Aust. J. Biol. Sci., 1983, **36**, 77–91

Intrinsic Noise and the Design of the Genetic Machinery

D. C. Reanney, D. G. MacPhee and J. Pressing

Department of Microbiology, La Trobe University, Bundoora, Vic. 3083.

Abstract

Darwinian theory envisages 'selection pressure' as a stress imposed on the genotype by the environment. However, noise in the replicative and translational mechanisms in itself imposes a significant 'pressure' on the adaptive fitness of the organism. We propose that the biosphere has been shaped by both *extrinsic* (environmental) and *intrinsic* (noise-generated) factors. Because noise has been a remorseless and ever-present background to the evolutionary process, adaptations to this intrinsic pressure include not only a variety of familiar genetic mechanisms but also many anatomical and life-style characteristics that focus on the transmission of information between generations.

Introduction

No system of information transmission can be 100% efficient. Copying errors occur during polynucleotide replication (Drake and Baltz 1976; Eigen and Schuster 1977) while both resting and replicating DNA can accumulate mutations as a result of exposure to a variety of agents such as heat (Drake and Baltz 1976) radiation (Schendel 1981) and mutagenic chemicals (Schendel 1981). Also, a variety of non-inheritable errors occur during transcription and translation. These non-inheritable errors normally outnumber inheritable errors. For example the error rate in protein synthesis at the level of discrimination between similar amino acids is about 3 in 10^4 while the error rate in DNA synthesis is about 1 in 10^8–10^9 (Hopfield *et al.* 1976).

Errors which occur during the storage, copying or expression of genetic information constitute *noise* in the sense implied by Shannon (1949) in his classic theorem on information transmission. It is the purpose of this paper to describe not the source of such noise but its consequences.

Consequences of Noise in Information Transmission

We find that noise has strongly influenced the design of the genetic machinery throughout evolution. Four different strategies have evolved:

(1) noise compensation;

(2) noise correction;

(3) noise reduction;

(4) special selection against noise-damaged genes.

0004–9417/83/010077$02.00

It is easy to confuse these in any evaluation of genetic structure because many contemporary organisms use all four strategies. Each process is however a distinct adaptive response to the noise factor.

Noise Compensation

The deleterious effects of noise can be masked if the damage is dispered across a number of 'information modules'. We call this strategy noise *compensation*. The most widespread form of compensation is redundancy.* According to Shannon's theorem this may take two forms: *message* redundancy or *code* redundancy. The qualitative effects of the two forms of redundancy are similar and we have chosen to give a detailed model only for message redundancy.

Consider an information module consisting of L nucleotides. If q is the nucleotide copying fidelity, a fraction q^L of all such information modules will have the correct information. If two copies of a given module are 'coupled' (this coupling need not necessarily involve covalent linkage) then the correct information will be absent only when both modules are damaged. Hence the fraction of correct copies in this case is

$$\begin{aligned} f &= 1-(\text{fraction with } \geqslant 1 \text{ error in each copy}), \\ &= 1-(1-q^L)^2, \\ &= q^L(2-q^L), \end{aligned}$$

and for the general case of n-fold redundancy

$$f = 1-(1-q^L)^n. \tag{1}$$

Hence the selective advantage, K, of information transfer due to redundancy is

$$K = f/q^L = [1-(1-q^L)^n]/q^L. \tag{2}$$

A plot of n (copy number) against K (selective advantage of information transfer due to redundancy) shows that K typically rises rapidly for small n values and approaches q^{-L} asymptotically (Fig. 1). K also rises with $1-q$ and L so that the selective value of redundancy increases as the amount of noise in the system gets larger.

A similar mathematical model may be provided for redundancy at the level of the genetic code. The synonym codons for amino acids like glycine can be considered as redundant copies of a module of information in which the expected errors generated by noise have already occurred in positions not subject to selective constraint: thus *GG*U, *GG*C, *GG*A, *GG*G. The mathematical treatment in this case shows that n-fold redundancy is most effective when variation is restricted to one codon position (Sonneborn 1965; Goldberg and Wittes 1966). Selection has evidently balanced n against K in different ways for different amino acids, depending on the physico-chemical features of the amino acids themselves. The amino acid which occurs most rarely, tryptophan (Dayhoff 1976) has an n value of 1 while glycine, the most abundant of all amino acids in proteins (Dayhoff 1976), has an n value of 4. The mid-frequency amino acid proline (Dayhoff 1976) also has an n value of 4, presumably because proline occurs at 'corners' in the α-helical structure of proteins (Dickerson and Geis 1969). The strong hydrophobe leucine has an n value of $4+2$. If the hydrophobes

* We use this term as a synonym for 'reiteration' without necessarily implying superfluousness.

are considered as a group whose detailed features are less important than their common characteristic of hydrophobicity then the redundancy value rises to a peak of $n = 16$ since invariance is restricted to the central residue. The reason for this high redundancy is almost certainly the determinative role played by hydrophobic residues in protein tertiary structure (Perutz *et al.* 1965). STOP triplets have an n value of 3 in the conventional code and 2 in the code of mitochondria (Barrell *et al.* 1979). It may be of evolutionary significance that the high n values for eight amino acids with fourfold redundancy can be achieved by single tRNA species in mitochondria (Bonitz *et al.* 1980).

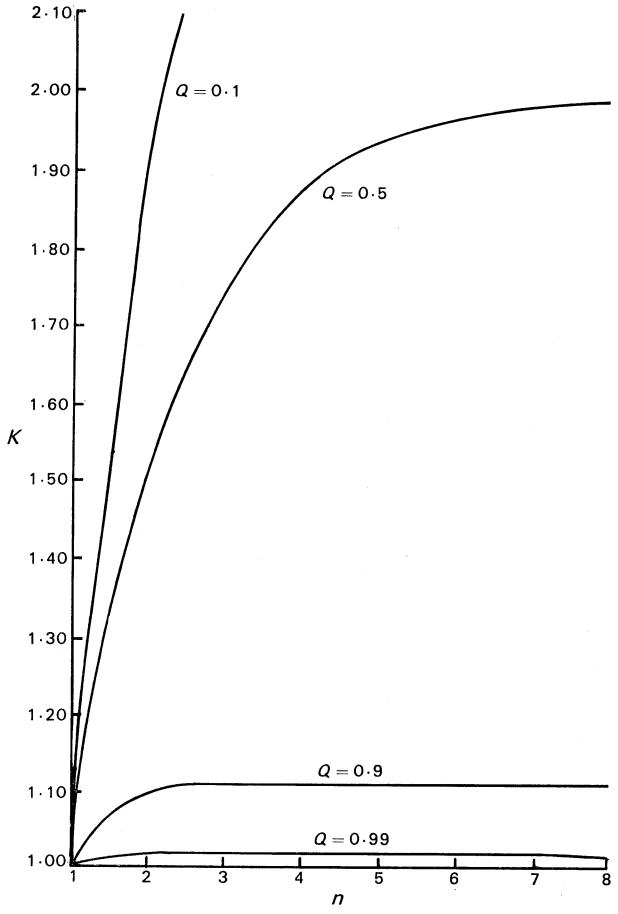

Fig. 1. Selective advantage of information transfer due to redundancy, K, as a function of number of copies, n, for selected values of $Q = q^L$ (cf. equation 2).

There is widespread consensus that redundancy patterns in the genetic code are an adaptive response to noise in the replicative (Sonneborn 1965; Goldberg and Wittes 1966) and decoding (Woese 1967; Ycas 1969) systems in cells. It is puzzling, therefore, that this principle has not been carried over to explanations of redundancy at the level of the gene and genome. Consider gene duplication.* Differentiated cells

* The above arguments do not appear to apply to 'repeat' DNA fractions in eukaryotes.

may contain 50 000 copies of the mRNA for their characteristic protein(s) even though the gene is represented once only in the genome (Long and Dawid 1980). So dosage effects need not be invoked for most families of repeated genes apart from those for ribosomal RNAs, transfer RNAs, histones and perhaps a few others (Long and Dawid 1980). The conventional explanation for gene reiteration is that 'duplication' provides for further evolution by liberating one copy from the policing action of natural selection (Ohno 1970). However, the immediate and obvious effect of duplication is to protect the system by ensuring that mutations which would be harmful in a single-copy situation no longer affect the survival chances of the organism (in the extreme case 'lethal' mutations are no longer lethal). This protective effect must be considered as the *primary* selective advantage of duplication since a very long period of evolutionary time may elapse before accumulated mutations in either copy generate a protein with a new phenotype which gratuitously enhances the adaptive fitness of the organism.

In assessing the protective effects of genetic redundancy it is essential to take into account those factors which might limit values of n. One can imagine several such constraints but we focus here upon the fact that increasing values of n place an increasing burden on the cell's energy reserves. The average number of correct information sites, i.e. nucleotides, in the n-fold redundancy case is increased over that in the non-redundant case to Lf sites. At some stage the total energy cost to the cell will be Ln nucleotides (remembering that nucleotide triphosphates, especially ATP, are both nucleic acid precursors *and* sources of usable energy). Let e be an efficiency of information transfer. Then

$$\begin{aligned} e &= \text{length of correctly transmitted information} \div \text{length of message} \\ &= \text{number of correct nucleotides} \div \text{number of nucleotides} \\ &= Lf/Ln = [1-(1-q^L)^n]/n. \end{aligned} \tag{3}$$

Inspection of this equation shows that the energy cost to the system will vary significantly with L, the length of the information module. This allows us to distinguish two strategies of genetic redundancy: (i) relatively short information modules repeated on the same DNA molecule, and (ii) repetition of the entire genome.

The first strategy is that of gene reiteration in the sense that tandemly repeated sequences are present on the same DNA chain (Ohno 1970). Redundancy due to the reiteration of individual genes has the advantage that protection can be achieved at a relatively small energy cost, where the genome is large. However, this advantage must be balanced against some important disadvantages: unless *each* gene has an appropriate value of n, harmful mutations in non-redundant genes can still adversely affect the survival fitness of the organism. The values of n needed to protect target genes could be unrealistically high in systems with a low copying fidelity. Moreover, even in accurately copied systems, repeated genes are unstable because the mechanisms that lead to their genesis (e.g. by crossing-over—Edlund and Normark 1981) can also lead to their loss by the same mechanism (Ptashne and Cohen 1975; Potter and Dressler 1977).

The second strategy extends redundancy to the whole DNA molecule. A valuable insight into the protective effect of such a strategy may be provided by a consideration of recent findings concerning the extremely radiation-resistant prokaryote *Dienococcus radiodurans* (formerly *Micrococcus radiodurans*). Renaturation kinetics indicate that

the genome of *D. radiodurans* has a molecular weight $2 \pm 0 \cdot 3 \times 10^9$ (Hansen 1978). The number of genome equivalents calculated for this bacterium works out at 4–10 whole DNA molecules per cell (Tirgari and Moseley 1980), and it has been suggested that these genome copies must be non-associated and capable of independent segregation. In this situation chance works for the preservation of genetic identity following exposure of *D. radiodurans* to ionizing radiation since the random nature of mutation ensures that the chances of a given *single* gene being 'hit' in *all* reiterated copies are remote. In addition, it has been suggested that *D. radiodurans* gains additional protection against radiation damage by the presence of all four genomes (in resting cells) in a single 'nucleus' (Tirgari and Moseley 1980). In this conformation, recombinational repair between homologous regions of all four chromosomes could well be possible, thereby contributing to the removal of radiation damage from one or more of the four chromosomes which eventually segregate as independent units.

In the above case, genome redundancy may be seen as an adaptation to externally provoked noise (radiation), but since the ultimate deleterious effects of radiation on DNA molecules to a considerable extent mimic those of copying infidelity, it can also be seen as a model of an adaptation to intrinsic noise.

The relatively high value of n for the *D. radiodurans* genome required to achieve a sufficient overall fidelity has evidently been 'bought' at a high energy cost (see equation 3). To consider more normal cases equation (3) may be used with e as a function of n to obtain a general expression for $e(2)-e(n)$ which may readily be shown to be >0 for all values of $n > 2$. It follows therefore that *twofold redundancy is the optimal energy-efficient strategy*. We believe this is a major reason why *diploidy* is the favoured mechanism of error compensation in eukaryotes, which contain more noise per genome than prokaryotes, if only because eukaryotic genomes are larger than prokaryotic. Furthermore, from equation (3), $e(1)-e(2) = \frac{1}{2}Q^2$, with $Q = q^L$, which means that the reduction of efficiency through redundancy is nearly 50% in high-fidelity systems. Since the increase in K is small for high values of Q, *for sufficiently high Q, redundancy must fail to be advantageous*. We belive the haploid ($n = 1$) nature of most prokaryotic genomes and their relative lack of gene repetition are consequences of the efficiency of bacterial corrective processes which allow prokaryotes to exploit their energy reserves for information transmission in a most effective way. The relationship between e and n is shown in Fig. 2.

The relationship between repeated genes and 'cost effectiveness' is confused by the fact that many eukaryotic genomes contain large amounts of repetitive DNA whose function(s), if any, are unknown. While the values of n for these repeat sequences can be very high, we do not suggest that redundancy of this particular type has 'protective' value. We wish to make the point, however, that selection cannot, in our view, remain completely indifferent to the energy burden which the replication of 'useless' DNA imposes on the system over long periods of evolutionary time. Within groups of organisms that have diverged from a common ancestor there is a well-established trend for the more recently evolved forms to contain *less* DNA than the 'primitive' ones (Mirsky and Ris 1951; Stebbins 1966; Hinegardner 1968; Bostock 1971). Further, competition experiments between plasmid-plus and plasmid-minus strains of the same bacterium in chemostats suggest that the bacterium with the extra DNA is at a competitive disadvantage (Melling *et al.* 1977; Wouters *et al.* 1978; Godwin and Slater 1979). These data are most easily rationalized in terms of energy economy arguments such as we have presented above. Indeed, one can argue

that the relatively small, non-repetitive genomes of prokaryotes are the end results of processes in which virtually *all* redundant sequences have been eliminated so as to 'streamline' the genome and optimize its reproductive fitness (Reanney 1974; Doolittle 1979, 1982). The same reasoning can be applied to viruses, which are undoubtedly under severe pressure to compact their genetic information (Doolittle 1982; Reanney 1982).

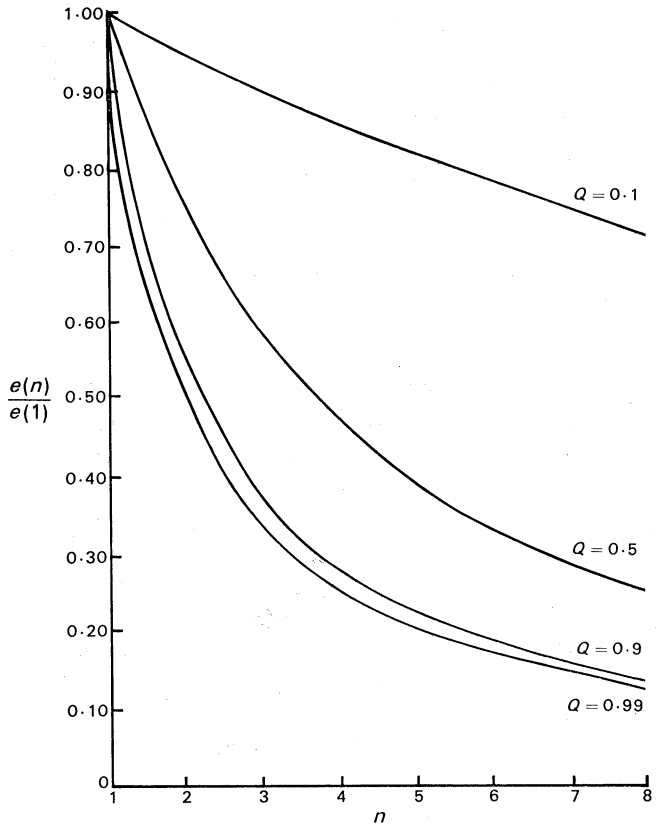

Fig. 2. Relative efficiency of error compensation due to redundancy, $e(n)/e(1)$, as a function of number of copies, n, for selected values of $Q = q^L$ (cf. equation 3).

The issue of repetitive sequences in eukaryotes is complicated by the possibility that the generation or (short-term) maintenance of these sequences may not be curtailed by selection at all, at least not in a conventional sense. Dover's concept of 'molecular drive' allows sequences to spread through a population in a non-Mendelian fashion by biased gene conversion or other mechanisms (for discussion, see Dover 1982). The same may be true of the spread of modular units of 'selfish' DNA (Doolittle and Sapienza 1980; Orgel and Crick 1980).

While genetic redundancy is the most widespread of all error compensating mechanisms developed during evolution it is not the only one. It is possible to enhance the efficiency of information transfer not by repeating the information (i.e. by expanding the length of the message), but by dividing the information into smaller units. The smaller a module of information the greater its chances of passing undamaged

through a noisy channel. This strategy of genome segmentation has been followed by many RNA viruses and it is significant that viral RNAs, which lack corrective copying functions, have high noise levels of 10^{-3}–10^{-4} errors per nucleotide per generation (Holland *et al.* 1982; Reanney 1982).

A mathematical treatment developed by Pressing and Reanney (1983) shows that the selective advantage, K, of genome subdivision initially increases as the number n of different modules gets larger. Indeed, the graphical relationship of K versus n for segmental genomes has a similar form to Fig. 1 (Pressing and Reanney 1983), provided that the various modular RNAs are united in a single viral capsid (monoparticulate viruses). Where the modular RNAs are separately encapsidated (multiparticulate viruses) the equation describing the selective advantage of genome subdivision has the form

$$K = \lambda^{n-1} q^{[1/(n-1)]L},$$

where λ is an average fraction of cells infected.

It can be shown that the predicted values of n for modular RNAs are relatively high in the cases of monoparticulate viruses (the observed range is 4–12) and relatively low in the case of multiparticulate viruses (the observed range is 2–3) (see Pressing and Reanney 1983).

Noise Correction

The copying of polynucleotides in the absence of enzymes has been studied by Orgel (1973) and others as part of the pre-life program. Non-enzymatic copying is very mistake-prone with an error frequency of about 10^{-1}–10^{-2} mutations per base per replication (Eigen and Schuster 1977). This high noise level limits the amount of information which can be reproducibly copied to about 100 nucleotides or approximately the size of a tRNA precursor (Eigen and Schuster 1977). The introduction of polymerases reduces the noise level to 10^{-3}–10^{-4} (Eigen and Schuster 1977) which is the measured copying fidelity of contemporary RNA viral replicases (Holland *et al.* 1982). An error rate of 10^{-4} would allow genome sizes to expand to the values found in non-divided RNA viral genomes (about 20000 bases) (Eigen and Schuster 1977). DNA replicases may not of themselves be appreciably more accurate than RNA replicases. *E. coli* DNA polymerase I, defective in its repair properties, has about the same replicative fidelity as RNA replicases (Eigen and Schuster 1977) and mammalian DNA polymerases *in vitro* catalyse non-complementary base substitutions at frequencies of between 3×10^{-5} (DNA polymerase α) and 1–3×10^{-4} (DNA polymerases β and γ) (Kunkel and Loeb 1981). The enhanced fidelity of DNA synthesis in modern prokaryotic and eukaryotic cells is due almost exclusively to the evolution of corrective mechanisms (Eigen and Schuster 1977). The *E. coli* DNA polymerase I has an associated proof-reading component which allows miscopied information to be excised and recopied (Kornberg 1980) and this activity together with the action of accessory (single strand binding) proteins lowers the error rate to about 10^{-7} (Kunkel *et al.* 1979). Post-replicative mismatch corrections can reduce the noise level by a further 1–3 orders of magnitude (Nevers and Spatz 1975).

Among the most thoroughly studied corrective processes are those which repair lesions introduced into DNA by external agents such as ultraviolet light or various mutagenic chemicals (Schendel 1981). Leaving aside the phenomenon of photoreactivation (which appears to involve direct enzymatic cleavage of ultraviolet-

induced pyrimidine dimers on a single-strand of DNA), we note that the various known DNA repair processes appear to depend quite fundamentally upon the already discussed phenomenon of redundancy for their various mechanisms of action. For example, both the conventional excision repair system initiated by the products of the *uvrA*, *uvrB* and *uvrC* genes, and the later steps in the methylation-directed mismatch repair mechanism, appear to involve enzymatic filling of a gap in one chain of a DNA duplex using the information specified by the intact opposite strand (Radman *et al.* 1978; Howard-Flanders 1981). Further redundancy appears to be required for post-replication repair, a process in which *both* the daughter double-helical molecules, generated when cells attempt to replicate damaged parental strands containing non-coding lesions (such as pyrimidine dimers), are involved in recombination-like interactions. In eukaryotes, the observation that many agents which damage DNA also increase the frequency of sister chromatid exchange (Perry and Evans 1975) supports the concept that this type of reciprocal recombination (between a newly duplicated chromatid and its sister—see Taylor *et al.* 1957) is another manifestation of what is basically a DNA repair process. Such a suggestion has previously been advanced by Sasaki (1977).

Overall, most lesions in even quite heavily damaged DNA appear to be repaired accurately (if they are repaired at all). Thus the many repair and regulatory systems which deal with damaged DNA appear to be the result of very strong selection to eliminate unacceptable lesions in, or scramblings of, the genetic information. If this is the case, and if in particular post-replication repair involving a recombination-like process is a mechanism which preserves the *integrity* of the information encoded in DNA, one is led to ask whether the whole system of generalized (reciprocal) recombination should not be viewed as *essentially a corrective mechanism* (Reanney 1977, 1979). Generalized recombination occurs at all levels of genetic organization from the *red* system of the *E. coli* phage λ to the synapsed chromosomes of plant and animal cells (Reanney 1977). A popular explanation of the phenomenon is that reciprocal recombination exchanges sequences from different lineages and so expands the pool of variation available to selection (Maynard-Smith 1977). However, the strand isomerization and branch migration which occur during reciprocal recombination show up mismatched sequences as heteroduplex lesions (Reanney 1977; Kornberg 1980). The *E. coli* example shows that such mismatches can be corrected by using methyl groups as markers to discriminate between parental and newly made strands (Radman *et al.* 1978). The notion that so widespread and complex a process as reciprocal recombination had its origins in the need to generate variation must now be viewed with some suspicion since illegitimate recombination, transposition and other factors (Reanney 1976; Campbell 1981) seem capable of generating more novelty than is required to sustain evolution at the observed rates.

Correction due to heteroduplex repair should not be equated with 'concerted evolution'—the apparent inability of certain families of repeated genes to evolve independently (see Zimmer *et al.* 1980; Jeffreys 1982). Concerted evolution may occur through the 'homogenization' of sequences by 'gene conversion' in the sense that fungal geneticists often use the term. But conversion in mammalian cells may not be automatically associated with reciprocal recombination and concerted evolution in higher plants and animals may be mediated by unequal crossing-over or transposition or both (Dover 1982).

Noise Reduction

Most corrective mechanisms exploit stereochemical selection for the 'correct' nucleotides or against 'incorrect' conformations. However, noise levels are affected by *milieu* effects such as the concentrations of cations like Mn^{2+} (Burnet 1974). The reduction in error rates which can be achieved by draining a replicating complex of Mn^{2+} (for example) has parallels at other levels. Consider the heat-induced deamination of C to U in mammal DNA. Baltz *et al.* (1976) have calculated that the minimal level of heat-induced noise in the genome of a large organism like man is about 100 mistakes per diploid cell per day. While some of these mistakes may be repaired by human homologues of the *N*-glycosidases found in *E. coli* (Lindahl 1974), enough noise remains to pose an unacceptable burden to future generations.

Baltz *et al.* (1976) point out that the temperature of the nude scrotum in males is 4–5°C less than that of somatic organs. This lower temperature results in a twofold reduction in the amount of heat-induced mutation. The location of the mammalian testes *outside* the body thus has the effect of *halving* the level of heat-induced noise in resting DNA in the germinal cells to which the future of the species is committed. This example is important in the context of our argument because it suggests that major features of anatomical evolution may be viewed, at least in part, as adaptive responses to intrinsic noise.

Special Selection against Noise-damaged Genes

To the above three distinct strategies which minimize the effects of genetic noise we must now add a fourth. In most higher organisms most genes (over 99% of the genetic mass of the organism) are contained in somatic cells which make no known contribution to future evolution (except in rare cases of viral pick-up and transmission). Continuity is relegated to a minority of genes—those contained in the germ cells. As Orgel (1963) and others have pointed out, there must be some special mechanism in germ-line tissue which maintains the functional integrity of gametic genes.

Germ-line DNA has unique features which reflect the influences of intrinsic selection pressure. First, germ-line cells are laid down early in embryogenesis (Edwards 1980; Mann and Lutwak-Mann 1981). This means that germ-line genes, especially in females, pass through a smaller number of generations than most somatic genes. Since copying is a major source of genetic noise, germ-line DNA's are less likely to accumulate errors than their somatic derivatives. The second and key feature of the germ-line DNA's survival strategy is embodied in the very nature of the sexual mechanism itself. Sexual reproduction relegates the species blueprint to *two independently evolving* lines, male and female (in effect this brings the 'true' redundancy value of the species autosomal DNA up to 4 since both partners are needed for continuity). This division allows ordinary Darwinian selection to evaluate the fitness of the genotype at the $2n$ level. The radical aspect of the sexual strategy is the subsequent *elimination of redundancy* in the haploid (n) gametes and the provision for a *special form of intense selection* on just those DNAs most likely to contribute to future generations. This allows key genes to be screened for errors in the absence of protective partners.

Our thesis can be illustrated by examining spermatogenesis in mammals. The primordial germ cells divide into $2n$ spermatogonia which then give rise to primary and secondary spermatocytes and finally spermatids with the haploid DNA complement. The elimination of redundancy precedes the differentiation of the motile

spermatozoa (Mann and Lutwak-Mann 1981). *It is evident then that the male gametes are built by single-copy* (n) *DNAs in such a way that the noise accumulated during previous divisions can be expressed in target genes.* How are these damaged genes monitored? In most cases known to us reproductive strategy requires a (large) population of sperm to pass multiple barriers before reaching the egg. The barriers may be spatial (e.g. distances between source of sperm and eggs), physical (e.g. the viscosity of the supporting medium, chill factors and currents in the surrounding milieu) or biological (e.g. eggs are usually surrounded by membranes which can usually only be penetrated by 'successful' sperm). Whatever the details it is obvious that the chances of successful fertilization are not randomly distributed among the population of sperm—they are heavily skewed towards its most 'active' members (Mann and Lutwak-Mann 1981). The sexual process in effect forces male gametes through a kind of 'obstacle course' which allows selection to evaluate the fitness of key genes. Defects in the genes of the protein-synthesizing apparatus, for example, will normally be reflected in faulty tail proteins and hence in suboptimal motility (note that the efficiency of the whole translational apparatus can be monitored by testing the performance of just a few proteins).

Does a similar selection operate during oogenesis? A healthy human infant female, for example, contains about 500 000 follicles but only 400 or so of these mature into haploid ova (Edwards 1980). The selection of these 400 single-copy DNA eggs from the potential pool of half a million precursors provides multiple possibilities for discrimination against damaged genes (expressed perhaps as unacceptably altered membrane proteins that could be targets for immunological assault). The formation of the mature egg is preceded by a process during which the primary oocyte divides into a large secondary oocyte and a small first polar body, both of which have non-redundant DNA (Edwards 1980). The second oocyte then divides, again with unequal division of cytoplasm, into the mature ovum and a second polar body. The polar bodies always degenerate and die (Edwards 1980). The 'choice' of one set of single-copy genes as the egg may occur randomly, but we speculate that this 'one-in-four' selection allows a final opportunity to screen out DNAs unfitted to carry on the species blueprint. These aspects of oogenesis deserve further investigation.

Even after fertilization, selection against damaged genes may continue. Numerous studies have shown that 50–60% of spontaneously aborted foetuses in the first trimester are chromosomally abnormal (Simpson 1980). In a systematic study Kajii *et al.* (1980) found, for example, that 53·5% of 402 unselected consecutive aborted foetuses had chromosomal abnormalities. By contrast, the value for the chromosomal abnormalities among 24 468 consecutive live-born children was 0·51% (Jacobs 1972). Early elimination of a non-viable embryo would be a favoured strategy in evolution as it would minimize the reproductive burden on the mother and shorten the time before a second pregnancy could occur. How the severely damaged genomes bring about the spontaneous expulsion of the foetus [418 of the 563 aborted foetuses examined by Kajii *et al.* (1980) were spontaneously expelled] is unknown. More subtle forms of genetic damage may also lead to abortion due to either (i) failure to implant or (ii) early elimination after implantation.

In addition to the separation of sexual roles that occurs in vertebrates, the familiar alternation of generations between $2n$ and n phases in ferns and in (some) green algae is, if our hypothesis is correct, equally a result of intrinsic pressures and a testi-

mony to the importance of noise in evolution. Likewise the *independent assortment* of single-copy gametic DNAs can be viewed as a conserving mechanism since the random shuffling of any two heterozygous alleles A (dominant, i.e. non-mutated) and *a* (recessive, i.e. mutated) limits the chances of appearance of the *jointly damaged* homozygote *aa* to 1 in 4. We find it surprising that the generally excepted explanation of the ploidy–sex phenomenon continues to be the teleological argument that sex provides variation 'for' future evolution. The 75% probability of protecting function through the *AA Aa* and *aA* combinations of 'good' and 'bad' genes, is much more striking in evolutionary terms than the 25% chance of exposing a new and usually faulty phenotype (*aa*) to the test of Darwinian selection.

We conclude then that 'sex' as a phenomenon is important in evolution chiefly because it counteracts the effects of damaged genes through error compensation and because it screens out faulty signals in the most crucial survival channel of all—the link between the generations.

Influence of Noise in Evolution

As far as we are aware, the first mention of noise as a major factor in evolution was made by Orgel (1963) in his classical statement of the error-catastrophe hypothesis. A relationship inversely linking copying error with the maximum information capacity (v_{max}) of a genetic system was proposed by Eigen and Schuster (1977) but this failed to develop or explain the importance of noise in the design of the genetic machinery. Indeed, the Eigen-Schuster thesis—which states that the greater v_{max} of the $n = 2$ eukaryotes over the $n = 1$ prokaryotes is related to their improved fidelity of information transfer—is inconsistent with the fact that anti-mutator strains of some coliform bacteria can achieve error rates as low as 10^{-10} (Kornberg 1980). This would allow bacteria to attain a theoretical v_{max} value of at least 10^{10} (i.e. more than the average genome size of mammals). In terms of our model the non-redundant, haploid character of prokaryotic genomes was only made *possible* by the high fidelity of their information-transmitting processes and the small sizes of their chromosomes probably reflect strong selection for genome economy (Reanney 1974).

The chief thrust of our argument in this paper is that key aspects of the information-transmitting machinery have been moulded chiefly by *intrinsic* factors, i.e. noise rather than by the conventional *environmental* pressures envisaged in the Darwinian theory of evolution. This 'intrinsic' form of selective pressure seems likely to have been the overriding factor during the early period of evolution when noise levels were high (from 10^{-2}–10^{-5}) and has, in our view, been responsible for the evolution of most familiar genetic mechanisms, including gene iteration, ploidy, meiosis, crossing-over and generalized recombination and the basic elements of the sexual strategy. We regard with suspicion the arguments which see these phenomena as an adaptive response to a need to provide variation for evolution. This is not to deny that gene duplication, for example, paves the way for divergence of function or that sex stores variations, rather to emphasize that the *mechanisms* which promote crossing-over and sex were chiefly evolved in response to unacceptable levels of noise. That they should inadvertently become generators of diversity was inevitable, given the stochastic character of noise itself. In an important sense, the variety of life is a surface phenomenon, superimposed on mechanisms of genetic 'homeostasis' by the very noise those mechanisms were designed to combat.

References

Baltz, R. H., Bingham, P. M., and Drake, J. W. (1976). Heat mutagenesis in bacteriophage T4: the transition pathway. *Proc. Natl Acad. Sci. U.S.A.* **73**, 1269–73.

Barrell, B. G., Bankier, A. T., and Drouin, J. (1979). A different genetic code in human mitochondria. *Nature* (*London*) **282**, 189–94.

Bonitz, S. G., Berlani, R., Coruzzi, G., Li, M., Macino, G., Nobrega, F. C., Nobrega, M. P., Thalenfeld, B. E., and Tzagoloff, A. (1980). Codon recognition rules in yeast mitochondria. *Proc. Natl Acad. Sci. U.S.A.* **77**, 3167–70.

Bostock, C. (1971). Repetitious DNA. In 'Advances in Cell Biology'. Vol. 2. (Eds D. M. Prescott, L. Goldstein and E. McConkey.) pp. 153–224. (Appleton-Century-Crofts: New York.)

Burnet, F. M. (1974). 'Intrinsic Mutagenesis: A Genetic Approach.' (Medical and Technical Publishing Co. Ltd.: Lancaster.)

Campbell, A. (1981). Evolutionary significance of accessory DNA elements in bacteria. *Ann. Rev. Microbiol.* **35**, 55–83.

Dayhoff, M. O. (1976). 'Atlas of Protein Sequences and Structure.' Vol. 5. Suppl. 2. p. 301. (National Biomedical Research Foundation: Washington, D.C.)

Dickerson, R. E., and Geis, I. (1969). 'The Structure and Action of Proteins.' (Harper and Row: New York.)

Doolittle, W. F. (1979). Genes in pieces: were they ever together? *Nature* (*London*) **272**, 581–2.

Doolittle, W. F., and Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature* (*London*) **284**, 601–3.

Doolittle, W. F. (1982). Selfish DNA after fourteen months. In 'Genome Evolution'. (Eds G. A. Dover and F. B. Flavell.) pp. 3–28. (Academic Press: New York.)

Dover, G. (1982). Molecular drive: a cohesive mode of species evolution. *Nature* (*London*) **299**, 111–17.

Drake, J. W. (1974). The role of mutation in microbial evolution. *Symp. Soc. Gen. Microbiol.* **24**, 41–58.

Drake, J. W., and Baltz, P. H. (1976). The biochemistry of mutagenesis. *Ann. Rev. Biochem.* **45**, 11–37.

Edlund, T., and Normark, S. (1981). Recombination between short DNA homologies causes tandem duplication. *Nature* (*London*) **292**, 269–71.

Edwards, R. G. (1980). 'Conception in the Human Female.' (Academic Press: London.)

Eigen, M., and Schuster, P. (1977). The hypercycle. *Naturwissenschaften* **64**, 541–65.

Godwin, D., and Slater, J. H. (1979). The influence of growth environment on the stability of a drug resistance plasmid in *Escherichia coli* K12. *J. Gen. Microbiol.* **111**, 201–10.

Goldberg, A. L., and Wittes, R. E. (1966). Genetic code: aspects of organization. *Science* (*Wash., D.C.*) **153**, 420–4.

Hansen, M. T. (1978). Multiplicity of genome equivalents in the radiation-resistant bacterium *micrococcus radiodurans*. *J. Bacteriol.* **134**, 71–5.

Hinegardner, R. (1968). Evolution of cellular DNA content in teleost fishes. *Am. Nat.* **102**, 517–23.

Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., and Vandepol, S. (1982). Rapid evolution of RNA genomes. *Science* (*Wash., D.C.*) **215**, 1577–85.

Hopfield, J. J., Yamane, T., Yue, V., and Coutts, S. M. (1976). Direct experimental evidence for kinetic proofreading in amino acylation of $tRNA^{Ile}$. *Proc. Natl Acad. Sci. U.S.A.* **73**, 1164–8.

Howard-Flanders, P. (1981). Inducible DNA repair. *Sci. Am.* **245**, 56–64.

Jacobs, P. A. (1972). Chromosome mutations—frequency at birth in humans. *Humangenetik* **16**, 137–40.

Jeffreys, A. J. (1982). Evolution of globin genes. In 'Genome Evolution'. (Eds G. A. Dover and R. B. Flavell.) pp. 157–76. (Academic Press: New York.)

Kajii, T., Ferrier, A., Niikawa, N., Takahara, H., Ohama, K., and Avirachan, S. (1980). Anatomic and chromosomal anomalies in 639 spontaneous abortuses. *Hum. Genet.* **55**, 87–98.

Kornberg, A. (1980). 'DNA Replication.' (W. H. Freeman and Co.: San Francisco.)

Kunkel, T. A., Meyer, R. R., and Loeb, L. A. (1979). Single-strand binding protein enhances fidelity of DNA synthesis *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 6331–5.

Kunkel, T., and Loeb, L. A. (1981). Fidelity of mammalian DNA polymerases. *Science* (*Wash., D.C.*) **213**, 765–7.

Lindahl, T. (1974). An N-glycosidase from *Escherichia coli* that releases free uracil from DNA containing deaminated cytosine residues. *Proc. Natl Acad. Sci. U.S.A.* **71**, 3649–53.

Long, E. O., and Dawid, I. B. (1980). Repeated genes in eukaryotes. *Ann. Rev. Biochem.* **49**, 727–64.

Mann, T., and Lutwak-Mann, C. (1981). 'Male Reproductive Function and Semen.' (Springer Verlag: New York.)

Maynard-Smith, J. (1977). Why the genome does not congeal. *Nature* (*London*) **268**, 693–6.

Melling, J., Ellwood, D. C., and Robinson, A. (1977). Survival of R factor carrying *Escherichia coli* in mixed cultures in the chemostat. *FEMS Microbiol. Lett.* **2**, 87–9.

Mirsky, A. E., and Ris, A. (1951). The deoxyribonucleic acid content of animal cells and its evolutionary significance. *J. Gen. Physiol.* **34**, 451–62.

Nevers, P., and Spatz, H.-C. (1975). *Escherichia coli* mutants *uvrD* and *uvrE* deficient in gene conversion of λ-heteroduplexes. *Molec. Gen. Genet.* **139**, 233–43.

Ohno, S. (1970). 'Evolution by Gene Duplication.' (George Allen and Unwin: London.)

Orgel, L. E. (1963). The maintenance of the accuracy of protein synthesis and its relevance to ageing. *Proc. Natl Acad. Sci. U.S.A.* **49**, 517–21.

Orgel, L. E. (1973). 'The Origins of Life'. (Chapman and Hall: London.)

Orgel, L. E., and Crick, F. H. C. (1980). Selfish DNA: the ultimate parasite. *Nature* (*London*) **284**, 604–7.

Perry, P., and Evans, H. J. (1975). Cytological detection of mutagen-carcinogen exposure by sister chromatid exchange. *Nature* (*London*) **258**, 121–5.

Perutz, M. F., Kendrew, J. C., and Watson, H. C. (1965). Structure and function of haemoglobin. *J. Mol. Biol.* **13**, 669–78.

Potter, H., and Dressler, D. (1977). On the mechanism of genetic recombination: the maturation of recombination intermediates. *Proc. Natl Acad. Sci. U.S.A.* **74**, 4168–72.

Pressing, J., and Reanney, D. C. (1983). Divided genomes and intrinsic noise. *Proc. Natl Acad. Sci. U.S.A.* (In press.)

Ptashne, K., and Cohen, S. N. (1975). Occurrence of insertion sequences (IS) regions on plasmid deoxyribonucleic acid as direct and inverted nucleotide sequence duplications. *J. Bacteriol.* **122**, 776–81.

Radman, M., Villani, A., Boiteux, S., Kinsella, A. R., Glickman, B. W., and Spadari, S. (1978). Replicational fidelity: mechanisms of mutation avoidance and mutation fixation. *Cold Spring Harbour Symp. Quant. Biol.* **43**, 937–46.

Reanney, D. C. (1974). On the origin of prokaryotes. *J. Theoret. Biol.* **48**, 243–51.

Reanney, D. C. (1976). Extrachromosomal elements as possible agents of adaptation and development. *Bacteriol. Rev.* **40**, 552–90.

Reanney, D. C. (1977). Genetic engineering as an adaptive strategy. *Brookhaven Symp. Biol.* **29**, 248–71.

Reanney, D. C. (1979). RNA splicing and polynucleotide evolution. *Nature* (*London*) **277**, 598–600.

Reanney, D. C. (1982). The evolution of RNA viruses. *Ann. Rev. Microbiol.* **36**, 47–73.

Sasaki, M. S. (1977). Sister chromatid exchange and chromatid interchange as possible manifestations of different DNA repair processes. *Nature* (*London*) **269**, 623–5.

Schendel, P. F. (1981). Inducible repair systems and their implications for toxicology. *CRC Crit. Rev. Toxicol.* **8**, 311–61.

Shannon, C. E. (1949). In 'The Mathematical Theory of Communication'. (Eds C. E. Shannon and W. Weaver.) (Univ. of Illinois Press: Urbana, Illinois.)

Simpson, J. L. (1980). Genes, chromosomes and reproductive failure. *Fertil. Steril.* **33**, 107–16.

Sonneborn, T. M. (1965). In 'Evolving Genes and Proteins'. (Eds V. Bryson and H. J. Vogels.) p. 377. (Academic Press: New York.)

Stebbins, G. L. (1966). Chromosomal variation and evolution. *Science* (*Wash., D.C.*) **152**, 1462–9.

Taylor, J. H., Woods, P. S., and Hughes, W. L. (1957). The organization and duplication of chromosomes as revealed by autoradiographic studies using tritium-labelled thymidine. *Proc. Natl Acad. Sci. U.S.A.* **43**, 122–7.

Tirgari, S., and Moseley, B. E. B. (1980). Transformation in *Micrococcus radiodurans*; measurement of various parameters and evidence for multiple, independently segregating genomes per cell. *J. Gen. Microbiol.* **119**, 287–96.

Woese, C. R. (1967). 'The Genetic Code.' (Harper and Row: New York.)

Wouters, J. T. M., Rops, C., and Van Andel, J. G. (1978). R-plasmid persistence in *Escherichia coli* under various environmental conditions. *Proc. Soc. Gen. Microbiol.* **5**, 61.

Ycas, M. (1969). 'The Biological Code.' (North Holland Publ. Co.: Amsterdam.)

Zimmer, E. A., Martin, S. L., Beverley, S. M., Kan, Y. W., and Wilson, A. C. (1980). Rapid duplication and loss of genes coding for α chains of haemoglobin. *Proc. Natl Acad. Sci. U.S.A.* **77**, 2158–62.

Manuscript received 7 July 1982, accepted 18 December 1982