BIOLOGICAL EVOLUTION *

F. HOYLE and N.C. WICKRAMASINGHE
Department of Applied Mathematics and Astronomy, University College, Cardiff, U.K.

Abstract. The known facts relating to the evolution of terrestrial biology points strongly in the direction of an external input of genetic information.

1. Evolution, a Brief History of the Darwinian Theory

It is obvious to the eye that remarkable similarities exist between animals and plants which yet do not normally interbreed with each other, between related species as one says, and this fact must have been known for thousands of years. When the idea first suggested itself to some person that apparently related species really had been related in the sense of being derived from a common ancestral species is not known, although towards the end of the seventeenth century Robert Hooke, who coined the word 'cell' used so widely in modern biology, is said to have been of this opinion. By the latter half of the eighteenth century the evolutionary view had become widespread, particularly in France, to a degree where the systematist Linnaeus accepted it around the year 1770 in order it seems to avoid being castigated by his contemporaries as a fuddy-duddy.

The first widely-discussed evolutionary theory was published in 1809 under the title *Philosophie Zoologique* by J-B de M. Lamarck. The theory rested on the postulate that special characteristics acquired by struggles for existence during the lives of parents tend to be transmitted to their offspring. If this postulate had been true, the theory itself would have been logically viable, but many subsequent experiments have shown Lamarck's axiom to be wrong, unfortunately for him.

British naturalists did not begin in the first third of the nineteenth century with a view as wide as the French had held in the eighteenth century, perhaps because of a distrust in Britain, following the Revolution of 1791–94 and the Napoleonic Wars, of everything French. The initial concern of British naturalists was to understand the factors in nature which control the balance of the varieties of a single species. Since the varieties could be observed actually to exist, they were accepted as given entities, requiring no explanation, thus avoiding the pitfall of Lamarck.

It has been said that the first mention of natural selection was made by William Wells at a meeting of the Royal Society of London as early as the second decade

* Published as Chapter E in 'Proofs that Life is Cosmic', *Mem. Inst. Fund. Studies Sri Lanka*, No. 1, 1982.

Astrophysics and Space Science is the original source of publication of this article. It is recommended that this article is cited as: *Astrophysics and Space Science* **268:** 55–75, 1999.

of the nineteenth century. The phrase 'natural process of selection' was explicitly coined by Patrick Matthew in *Naval Timber and Arboriculture* published in 1831 (Edinburgh). The idea of natural selection is really no more than a tautology:

If among the varieties of a species there is one better able to survive in the natural environment, that particular variety will be one which best survives. The powers of invention required to perceive this truism could not have been very great.

If evolution leading to the divergence of species from a common ancestor was suspected, and if the concept of natural selection was available, why was the theory of evolution of species by natural selection not under discussion already in the 1830's? The answer is that it was, as can be seen from the second of two papers published in 1835 and 1837 by Edward Blyth (*The Magazine of Natural History*). The first of these papers, *The Varieties of Animals*, is a classic. Besides the clarity with which Blyth addressed his main topic the paper contains passages which foreshadow the later work of Gregor Mendel. In his second paper, Blyth considered the theory of evolution of species by natural selection, telling us in passing that the matter had frequently been dealt with by abler pens than his own. The difficulty for Blyth was that, if 'erratic adaptive changes' as he called the modern concept of mutations could arise spontaneously in a species, why were species so sharply defined? Why was the common jay so invariant over the large latitude range from S. Italy to Lapland, when surely it would be advantageous for appreciable variations of the jay to have developed in order to cope better with such large fluctuations in its environment? So quite apart from the unsolved question of the source of the supposed mutations it seemed to Blyth as if the evidence did not support the concept of evolution by natural selection.

The position remained unchanged in this respect for two further decades until the arrival of a new generation of British naturalists, a position analogous to that which occurred almost exactly a century later in respect of the thoery of continental drift. In spite of there being evidence in favour of continental drift, geologists and geophysicists convinced themselves in the 1930's that there were overriding reasons why the theory could not be correct. However, the evidence continued to accumulate to such a degree that by 1960 the situation became inverted. The evidence forced scientific opinion to accept the theory of continental drift, even though nobody understood why continents drifted. So it was with the theory of evolution by natural selection. The evidence forced belief in the theory, even though nobody understood why mutations occur or how the difficulties raised by Edward Blyth might be overcome.

The two crucial papers were both written by Alfred Russel Wallace, with titles that left little doubt of their author's intentions, in 1856 *On the Law which has Regulated New Species*, and in 1858 *On the Tendency of Varieties to Depart Indefinitely from the Original Type*. Unfortunately for Wallace and for scientific history, he chose to send both papers to Charles Darwin, who had himself been skirting the problem for many years in his personal writings, but who had published nothing nor even communicated his views to his closest friends. With Wallace's second

paper available to him however, Darwin then wrote his book *The Origin of Species* published in 1859. The surprise is that, in spite of the extreme clarity of Wallace's writing, Darwin still contrived to state the theory in a laborious confused way and with an erroneous Lamarckian explanation for the origin of mutations, an explanation which Wallace had himself explicitly eschewed (for a detailed discussion see C.D. Darlington, *Darwin's Place in History* (Oxford, 1959)).

If Wallace had published his papers quietly in the *Journal of the Linnaean Society* his views would probably have made as little immediate impact as did the now-classic paper of Gregor Mendel. It was the social prestige enjoyed by Darwin, his friends and supporters, that brought the theory of evolution by natural selection forcibly on the world's attention. As always seems to happen when media publicity becomes involved nobody was then interested in precise statements or in historic fact. Writers copied from each other instead of checking original sources, careers were based on the controversy, and attributions became falsified. So did it come about that the theory became known as Darwin's theory, just as two decades earlier the ice-age theory had become known as Agassiz' theory, after Louis Agassiz who propagandised effectively for that theory but did not invent it.

2. The Neo-Darwinians

The work of Gregor Mendel (published in 1866), was rediscovered early in the present century. The work showed that certain heritable characteristics, colours of peas in Mendel's case, were determined by a discrete unit, which was transmitted from generation to generation in accordance with certain simple mathematical rules. Generalising from the small number of characteristics involved in the early experiments, the view soon gained ground that all the gross characteristics of a plant or animal were determined by small discrete units, genes. At the suggestion of W. Johannsen in 1909, the inferred collection of genes for a set of identical individuals in a species became known as their genotype, and the plant or animal to which the genotype gave rise was called the phenotype.

Advances in microscopy pointed to certain discrete objects in the nuclear region of cells, the chromosomes, as the likely site of the genotype. Since the inferred number of genes was much greater than the number of chromosomes, the genes became thought of (correctly as it eventually turned out) as small structures carried on the chromosomes. Microscopy was not sufficiently refined, however, for individual genes to be distinguished, only the gross forms of the chromosomes. The gross forms for a particular organism became known as its karyotype. Grossly different organisms had readily distinguishable karyotypes, but similar species were often found to have karyotypes that could not be distinguished by the microscopic techniques then available. It was felt, however, that a detailed knowledge of the genes – if it were available – would distinguish between similar species, or even

between varieties of the same species. How far this has turned out to be true will be considered in Section 6.

Experiments of genetic significance in the first half of the century were mainly of two kinds, more complicated examples of the cross-breeding of varieties than those examined by earlier workers, and experiments designed to induce changes in the genotype. Since a gene is a material structure, it was argued, the structure must be changeable by violent means, through irradiation by X-rays for example. It was found possible in some cases to induce changes by such means without destroying viability, although for the great majority of changes viability was weakened in comparison with the original organisms. So genes could be changed, organisms could be altered, mutations could happen it was proved, even though the mutations were deleterious in the overwhelming majority of cases.

Since there could be mutagenic agents in the natural environment, for example the near ultraviolet component of sunlight and ionizing radiation from cosmic rays, mutations could arise in the wild. Besides which, it is surely impossible to keep on copying any object or structure without an occasional error being made. So quite apart from deliberate mutagenic agents there must be a non-zero copying error rate occurring in the genotype from generation to generation. Here at last therefore were the mutations required by the theory of evolution through natural selection. No matter that most of the mutations would be bad, since the bad ones could be removed by natural selection it was argued (erroneously as will be seen in Section 5). Such then was the position of the neo-Darwinians, who imagined themselves in a stronger position than the biologists of the nineteenth century had been, but the reverse was actually the case. The theory in the form proposed by Wallace would admit of mutational changes coming from anywhere, by additions to the genotype of a species from outside itself, for example through the addition of externally incident genes, as well as by changes to already-existing genes. The neo-Darwinians were confined, however, to the already-existing genes, and this had turned out to be an insufficient position, as will be demonstrated here and in Sections 5 and 6. The neo-Darwinians boxed themselves into a closed situation, whereas the theory of Wallace could be either closed or open.

The development of modern microbiology from the work of Oswald Avery in the mid-1940's, through that of Erwin Chargaff to the elucidation of the structure of DNA by Francis Crick and James D. Watson, added precision to the concept of the genotype. The genes were sequences of four kinds of base-pair, A-T and its reverse T-A, G-C and its reverse C-G, a typical gene being about a thousand base pairs long. The base-pairs were subsequently shown to be grouped in triplets with each triplet specifying a particular member of a set of 20 amino acids according to the so-called genetic code, the whole gene being a blue-print for the construction of a particular chain of amino acids, a protein or polypeptide. It is through the active chemical properties of its coded polypeptide that a gene expresses itself and is biologically significant.

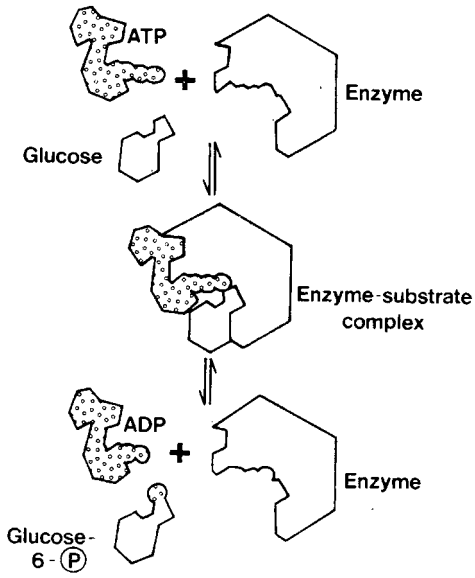

Figure 1. Enzyme action: formation of an enzyme-substrate complex, followed by catalysis.

A mutation to a gene could now be seen to consist in one or more base-pairs being changed to another member or members of the set of four possibilities, A-T, T-A, G-C; C-G, this happening to the initial cell at the germination or conception from which an individual of a species was derived. The chance of such a change occurring due to a copying error was measurable, and was found to be about 10^{-8} per base-pair per generation-i.e. about 10^{-5} for any base-pair to be changed for a whole gene with a thousand base-pairs. This result, was a death knell for neo-Darwinians since it forced evolution according to their views to be a one-step-at-a-time affair, a requirement which both experiment and commonsense showed to be impossible.

Figure 1 is a schematic representation of the mode of operation of an enzyme. An enzyme is a polypeptide which coils into an approximately spherical shape but with a highly specific site at its surface, a site shaped to hold the chemical substances in the reaction which it catalyses, chemical substances existing in many cases outside the biological system itself, chemical substances which do not evolve with the system. This fitting to the shapes of externally-defined substances is a constraint an enzyme must meet in order that it should fulfil its biological function. Exactly how many of the hundred (or several hundred) amino acids in the polypeptide chain of an enzyme must be explicitly defined in order that this shape criterion be satisfied is a matter for debate, but the number cannot be trivially small. If it were so, there would surely be far more variability of structure in the enzymes found catalysing the same chemical reaction in bacteria, humans, and in a potato. The number of amino acids in an enzymic polypeptide chain that

cannot be changed without destroying the function of an enzyme is probably at least a half and may in some cases be considerably more than a half. This demands that, hundreds of base-pairs be appropriately placed in the gene which codes for the enzyme. If one is given an initial situation in which these requisite base-pairs are already correctly placed, well and good, but if the requisite base-pairs are not correctly placed initially, it is essentially impossible that copying errors will ever lead to a functioning enzyme. The difficulty is that all the key base-pairs have to come right simultaneously, not one-at-a-time, because there is nothing to hold individual base-pairs right until the whole lot are right. Every $\sim 10^8$ generations the key base-pairs are randomly shuffled, with the consequence that as some come right others go wrong. The chance of n requisite base-pairs happening to come right at each random shuffling is 4^{-n}, so that with $\sim 10^8$ generations required for a shuffling the number of generations needed for a mutational miracle leading to a functioning enzyme to occur is $\sim 10^8.4^n$, which for n of the order of a hundred is a lot of generations. But not too many for the neo-Darwinians, who know their theory to be right by some kind of revelation, and who therefore are not embarrassed to offer the most unlikely proposals in its defence.

3. Punctuated Equilibria or Punctuated Geology?

If it were possible to circumvent the criticism of neo-Darwinism given at the end of Section 2, arriving at the complex structures of genes several hundreds of base-pairs long by mutations that obtained correct pairs one-at-a-time, with natural selection somehow holding each pair fixed as it came right, evolution would necessarily have to proceed in a very large number of tiny steps, hundreds of steps for each of tens of thousands of distinct genes. There-would be two ways to support this point of view. If both worked out well, one would be obliged to respect the neo-Darwinian position, but both ways turn out badly, as the criticism given at the end of Section 2 warns that they inevitably will. One way would be to demonstrate the mathematical validity of a small-step genetical theory (discussed in Section 5) and the other would be to obtain direct evidence from the paleontological record showing that markedly separated stages in an evolutionary chain are linked by many intermediate small steps. So far from this being found, new species arise abruptly in the paleontological record, forcing the neo-Darwinian theory again onto the defensive in exactly the place where it might hope to be strongest if it were true.

Defensively, it has been pointed out (for example, recently by T.H. van Andel: 1981, *Nature* **294**, 397) that present-day sedimentation rates, if maintained throughout geological history, would have resulted in greater depths of sediments than are in fact found from the various geological periods, implying it is argued either much erosion of sediments, in which case the fossil evidence has been largely destroyed, or it might have been that there was a cessation of sedimentation over much of geological history, in which case the fossil record would have been estab-

lished only sporadically. Evolution in small steps could then be made to appear as a sequence of jumps, simply by the discrete manner in which the evolution happens to be recorded in the presently available fossil record.

All this might be possible as a defensive manoeuvre, but the argument lacks the force of proof. When a curve is drawn through a number of points, the points themselves need occupy only a small fraction of the total range of the abscissa – what matters for constructing a curve is that there be enough points and that they be suitably distributed with respect to the form of the curve itself. Moreover, sediments are available from many geographical areas, and gaps in one place can be filled by available sediments in another place, unless erosion or a lack of sedimentation invariably conspired to be contemporaneous over all areas. For small evolutionary changes such a complementary association of different areas might be considered difficult to achieve but if we are looking for big changes, as from reptiles to mammals for example, a geological resource of this kind should be possible. One could see the defensive argument working in particular cases, but it is implausible to require it to work in every case, as it would need to do to explain the general abruptness of emergence of new species.

