New York: Wiley.




PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES