J Mol Evol (1992) 34:259–271

Journal of Molecular Evolution

© Springer-Verlag New York Inc. 1992

*Focus*

# Revisiting Junk DNA

Emile Zuckerkandl

Linus Pauling Institute of Science and Medicine, 440 Page Mill Road, Palo Alto, CA 94306, USA

**Summary.** The distribution of functions within genomes of higher organisms relative to processes that lead to the spread of mutations in populations is examined in its general outlines. A number of points are enumerated that collectively put in question the concept of junk DNA: the plausible compatibility of DNA function with rapid substitution rates; the likelihood of superimposed functions along much of eukaryotic DNA; the potential for a merely conditional functionality in sequence repeats; the apparent adoption of macromolecular waste as a strategy for maintaining a function without selective grooming of individual sequence repeats that carry out the function; the likely requirement that any DNA sequence must be "polite" vis-à-vis (compatible with) functional sequences in its genomic environment; the existence in germ-cell lineages of selective constraints that are not apparent in populations of individuals; and the fact that DNA tectonics—the appearance and disappearance of genomic DNA—are not incompatible with function. It is pointed out that the inverse correlation between functional constraints and rates of substitution cannot be claimed to be a pillar of the neutral theory, because it is also predicted from a selectionist viewpoint. The dispensability of functional structures is brought into relation with the concept of reproductive sufficiency—the survivability of genotypes in the absence of fitter alleles.

**Key words:** Junk DNA — Equifunctional substitutions — Functional involvement — Functional constraints — DNA tectonics — Dispensability of sequences — Reproductive sufficiency

Whereas prokaryotic genomes are known to be composed of functional sequences in quasi-uninterrupted apposition, eukaryotic genomes are considered to be constituted by functional DNA islands in a sea of nonfunctional DNA.[1] The present paper focuses on relationships, observations, and inferences germane to a problem that is far from having been laid to rest—the proportion of functional sequences in eukaryote genomes. Without denying the obvious, the existence of an islands/sea contrast between coding and noncoding sequences, it will be shown that available evidence does not refute, and increasingly supports, the view that the sea itself is to a significant extent permeated by function.

## Functional Space, Selection Space, and Other "Spaces" in Genomes

Functional structures and processes are those that play a role in the operation and persistence of an established biological system and that contribute to defining what the system is.

While selection expresses function, selection and function cannot be considered as congruent categories. Indeed, functional structures, e.g., a mutated nucleotide as compared with the nucleotide that it replaces, can spread in populations by random drift as equifunctional structures (Kimura 1983), namely as functional structures endowed with sufficiently similar positive selection coefficients for selection to fail to come into play under prevailing

[1] To quote George Russev

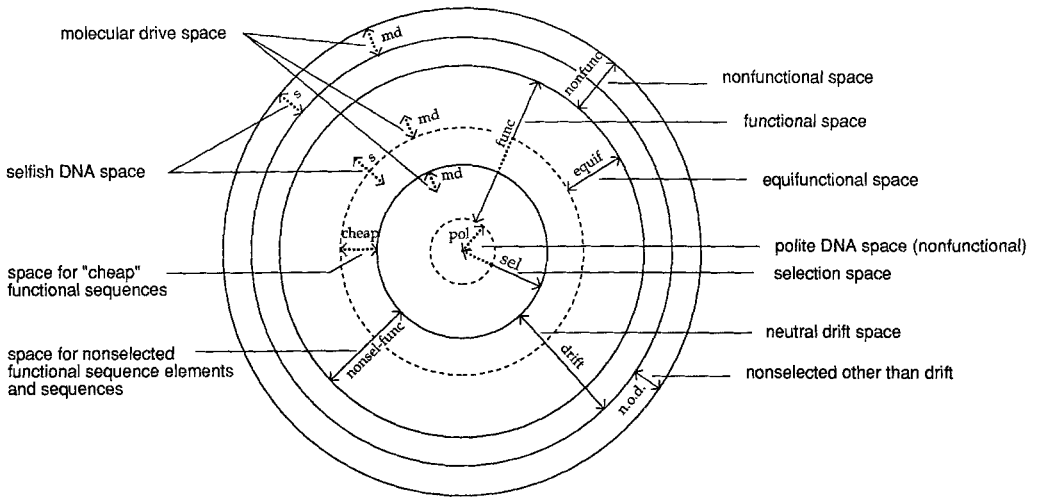


**Fig. 1.** Relationships in eukaryote genomes, described in terms of overlapping "phase spaces," between the functional status of sequences and modes of fixation of sequence change. Space boundaries and arrows spanning spaces are dotted where the phase spaces are only hypothetical or their recorded relations appear to be only partly, or conditionally, operative. The polite DNA space is a hypothetical space that is presumed to partake in both the selection space (for negative selection only) and the nonfunctional space. Although the segments of the figure's radius are not in a quantitatively meaningful relation to one another, they do suggest that the functional space and the drift space are particularly large.

conditions. Perhaps, also, functional structures can, in some situations, be perpetuated by being repeatedly regenerated from a functional and selected master sequence, without selection acting on the copies (Zuckerkandl et al. 1989).

There is in principle a population-genetical criterion for functional sites. To a first approximation, functional sites are those at which not *all* occurring mutations spread by neutral drift, i.e., sites at which mutations do *not* spread by drift at mutation rate. This criterion however does not take into account that there might be additional functional sites at which mutations do propagate at mutation rate (Zuckerkandl et al. 1989). Moreover, the criterion can be difficult to apply in the case of repetitive sequences, because of a possible interference by processes of molecular drive (see below).

There is an ongoing growth in our perception (e.g., Pardue et al. 1987; Trifonov 1989; Vogt 1990; Rogerson 1991; see also Zuckerkandl 1987) of what one may term the genome's functional space (Fig. 1). This is the proportion of an organism's DNA that fills a function of some kind. Though the functional space is no doubt larger than the selection space (because nonselected sequences may be functional), we do not know how much larger. It is unlikely for the functional space to grow without the selection space growing also.

Some examples of the slow but apparently steady invasion of the genome by function involve a few of the clearest cases there are of so-called junk DNA: purine–purine, pyrimidine–pyrimidine, and purine–pyrimidine dimer repeats. $(TG)_n$ repeats, with $n$ about 10–60 in humans, are spread in about 50,000 copies over the genome (Litt and Luty 1989). *Drosophila* has proteins that specifically bind to such sequences, not only in isolated DNA, but in nuclei (Gilmour et al. 1989). These long dinucleotide repeats correlate with overall transcriptional activity in chromosomal regions, with the ability of the regions to undergo meiotic recombination, and with increased dosage compensation (Lowenhaupt et al. 1989). TG repeats have been found close to the promoter and enhancer regions in a significant number of genes (see Boulikas 1991). They can function as a highly specific protein recognition feature (Vashakidze et al. 1988), and this property is probably linked to their ability to induce kinks in the DNA (Bolshoy et al. 1991). Though no direct proof of functions is available as yet, on this front nonfunctionalists seem to have poor prospects. $(CT)_n$ is another example. Alternating CT repeats are present in a number of genes (see Boulikas 1991). They bind certain proteins specifically (e.g., Yee et al. 1991). In *Drosophila,* there is a cooperative interaction between the $(CT)_n$ element upstream of the heat-shock gene hsp26 promoter and the proteins that bind to the TATA box (Elgin 1990). Most decisively perhaps as far as the existence of a function is concerned, the $(CT)_n$ repeat acts as a positive transcriptional regulator and its deletion leads to a four- to fivefold reduction in hsp26-inducible activity in vivo (Glaeser et al. 1990).

Figure 1 represents the relationship between some genomic "phase spaces." The figure indicates a conceptual, not a quantitative relationship between phase spaces. The radius of the outermost circle in the figure represents the total DNA of higher or-

ganisms. Different processes and characteristics are assigned to various fractions of the total DNA. It is shown that these processes and characteristics at times exclude each other, at times overlap, and in other cases are nested in one another. The sizes of the selection, drift, and equifunctionality DNA spaces can be expected to vary in relative amounts as a function of population size. Thus, quantitative concretizations of the diagram would require the specification of effective population sizes and also of other parameters.

Kimura (1983, p. 176) points out that his selectively neutral mutations are the selectively equivalent ones, and thus leaves the door wide open to their having a function.[2] Kimura uses phrases such as "neutral (i.e., equivalent)" (p. 196) and says: (The neutral theory) "holds that *most* evolutionary *changes* are caused by random genetic drift of mutant alleles that are selectively equivalent" (p. 54). To equate selectively equivalent with equifunctional is no doubt a good approximation. Admittedly, it can be the case that a selectively equivalent substituent, say a substituent amino acid compared to the substituted amino acid, has quite different properties, suggesting that it fills a different function. This occurs when the substituent amino acid contributes to a general molecular function (Zuckerkandl 1976a), such as the polarity of the protein. For instance, the replaced amino acid residue could be charged and the replacing amino acid residue, apolar. Despite the large difference in structure and properties between the amino acid residues successively found at the molecular site in regard to the particular functional characteristic considered, namely overall polarity, the site continues to fill the same global function. When the substituent is equifunctional, it means that the difference in polarity and charge introduced by the amino acid replacement, in terms of the whole protein or of a region of it, is small enough under the circumstances so as to be compatible with neutral drift.

The equifunctional is an interface between the functional and the neutral. An equifunctional substituent is functional but must spread by neutral drift. The concept is most pertinent in its form of near-equifunctionality, paralleling near-neutrality. The size in the genome of the equifunctionality space (Fig. 1) is an unsolved question. The number of possible equifunctional substituents at any one site must often be constrained by substituents at other sites. Equifunctionality no doubt often involves a complex combinatorial game that is arduous to explore experimentally. One equifunctional substitution may render nonequifunctional another that would have been equifunctional had it occurred first. Particularly in coding sequences and perhaps more generally, the equifunctionality space may well be smaller than had been anticipated (e.g., Barnard et al. 1972; Zuckerkandl 1975, 1987), especially when effective population sizes are substantial and smaller differentials in selection coefficients can become effective.

The neutral drift space, which contains functional structures besides nonfunctional ones, is higher up on the totem pole of potential (future evolutionary) functional impact than the nonfunctional space, if it be true that, except in very early evolution, functional structures mostly originate from other functional structures (Zuckerkandl 1975).

The nonselected functional space reflects the tenet that a significant portion of the total DNA is likely to fill functions that require the interaction of specific proteins with degenerate repeated short sequence motifs. These latter are thought to arise mostly not through selection of individual nucleotides, but via sectorial controls of DNA base composition (Zuckerkandl and Villet 1988), whatever their mechanism. The existence of such controls has been established by Bernardi and associates (Bernardi et al. 1988). One part of this nonselected functional space is represented by the hypothetical cheap

---

[2] Kimura (1983, p. 50) wrote: "The neutral theory by no means claims that the genes involved are functionless as mistakenly suggested by Zuckerkandl (1978)." There obviously has been some misunderstanding, as I made no such statement. The misunderstanding apparently arose because I pointed out, not using exactly this wording, that there could be no nonfunctional gene (except duplicates in the process of becoming pseudogenes, which may yet fill functions other than the protein-coding function), and that all genes, in order to become what they are, had to be selected for. There is hardly an occasion for disagreement here. Although substituents at a polypeptide site could spread by neutral drift, I remarked, the site itself presumably will not be neutral, neither in the sense of functional neutrality nor in that of neutral drift. (There could be exceptions to this.) As mentioned, equifunctionality rather than functional neutrality are emphasized by the neutral theory. Yet the frequency of equifunctional substitutions is said to increase as a molecule or a part of a molecule becomes functionally less important (Kimura 1983, p. 157). At the limit the function disappears, corresponding to total indifference to substitution, all substitutions then being subject to neutral drift. The point deserved to be put forth (it is made more clearly here than in 1978) that whole proteins and also, according to numerous indications, most protein sites cannot be expected to develop a functional indifference to amino acid replacements approximating the zone of extreme functional insignificance where equifunctionality becomes congruent with nonfunctionality. It is true that there may be sites in loops or at ends of polypeptide chains that are not functional, yet still will not accept all substituents, because some are not compatible with the molecule's function as carried out by other sites. Such substituents are impolite (see definition of politeness, p. 262). Most sites of a protein tend to be conserved either as a character of function or a character of politeness. Assume that an amino acid residue in an $\alpha$-helix fills no function [neither a specific nor a general one (Zuckerkandl, 1976a)]—an unlikely eventuality. Even then, the elimination of the amino acid site would be very impolite—it would compromise the function of other sites in the helix—and would no doubt be selected against

genes, which are multiple sequence copies generated from a selected progenitor sequence and presumed originally to be capable of filling a function, yet abandoned to a random decay that eventually abolishes the daughter sequences' function(s) and turns these sequences into pseudogenes (Zuckerkandl et al. 1989).

Because of constraints of representation, the nonfunctional space (which probably is in part a conditionally functional space, see point 3 below) is represented twice in Fig. 1: once in order to show that the nonfunctional space of course participates in the drift space, and a second time to emphasize that the same nonfunctional space may well also be part of the selection space. Namely, it may, under certain conditions, be subject to negative (not to positive!) selection, when it fails to qualify as polite DNA.

The "polite" DNA space (Zuckerkandl 1986) is a phase space of sequences, sequence elements, or base compositional characters that are not intrinsically functional but conform to structural imperatives in the interest of functions carried out elsewhere in the DNA sector. Though not part of the functional space proper, polite DNA encroaches upon the neutral drift space,[3] which is thus decreased. In Fig. 1 polite DNA is therefore represented outside of both the functional and the neutral drift spaces. At the same time it is presumed to belong half-way to the selection space. For this reason it is placed at the center rather than with other nonfunctional sequences at the periphery of the figure. Politeness as such, separately from function, cannot be positively selected, yet it is expected that marked impoliteness is negatively selected. Positively selected structural features will in general also be polite.

Selection as represented in the figure is selection at the level of the individual. Selfish DNA, another important category (Doolittle and Sapienza 1980; Orgel and Crick 1980), is nonselected at the level of the individual, but it is selected at the level of the genome. It implies, therefore, a dimension that is not represented in the figure, as Ford Doolittle remarks (personal communication). In its capacity of self-perpetuating, self-spreading parasitic DNA, selfish DNA has to be integrated into the figure as part of the nonselected-other-than-drift space. In this sense it is given in the figure as sharing its space with molecular drive (see below). A sequence that selects itself selfishly in the genome could spread in a population at a rate much higher than neutral drift would allow "because its ability to transpose itself would ensure that all progeny chromosomes from a hybrid mating would be infected with the element" (Ford Doolittle, personal communication). On the other hand, selfish DNA is really not to be excluded from any other category, exept from selection at the level of the individual. For instance, it may coincidentally be functional. Also, cheap genes, to the extent to which they are generated by reverse transcription followed by retroposition (cf. Rogers 1985), certainly have a selfish trait.

The phase space for the processes subsumed under the term molecular drive (Dover 1986) is represented in Fig. 1 in several bits and pieces. The concept deals with mechanisms that keep certain repeat sequences identical in the genome. Inasmuch as the effects of molecular drive are brought about by processes that express neither selection nor drift, molecular drive has a genome space of its own, represented in the diagram, jointly with selfish DNA, as the outermost shell. Like selfish DNA, DNA spread by molecular drive may involve sequences that are part of the functional as well as the nonfunctional spaces. Negative selection could intervene to block biased gene conversion. Positive selection could favor individuals that have a larger or smaller number of identical repeats of a kind. Differences in these repeat numbers are liable to be selectively significant when the numbers are small. When copy numbers have grown large, then, above a certain threshold of copy number, selection could intervene in the "forward creep—back leap" mode (Zuckerkandl 1986). In this mode, slightly deleterious mutations are assumed to accumulate locally and progressively by neutral drift, until the summed substitutions sufficiently lower relative fitness for the DNA segment to be eliminated, or replaced, by selection. In the present case slightly deleterious further increases in copy number of a given sequence would eventually lead to the elimination of a group of the copies in a single event.

Processes of molecular drive, like selfish DNA, could also occupy some of the equifunctional phase space, as suggested in the figure. Indeed, a given type of repeat sequence may be functional, yet function and selection may be indifferent to the identity of the precise variant of the sequence type that acts as the master sequence and insures the identity of the repeats.

The imagined proportions of various genomic spaces of the kind referred to—with none of the proportions effectively determined so far—are shifting in the minds of at least some evolutionists. In particular, as mentioned, the functional space and the selection space seem to be growing. What becomes in this process of one further space, one not represented in the figure, even though it has been in the limelight for many years, namely the genomic space of junk DNA?

[3] The phrase neutral space was used by Takahata (1987)

## Circumstances that Discredit the Junk DNA Concept

Most of the noncoding DNA sequences are supposed to represent the genome's "junk space" (Ohno 1972). Investigators who think that, for example, satellite DNA is junk emphasize the rapid changes in sequence and sequence organization in satellite DNA, the variations from individual to individual, the enormous quantitative variations from species to species in the genomic multiplicity of any particular type of satellite sequence, and the fact that their presence or absence has no obvious phenotypic effect (except at times in the germ line) nor any discernable functional impact (Miklos 1985).

Statements about the absence of phenotypic effects are often unsatisfactory. Few if any measurements have been made to ascertain whether the regulation of some cellular mechanisms such as transcription of certain genes and replication has been modulated. Nor have long-term genetic effects extending over many generations been evaluated. Also, possible phenotypic effects during development are rarely considered. Finally, possible effects on fitness of the loss of intergenic sequences, e.g., of an important fraction of a type of satellite DNA, have failed to be subjected to an essential test: to extensive experiments putting mutant and wild-type populations in competition with one another. Wu (1990) likewise has voiced criticisms of this kind.

According to one's assumptions about functional strategies in genomes, the observations regarding satellite DNA can be interpreted differently. All the observations cited are compatible with function. The junk concept can be confronted with a series of inferred or presumptive features of molecular evolution that collectively seem capable of pushing the junk DNA paradigm out of the central position that in many quarters it is still considered to occupy.

1) In noncoding DNA,[4] high rates of sequence change are compatible with functionality, through a) large contributions of equifunctional substitutions and b) low functional involvement.

For a DNA sector, functional involvement[5] may be defined as the proportion of nucleotide sites that are involved in functions. Considerable work is required before functional involvement in noncoding regions of DNA can be estimated, outside of some short, specific regulatory elements (Zuckerkandl 1986). When functional involvement is low and the rate of equifunctional substitutions is high, a fast evolutionary rate of sequence change is in principle compatible with the presence of a function in a given segment of DNA. Thus, high rates of sequence change do not by themselves strongly suggest, let alone demonstrate, nonfunctionality. The convergence of rates of base substitutions in different sectors of the genome on a given maximum value has been taken as a presumption in favor of that value expressing the neutral drift rate (Roy Britten, personal communication, 1989). This presumption is not strong. Different functions, if their DNA substrata share in relatively low functional involvement and/or high incidence of near-equifunctionality, can in principle be characterized by similar fast substitution rates, not much different from true neutral drift rates. The latter may never have been actually measured (Zuckerkandl 1986), for lack of identifying functionally completely unconstrained sectors of DNA.

Functionally involved nucleotides need to be in certain spatial relationships with respect to one another. They will therefore often be separated by interspacing nucleotides whose function is to bring the active nucleotides into the proper mutual spatial relationship. In an assessment of functional involvement, interspacing sequences that function exclusively as such should be counted as equivalent to single functional nucleotides irrespective of their length, because their functional contribution, if any, is that of a unit component. Thus, the longer the interspacing sequences, the more they lower the value of functional involvement of the DNA sector in which they occur. On the other hand, one functional sequence element may be another's interspacing sequence and vice versa, with overlaps and nesting presumed to be common occurrences. In such cases, functional involvement is increased in the DNA sector. For instance, considering DNA binding sites for glucocorticoid receptor and progesterone receptor in the promoter of mouse mammary tumor virus, a promoter-proximal and promoter-distal group of binding sites lie on the same side of a nucleosome, separated by only 2 nm between two turns of the DNA helix around the nucleosome (Piña et al. 1990). This proximity could favor interactions between steroid receptor molecules that would not occur along linear DNA. Thus, with respect to these receptor binding sites, the intermediate 96 bp would appear to function as an interspacing sequence; yet, within this interspacing sequence, other functions are likely to be present, such as those of nucleotides that determine the position of the nucleosome to be where it is found (Piña et al. 1990).

A low functional involvement of a segment of

[4] The phrase noncoding DNA is used here in the sense of non-protein-coding and non-functional RNA-coding, even though many different codes are used in genomes (Trifonov 1989)

[5] The phrase functional involvement is used here rather than functional density. Functional density has been defined as the proportion of molecular sites involved in specific functions (Zuckerkandl 1976a, 1986). Functional involvement can in the future be estimated for sectors of DNA without regard to a distinction between specific and general DNA functions

DNA does not imply a lowly function for that segment; it only implies that the function, however important it may be, requires or accommodates a large contingent of interspacing sequences between the active nucleotides.

2) Non-protein-coding sequences probably contain different messages (codes) of which only some are known at present. For example, many introns exclude hairpin structures in favor of mirror-symmetric nucleotide sequences (Trifonov 1989). We seem to be moving toward the discovery not only of new functions, but of superimposed functions. The existence of superimposed protein-coding sequences has been known for many years. There are examples, and Trifonov (1989) has produced the best so far, of protein-coding sequences upon which several other codes are superimposed. An additional example worth mentioning may result from structural requirements that are in part mutually exclusive: in eukaryotes, coding sequences need to satisfy simultaneously the imperatives of protein structure and of chromatin structure. A way to reconcile these competing imperatives is to keep uninterrupted runs of codons short enough so that they do not significantly destabilize chromatin structure. The insertion into, or conservation within, structural genes of intervening sequences appears to resolve the conflict (Zuckerkandl 1981; Beckmann and Trifonov 1991)—a suggestion to the effect that intervening sequences are *not* junk. If they are functional with respect to chromatin structure, overall substitution rate in intervening sequences probably does not measure true mutation rate.

We may expect to find superimposed sequence constraints of different kinds to be widespread in the intergenic parts of genomes also. There can be little doubt that, at least during some phase of development or of the cell cycle, every piece of polynucleotide sequence is involved in multiple interactions (DNA/protein, DNA/RNA, RNA/protein, and corresponding ternary interactions) (Trifonov 1989). An unknown, but probably not small, proportion of these interactions is likely to be accompanied by some constraints on nucleotide substitutions, as required, for instance, for satisfactory packaging of DNA. Such constraints may usually be very weak for individual nucleotide positions but may be significant at the level of the DNA sector.

An example of superimposed functions in noncoding DNA has already been given: of a spacer function and a nucleosome positioning function. One can conceive of further kinds of superimposed sequence constraints in noncoding DNA. For example, the same sequence that is recognized by a certain DNA-binding protein when the sequence is in one structural state must also be such as to favor a different structural state in which the sequence does not preferentially bind this protein but another, or both. This is likely to apply to a DNA domain (a chromatin loop) when its structural state changes, as indicated by a switch from relative DNase I insensitivity to a generalized moderate ("intermediate") DNase I sensitivity (Stalder et al. 1980; Felsenfeld et al. 1982; Stumph et al. 1982), extending over, say, 100 kb. One would anticipate that in its two states, the state relatively insensitive to endonucleases and the moderately sensitive state, the chromatin loop, in which noncoding sequences predominate, does not on average bind the same mix of specific mass-binding (multiple-site-binding) protein factors, such as histone H1 and some nonhistone proteins. This matter has apparently not yet been explored.

Examples involving more limited regions of DNA are better known. It appears that in vertebrates generally the binding of either an inhibitory nucleosome or of the TATA box-binding transcription factor TFIID to the same DNA sector determines whether the structural state of chromatin is or is not compatible with the binding of RNA polymerase II and with transcription (Workman and Roeder 1987; Knezetic et al. 1988; Elgin 1990). A similar situation applies to the action of polymerase III (see Workman and Roeder 1987). In the promoter of the mouse mammary tumor virus the binding of nucleosomes prevents the binding of nuclear factor 1 (NF1), which is part of the transcription initiation complex, whereas glucocorticoid receptor binds to its germane DNA element when this element is engaged in a nucleosome core (Archer et al. 1991). One may expect that many noncoding sequences need to be such that they bind different proteins when in different structural states. Each of these states implies certain sequence requirements for binding the proteins. These requirements need not imply the adherence to a uniquely defined repeated sequence for binding the proteins. A certain degeneracy of these sequences is expected (Trifonov 1983), for instance in intergenic regions beyond the sequences adjacent to the coding sequences. Such degeneracy is in principle compatible with high specificity of binding of a protein or a protein complex over a sector of DNA (Zuckerkandl and Villet 1988).

3) For the generation and maintenance of functional units, the genome may well use two different strategies[6] that are probably somewhat comparable

[6] The term strategy of course does not refer to a real purposiveness, but to purely mechanistic processes expressing natural selection, as understood by Ernst Mayr (1988) in his Essay Three. At this point of development of the evolutionary sciences one should not have to forego the convenience, at times, of finalistic language for fear of literal and retrograde interpretations of images, in an area that scientific thinking has long since clarified

to K-selection and r-selection in populations, namely the individual grooming of units through natural selection or alternatively the repeated production of a large number of potentially functional units that are permitted to drift. It is likely that the same repeated sequences can be either pivotal for a function or only conditionally functional and effectively superfluous, according to the number of these sequence repeats available in the genome and to their distribution. This conclusion is suggested by the analysis of the structure and function of enhancers and is probably of more general applicability (Zuckerkandl 1986). Enhancers are frequently made of repeats of different sequence motifs that can replace each other in various combinations (Serfling et al. 1985). For a given type of repeat, the effectiveness of the enhancer increases with the repeat number up to a limit. If $z$ is a particular repeat added after this limit has been reached, $z$ can be considered as functionless. However if, say, repeats $x$ and $y$ are lost, $z$ becomes functional. The effective functionality of a noncoding sequence is thus expected to depend frequently not only on its own characteristics, but on the amounts of similar or comparable sequences that are available in the genomic neighborhood. Though complete examples in support of this expectation may not yet be available, part of it has been verified. A case in point is the so-called GC boxes, which bind the transcriptional activator Sp1. Although one or two GC boxes are sufficient for low levels of transcriptional activation by Sp1, high levels of stimulation require multiple Sp1 binding sites close to the transcriptional start site (Courey and Tjian 1988).

The intensity of constraints imposed by natural selection on an individual sequence is likewise expected to be determined in part by the degree of redundancy of the sequence, for instance in satellite DNA. The constraints in an individual repeat may tend toward zero when the repeat frequency is high, and yet collectively such sequences may play a functional role. Are redundant repeats that are only conditionally functional to be considered as junk? No, as a consideration related to the present one suggests.

4) Indeed the genome probably has a tendency to generate macromolecular waste, in the form of a large excess of certain potentially functional sequences. This is a trend that can favor biological success with little penalty attached to overshooting. One is reminded of cellular waste in the production of gametes. In many male organisms, more sperm are produced than are required for the survival of the species. On the other hand, the amount of sperm produced in a species must be assumed to be regulated; not so precisely, however, that the presence or absence of a large absolute number of individual sperm would make a detectable difference. Overshooting in the production of potentially functional units does not prevent the individual units—gametes or satellite sequence modules—from being potentially functional; it does not turn the excess units into junk. Each gamete in excess is still a sophisticated biological unit that is ready to function, though individually deprived of a participation in shaping the future of the species.

The mechanisms of generating wasted functional sequences may well be in part those invoked for the spread of selfish DNA (Doolittle and Sapienza 1980; Orgel and Crick 1980). (Selfishness and altruism are not necessarily mutually exclusive; quite to the contrary, they may be mutually supportive.) If a DNA sequence capable of multiplying in the genome "parasitically" is also endowed with some functional properties, the function, or a potential for it, will spread in the genome, though much of this potential may go to waste.

5) DNA sequences are probably constrained to be polite vis-à-vis their sequence environment (Zuckerkandl 1986; Holmquist 1989). Saying that politeness is required on the part of DNA is a way of saying that DNA sequences must not, by their direct or indirect effects, be disruptive of functional structures located in the same DNA sector or in interacting sectors. Thus, even sequences that would not be responsible for participating in any function are expected to display structural and functional compatibility. This in turn is expected to impose some constraints on base sequence and/or composition.

The constraints of politeness can be mutual. This occurs when two adjacent DNA segments have to be mutually compatible, because each has its own requirement. For example, the nucleosome linker sequence needs to keep out the nucleosome, and the nucleosome core DNA sequence needs to keep out the linker sequence. Autonomously replicating sequences, ARS, are a case in point. It appears to be functionally important for ARS in yeast to be located in the more accessible linker region rather than in the core region, where they cannot function properly (Simpson 1990). Exclusion of nucleosomes from the linker region can be brought about by runs in the linker region of poly(dA)·poly(dT) (Nelson et al. 1987) and therefore by a contribution to function of typical junk DNA. The presence of such sequences in the core region would be impolite. Likewise, in the neighborhood of coding sequences, promoter and enhancer regions, characterized in part by hypersensitivity to DNase I, need to remain nucleosome-free in preparation for and during transcriptional activation (Elgin 1990; Archer et al. 1991; Straka and Hörz 1991). The contrary requirement, that of a particular sequence feature to be included

in a nucleosome core, presumably as a protection against DNA damage, appears to apply to the exon–intron junctions (Beckmann and Trifonov 1991). A DNA insertion that would compromise this relationship between nucleosome and exon–intron junction again would be impolite.

In the midst of what has been presumed to be nonsense DNA, more sense is becoming apparent progressively, and mandatory politeness of DNA is presumably a part of it.

6) Functions do not cease being functions when they are dispensable. *DNA tectonics,* namely the bodily disappearance of old DNA and the appearance of new DNA, are compatible with functionality. DNA tectonics have been examined in the past, notably by Hinegardner (1976). The disappearance of functions from the genome becomes a threat to the persistence of the species only below the level of *reproductive sufficiency* of the average phenotype (Zuckerkandl 1976b, 1991). Reproductive sufficiency characterizes the mutational state of an allele when a population of individuals homozygous for this allele can maintain itself in the absence of a fitter allele. The disappearance from the genome of a sequence or of a type of repetitive sequence may indeed not inferfere with the reproductive sufficiency of the genome; yet, when present, the sequence or sequences may have a function and increase the fitness of the organism. Even a high turnover rate of sequences, in terms of their bodily disappearance and their generation through the various processes of DNA tectonics, is likely to be compatible with function of the sequences concerned and does not justify their being considered as junk.

7) There are likely to be many selective events not expressed in the postembryonic phenotype, namely selective events at hierarchical levels lower than that of the individual (Holmquist 1989). These events are expected to occur in part among germ-cell lineages. At the levels of gametes and embryos, selective constraints may come into play that are not manifest later, and these constraints at times might involve noncoding sequences in addition to coding sequences. Certain noncoding sequences that are considered to be junk in the adult could have their functional effects at early developmental stages. Perhaps heterochromatin in the grasshoppers of the genus *Atractomorpha* furnishes an example. Here quantitative variation over different species has significant germ-line effects on recombinational probabilities (John and Miklos 1979; Miklos 1985). No phenotypic effects of variations in heterochromatin have been detected in adult organisms by the same authors. So far we are dealing in this case only with an effect of heterochromatin and do not know as yet whether the effect is functional.

As a reassessment of junk DNA and of DNA function pushes much of the junk out of the genome, junk DNA is only in part being transformed into functional DNA. This should be clear from point 1 on functional involvement, point 3 on conditional functionality, point 4 on molecular waste, and point 5 on polite DNA. The view discussed here or in earlier papers (e.g., Zuckerkandl 1986), therefore, is not that all DNA is functional, contrary to what Li and Graur (1991) reported in their recent book. It only gives functions a more pervasive place in eukaryote genomes than is commonly believed.

I shall further comment here on point 1 (rates of sequence change and functionality) and point 6 (DNA tectonics and reproductive sufficiency).

## Substitution Rates and Functional Constraints

It has often been stated that the correlation between high substitution rates and low functional constraints is an important argument in favor of the neutral theory. In fact, the correlation is compatible with a selectionist as well as with a neutralist interpretation.

Kimura (1989) restates as follows what he considers to be an eminently proneutralist and antiselectionist feature of molecular evolution: "Molecules or portions of molecules that are subject to less functional constraint evolve faster (in terms of mutant substitutions) than those that are subject to stronger constraints."

Kimura and Ohta (1973) recognize, however, that "the fact that the rate of evolution is very low at functionally critical sites may also be explained readily by the traditional theory based on positive Darwinian selection"—plus, one should of course add, on negative selection. Indeed, from the very beginnings of polypeptide sequence comparisons, it was clear that a low rate of substitution must imply a high degree of functional constraint and vice versa. It is the vice versa, namely a high rate of substitution implying a low rate of functional constraints, that is claimed by Kimura and Ohta (1973) to be a pivotal criterion. The authors write: "Under the neutral theory, the less stringent the constraint, the larger the fraction of mutations that are selectively neutral (not harmful) and therefore the higher the evolutionary rate. Thus the upper limit of the evolutionary rate is set by the mutation rate. . . . On the other hand, such an interpretation is meaningless under the positive Darwinian theory. Rather, we should expect that rapidly evolving molecules have important functional significance and are undergoing very rapid adaptive improvements by accumulating many advantageous mutations." In fact, the high rate/low constraint correlation is not at all mean-

ingless from a selectionist viewpoint, even though under the selectionist interpretation the upper limit of the evolutionary rate presumably falls short of the mutation rate.

Rapid rates of substitution generated by positive selection do occur, but only exceptionally, it seems; only in few proteins; or perhaps in many proteins, but during relatively short periods of their evolutionary histories. Phases of rapid rates of positive selection may be expected in particular at the time some important functional change comes about in a protein (Zuckerkandl and Pauling 1965). Goodman et al. (1974, 1975) and Goodman (1981) found support for this view in certain ancestral segments of the reconstructed molecular phylogenetic tree for hemoglobin chains. Kimura (1981, 1983) discussed Goodman's conclusion and found that it was based on assumptions about evolutionary times of split of molecular lineages that were not warranted (see, however, Goodman et al. 1987). Hughes and Nei (1988, 1989) analyzed in the murine major histocompatibility complex those parts of genes that correspond to the antigen recognition sites. They found higher rates of nonsynonymous than of synonymous substitutions, an observation for which a selectionist hypothesis alone seems to be able to account. In general, however, higher amino acid replacement rates are associated with lower selective constraints, as stated. Not only within a neutralist, but also within a selectionist framework, the existence of a correlation between high replacement rates and low constraints is suggested by various circumstances: by the position of the most rapidly substituted amino acid sites mostly at the less functionally constrained surface of globular proteins; by the chemical or steric similarities between many of the frequently interchanged amino acid residues, similarities that should make it functionally easier for them to replace one another; by the mostly small effects of substituents at frequently substituted sites on tertiary structure; and by their mostly moderate effects on physico-chemical parameters associated with function, as shown particularly by studies of hemoglobin (Perutz 1983). The [high functional constraints/low substitution rates—low functional constraints/high substitution rates] correlation has been and can be considered as a generally valid selectionist correlation, with the exceptions to it (namely high functional constraints/high substitution rates) being relatively rare.

In the early days of the field of molecular evolution, it was indeed thought by many that selection probably had its hand even in those high rates of substitutional change that express low functional constraints. For an unknown, potentially important proportion of frequently substituted amino acid sites this continues in fact to be a possible view, as indirectly suggested, in particular, by the work of Kreitman and Hudson (1991) on the intervention of selective forces at linked silent sites. As pointed out above, rapid sequence change in itself does not tell us whether selection is altogether absent and, if absent, whether the neutral drift rate implies functionlessness. Thus, although the rapid evolutionary change of, say, pseudogenes is considered as lending support to the neutral theory, one should merely say that it is compatible with it. It should be clear that this compatibility does not settle the question of possible functions of the DNA included in pseudogenes.

The recurrent use in recent literature of the phrase nonfunctional DNA to designate non-protein-coding DNA, as though the equivalence of these phrases were a matter of course, reminds one of the opposite selectionist prejudices of an earlier time.

That functional constraints, and therefore natural selection, would slow evolutionary sequence changes in informational macromolecules has been assumed quite independently of the neutral theory. Perhaps the first formulation of this recognition was Ingram's. The human $\alpha$-hemoglobin chain binds $\beta$, $\gamma$, or $\delta$ chains, whereas the $\beta$ chain normally binds only one different type of chain, the $\alpha$ chain. In 1961, Ingram pointed out that hemoglobin chains that had to interact with several different partner chains were expected to evolve more slowly than hemoglobin chains that needed to interact with only one different type of partner chain. The idea clearly was that additional functional constraints would slow molecular evolution. A selectionist article published by Linus Pauling and me in 1965 assumed the same correlation.[7] Ingram's prediction was borne out (Derancourt et al. 1967; Kimura and Ohta 1973; Langley and Fitch 1973).

A further prediction is likewise common to the neutralist and selectionist theories and therefore also cannot be used as a test. Kimura (1983) stated: "Eventually, it will be found, if the neutral theory is valid, that molecules or parts of one molecule

[7] This is illustrated by the following quotes, which give by implication the two terms of the rates/constraints correlation: "... Numerous amino acid substitutions are possible because many of the amino acid residues found in proteins are so similar in structure and chemical properties to at least one other type of residue that transitions from one amino acid to the most closely related ones will usually represent a very small modification indeed." [In other words, the greater the frequency of a structural change is, the smaller is the functional change.] ... "We may venture the generalization that the outside of the globin molecule, and perhaps of globular proteins in general, is more variable than the inside. ... It is plausible that this should be so, since it is at the inside of protein subunits that the requirements for steric fit between residues should be the most generalized" [in other words, that special functional constraints are present]

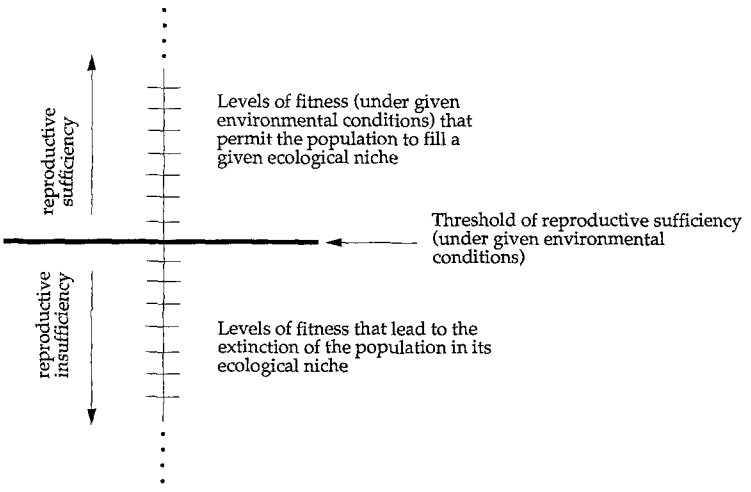


**Fig. 2.** Reproductive sufficiency as a threshold and as a domain of fitnesses above the threshold

which are more important in function and which therefore evolve more slowly, will show a lower level of heterozygosity." This is also expected from a selectionist point of view, as alternative amino acid residues at highly constrained sites will rarely be tolerated functionally. Most selected alleles, like neutral alleles, should therefore be found at the more changeable amino acid sites, where amino acid replacements are more frequently functionally fit than at the more highly constrained sites. Interestingly, Kimura and Ohta (1973) themselves seem to imply that this view is a possible one. "Under the theory [of positive Darwinian selection], the observed low evolutionary rate at functionally critical sites can be explained by saying that the fraction of definitely advantageous mutants is very small at such critical sites. In other words, the more stringent the functional requirements, the lower the chance that mutations turn out to be advantageous."

In conclusion, the rates/constraints correlation argument cannot be considered as an argument in favor of the neutral theory, nor of the selection theory. On account of the originally selectionist view of the rates/constraints correlation, the neutralists' confiscation of the correlation—their proprietary claim to it—came as a real surprise, one renewed every time that the claim is repeated.

## Reproductive Sufficiency and Dispensability of Functions

Both the fitness and the survivability of a genotype under given environmental conditions need to be considered. For this purpose, one may use the concept of reproductive sufficiency, as defined above, the survivability of genotypes in the absence of fitter alleles (called genetic sufficiency in Zuckerkandl 1978). Figure 2 shows that reproductive sufficiency is both a threshold and a zone that encompasses the upper domain of relative, and absolute, fitness. The concept of reproductive sufficiency is closely related to that of adaptedness, which was pertinently distinguished by Dobzhansky (1968) from relative fitness and defined as the ability of a genotype or of a population to survive and reproduce in a given environment. Ayala (1969) has further analyzed and in new ways quantified this concept and has measured adaptedness and its rate of increase by selection in *Drosophila.*

The concept of reproductive sufficiency emphasizes that relative fitnesses smaller than those of a number of existing genotypes can be compatible with adaptedness, provided that none of the fitter genotypes be present in the population. The concept also points to the fact that genotypes that are adapted per se to a given environment will disappear as they become deadapted in the presence of any of the fitter genotypes. The degree of adaptedness depends among other things on the presence or absence of competing species. The conditions for reproductive sufficiency also are both environmental and genetic. The environmental conditions perforce include ecological components. Among these might again be other species that compete for the same resources. Fitter genotypes of the same species are however excluded from these conditions by definition.

In nature, a "fitness treadmill" is operating, namely a constant trend to replace a population by another formed of fitter alleles. There are many levels of adaptedness over and above the population survival threshold. Reproductive sufficiency designates the domain above this threshold value of adaptedness, a threshold that will vary with the environment. Reproductive sufficiency thus is all-or-none. It can certainly be explored experimentally (Zuckerkandl 1978) and is of course to be used as a complement of the concept of fitness, not as an alternative to it.

It might be argued that the presence of a fitness treadmill and thus of alleles fitter than those that

correspond to the threshold level of reproductive sufficiency has a bearing on the chances of survival of the species. The fitness treadmill could offer the population an increased chance of long-term survival. It would do so by providing a buffer zone against untoward changes that would endanger the persistence of the population. This may be so at times and not so at others. An allele that increases the average fitness of the individuals that carry it will not usually happen to be preadapted to the next environmental challenge, the next threat to the persistence of the population. Alleles that were the fitter ones under preceding environmental conditions may be of little help for responding successfully to the new conditions. Thus, the threshold of reproductive sufficiency of a genotype under given environmental conditions relates to a minimum state of a genetic system that is compatible not only with short-term survival of the system but rather frequently, in principle, also with a normal-term survival. Such survival, however, is not testable because in nature the fitness treadmill does not remain fallow for long. Ascertaining reproductive sufficiency thus is a way of putting in parentheses part of biological reality, in order to identify another part, namely intrinsically essential biological traits for persistence of a population in a given environment. It seems that this type of observation is required, if eventually the relationships between genetic organizations and environments are to be thoroughly understood.

It is well known (see Wilson et al. 1977; Zuckerkandl 1991) that certain proteins whose sequence has quite obviously been selected for during evolution and whose structural adulterations have been selected against can be lost without this loss compromising the survival or fertility of the null mutant. Human serum albumin and lysozyme are such proteins. Dispensability and functionality thus can go hand in hand. Nonetheless, it is commonly thought that noncoding sequences that are present in some species and absent in other closely related ones clearly manifest their nonfunctionality through their ability to vanish. If even coding sequences can sometimes be lost, then less-constrained noncoding sequences surely can be lost very much more frequently, without their disappearance implying nonfunctionality.

It is important to realize that the loss of a sequence may not compromise reproductive sufficiency, whereas the presence of this sequence may increase the individual's fitness, as, in a proportion of cases, competition experiments would probably show. In other words, there can be function even when its absence fails to endanger the species. In higher organisms, function is likely often to be compatible with either sequence loss or sequence multiplication.

What results of mutational processes are to be considered as biologically important? Are functional sequences that are dispensable important because they are functional or unimportant because they are dispensable? Neither answer much helps our understanding. It appears that a concept of evolutionary importance needs to be linked to evolutionary innovation, i.e., to the appearance of a function that is new or regulated in a new way or altered more profoundly than by a variation, within a limited range, of some physico-chemical parameters.

The fact is that, upon a closer look, junk DNA and functional noncoding DNA are expected often to behave identically, be it in regard to rapid substitution rates, to "wasteful" multiplication of sequences, or to the continuous elimination and generation of sequences through the processes of DNA tectonics. Functionality in the noncoding genome therefore cannot be negated on the basis of these observations.

*Acknowledgments.* I am indebted for their valuable input to Francisco Ayala, Teni Boulikas, Ford Doolittle, Gerald Holmquist, Jerzy Jurka, George Russev, Greg Spicer, and Edward Trifonov. I express to them my gratitude, as well as to Martha Best for her kindness and patience in preparing the manuscript, to Rosemary Babcock, and to Jolanta Walichiewicz. I am also grateful to Francisco Ayala for having suggested that the phrase "genetic sufficiency" be replaced by "reproductive sufficiency."

## References

Archer TK, Cordingley MG, Wolford RG, Hager GL (1991) Transcription factor access is mediated by accurately positioned nucleosomes on the mouse mammary tumor virus promoter. Mol Cell Biol 11:688–698

Ayala FJ (1969) An evolutionary dilemma: fitness of genotypes versus fitness of populations. Can J Genet Cytol 11:439–456

Barnard EA, Cohen MS, Gold MH, Kim J-K (1972) Evolution of ribonuclease in relation to polypeptide folding mechanisms. Nature 240:395–398

Beckmann JS, Trifonov EN (1991) Splice junctions follow a 205-base ladder. Proc Natl Acad Sci USA 88:2380–2383

Bernardi G, Mouchiroud D, Gautier C, Bernardi G (1988) Compositional patterns in vertebrate genomes: conservation and change in evolution. J Mol Evol 28:7–18

Bolshoy A, McNamara P, Harrington RE, Trifonov EN (1991) Curved DNA without A–A: experimental estimation of all 16 DNA wedge angles. Proc Natl Acad Sci USA 88:2312–2316

Boulikas T (1991) Relation between carcinogenesis, chromatin structure and poly(ADP-ribosyl)ation. Anticancer Res 11:489–527

Courey AJ, Tjian R (1988) Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. Cell 55:887–898

Derancourt J, Lebor AS, Zuckerkandl E (1967) Séquence des acides aminés, séquence des nucléotides et évolution. Bull Soc Chim Biol 49:577–607

Dobzhansky T (1968) On some fundamental concepts of Darwinian biology. In: Dobzhansky T, Hecht MK, Steere WC (eds) Evolutionary biology, vol II. Appleton-Century-Crofts, New York

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

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

Elgin SCR (1990) Chromatin structure and gene activity. Curr Opinion Cell Biol 2:437–445

Felsenfeld G, McGhee J, Rau DC, Wood W, Nickol J, Behe M (1982) Chromatin conformation and gene activity. In: O'Malley BW (ed) Gene regulation. Academic Press, New York, pp 1–135

Gilmour DS, Thomas GH, Elgin SCR (1989) *Drosophila* nuclear proteins bind to regions of alternating C and T residues in gene promoters. Science 245:1487–1490

Glaeser RL, Thomas GH, Siegfried E, Elgin SCR, Lis JT (1990) Optimal heat-induced expression of the *Drosophila hsp26* gene requires a promoter sequence containing $(CT)_n \cdot (GA)_n$ repeats. J Mol Biol 211:751–761

Goodman M (1981) Globin evolution was apparently very rapid in early vertebrates: a reasonable case against the rate-constancy hypothesis. J Mol Evol 17:114–120

Goodman M, Moore GW, Barnabas J, Matsuda G (1974) The phylogeny of human globin genes investigated by the maximum parsimony method. J Mol Evol 3:1–48

Goodman M, Moore GW, Matsuda G (1975) Darwinian evolution in the genealogy of haemoglobin. Nature 253:603–608

Goodman M, Czelusniak J, Koop BF, Tagle DA, Slightom JL (1987) Globins: a case study in molecular phylogeny. Cold Spring Harbor Symp Quant Biol 52:875–890

Hinegardner R (1976) Evolution of genome size. In: Ayala FJ (ed) Molecular evolution. Sinauer, Sunderland MA, pp 179–199

Holmquist GP (1989) Evolution of chromosome bands: molecular ecology of noncoding DNA. J Mol Evol 28:469–486

Hughes A, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167–170

Hughes A, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci? Evidence for overdominant selection. Proc Natl Acad Sci USA 86:958–862

Ingram VM (1961) Gene evolution and the haemoglobins. Nature 189:704–708

John B, Miklos GLG (1979) Functional aspects of satellite DNA and heterochromatin. Int Rev Cytol 58:1–114

Kimura M (1981) Was globin evolution very rapid in its early stages?: a dubious case against the rate-constancy hypothesis. J Mol Evol 17:110–113

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge MA

Kimura M (1989) The neutral theory of molecular evolution and the world view of the neutralists. Genome 31:24–31

Kimura M, Ohta T (1973) Mutation and evolution at the molecular level. Genetics Suppl 73:19–35

Knezetic JA, Jacob GA, Luse DS (1988) Assembly of RNA polymerase II preinitiation complexes before assembly of nucleosomes allows efficient initiation of transcription on nucleosomal templates. Mol Cell Biol 8:3114–3121

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

Langley CH, Fitch WM (1973) The constancy of evolution: a statistical analysis of the $\alpha$ and $\beta$ hemoglobins, cytochrome *c*, and fibrinopeptide A. In: Morton NE (ed) Genetic structure of populations. University Press of Hawaii, pp 246–262

Li W-H, Graur D (1991) Fundamentals of molecular evolution. Sinauer, Sunderland MA

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397–401

Loewenhaupt K, Rich A, Pardue ML (1989) Nonrandom distribution of long mono- and dinucleotide repeats in *Drosophila* chromosomes: correlations with dosage compensation, heterochromatin, and recombination. Mol Cell Biol 9: 1173–1182

Mayr E (1988) Toward a new philosophy of biology. The Belknap Press of Harvard University Press, Cambridge MA

Miklos GLG (1985) Localized highly repetitive DNA sequences in vertebrate and invertebrate genomes. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum Press, New York, pp 241–319

Nelson HCM, Finch JT, Bonaventura FL, Klug A (1987) The structure of an oligo(dA)·oligo(dT) tract and its biological implications. Nature 350:221–226

Ohno S (1972) So much "junk" data in our genome. In: Smith HH (ed) Evolution of genetic systems. Gordon-Breach, New York, pp 366–370

Orgel LE, Crick FHC (1980) Selfish DNA: the ultimate parasite. Nature 284:604–607

Pardue ML, Lowenhaupt K, Rich A, Nordheim A (1987) $(dC\text{–}dA)_n (dG\text{–}dT)_n$ sequences have evolutionarily conserved chromosomal locations in *Drosophila* with implications for roles in chromosome structure and function. EMBO J 6:1781–1789

Perutz MF (1983) Species adaptation in a protein molecule. Mol Biol Evol 1:1–28

Piña B, Brüggemeier U, Beato M (1990) Nucleosome positioning modulates accessibility of regulatory proteins to the mouse mammary tumor virus promoter. Cell 60:719–731

Rogers JH (1985) The origin and evolution of retroposons. Int Rev Cytol 93:187–279

Rogerson AC (1991) There appear to be conserved constraints on the distribution of nucleotide sequences in cellular genomes. J Mol Evol 32:24–30

Serfling E, Jasin M, Schaffner W (1985) Enhancers and eukaryotic gene transcription. Trends Genet 1:224–230

Simpson RT (1990) Nucleosome positioning can affect the function of a cis-acting DNA element in vivo. Nature 343: 387–389

Stalder J, Groudine M, Dodgson JB, Engel JD, Weintraub H (1980) Hb switching in chickens. Cell 19:973–980

Straka C, Hörz W (1991) A functional role for nucleosomes in the repression of a yeast promoter. EMBO J 10:361–368

Stumph WE, Baez M, Lawson GM, Tsai M-J, O'Malley BW (1982) Chromatin structure of the ovalbumin gene domain. In: O'Malley BW (ed) Gene regulation. Academic Press, New York, pp 87–104

Takahata N (1987) On the overdispersed molecular clock. Genetics 116:169–179

Trifonov EN (1983) Sequence-dependent variations of B-DNA structure and protein-DNA recognition. Cold Spring Harbor Symp Quant Biol 47:271–278

Trifonov EN (1989) Viewpoint: the multiple codes of nucleotide sequences. Bull Math Biol 51:417–432

Vashadkidze RP, Chelidze MG, Mamulashvili NA, Kalandarishvili KG, Tsalkalamanidze NV (1988) Nuclear proteins from *Drosophila melanogaster* embryos which specifically bind to simple homopolymeric sequences poly [(dT–dG)·(dC–dA)]. Nucleic Acids Res 16:4989–4994

Vogt P (1990) Potential genetic functions of tandem repeated DNA sequence blocks in the human genome are based on a highly conserved "chromatin folding code." Hum Genet 84: 301–336

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

Workman JL, Roeder RG (1987) Binding of transcription factor TFIID to the major late promoter during in vitro nucleosome assembly potentiates subsequent initiation by RNA polymerase II. Cell 51:613–622

Wu C-I (1990) Book review. Mol Biol Evol 7:286–287

Yee HA, Wong AKC, van den Sande JH, Rattner JB (1991) Identification of novel single-stranded d(TC)$_n$ binding proteins in several mammalian species. Nucleic Acids Res 19: 949–953

Zuckerkandl E (1975) The appearance of new structures and functions in proteins during evolution. J Mol Evol 7:1–57

Zuckerkandl E (1976a) Evolutionary processes and evolutionary noise at the molecular level. I Functional density in proteins. J Mol Evol 7:167–183

Zuckerkandl E (1976b) Evolutionary processes and evolutionary noise at the molecular level. II A selectionist model for random fixations in proteins. J Mol Evol 7:269–311

Zuckerkandl E (1978) Multilocus enzymes, gene regulation, and genetic sufficiency. J Mol Evol 12:57–89

Zuckerkandl E (1981) A general function of noncoding polynucleotide sequences. Mol Biol Rep 7:149–158

Zuckerkandl E (1986) Polite DNA: functional density and functional compatibility in genomes. J Mol Evol 24:12–27

Zuckerkandl E (1987) On the molecular evolutionary clock. J Mol Evol 26:34–36

Zuckerkandl E (1991) Dispensability of parts of histones and the molecular clock. J Mol Evol 32:271–273

Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ (eds) Evolving genes and proteins. Academic Press, New York, pp 97–165

Zuckerkandl E, Villet R (1988) Generation of high specificity of effect through low-specificity binding of proteins to DNA. FEBS Lett 231:291–298

Zuckerkandl E, Latter GE, Jurka J (1989) Maintenance of function without selection: Alu sequences as "cheap genes." J Mol Evol 29:504–512

Received June 21, 1991/Revised and accepted September 10, 1991