If, on the other hand, evolution really does proceed in sudden steps which separate extended time intervals of near-constancy, punctuated equilibria as such an evolutionary process has been called, one would expect to find examples of abrupt changes within continuous ranges of sediments. The question of whether sediments were really laid down continuously or discretely in the manner discussed above, is a matter for the judgements of professional geologists and palaeontologists. If we have understood their findings correctly, punctuated equilibria exist (for example, P.G. Williamson: 1981, *Nature* **293**, 437).

Although neo-Darwinians appear to have convinced themselves that they can explain such findings, we are at a loss to understand their point of view. One might attempt to conceive of many small mutations being accumulated during a time interval of near-constancy of a species, of the mutations establishing a potential for sudden change in a species, like the slow winding of a catapult and of the catapult eventually being suddenly released. But many small mutations established without regard for selective control would mostly be bad, and if there was indeed selective control we should simply be back again with the previous state of affairs, slow evolution in small steps, not punctuated equilibria. Large advantageous mutations could explain the findings, but large advantageous mutations requiring many base-pair changes in the DNA structure of a gene or genes are exceedingly improbable for the reasons discussed at the end of Section 2. Large advantageous mutations requiring only a few base-pair changes might be postulated, but this would be to suppose that genes hover on the edge of marked advantage for species without natural selection having established them in such a critical position. In effect, a *deus ex machina* would be implied. In effect, the theory would have become open in the sense of Section 2, not closed as it is supposed to be in the neo-Darwinian

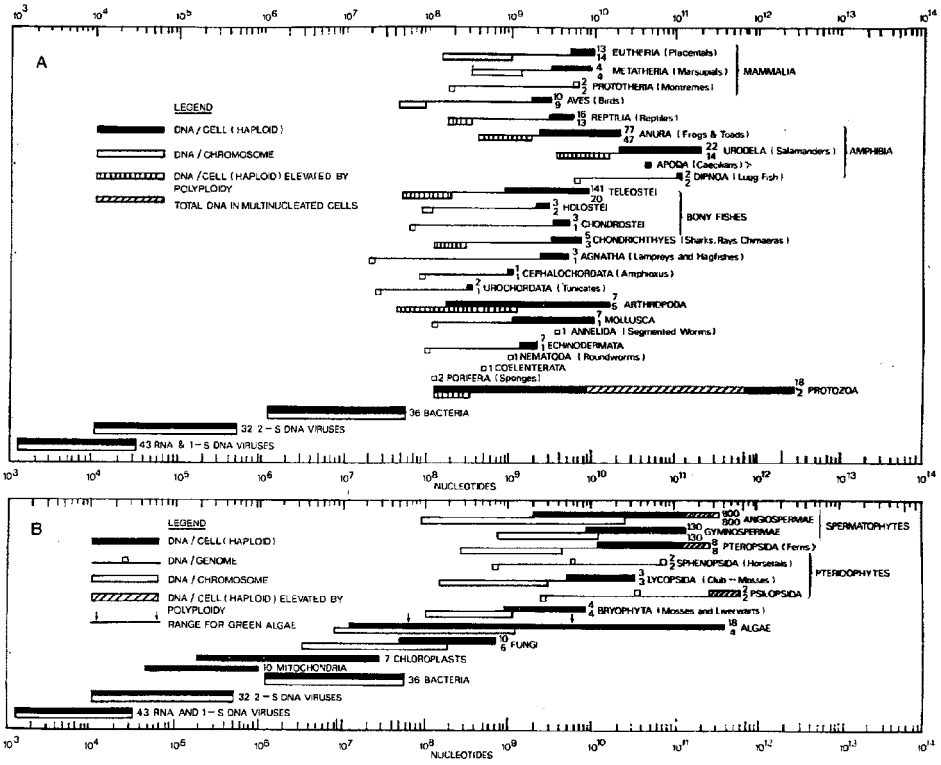

Figure 2. DNA content per cell and per chromosome of various organisms.

theory. The position then comes close to our own point of view, to be explained in Section 4.

Could abrupt changes to a species be caused by sudden geological changes one might ask? Only to the extent that changes in the physical environment produced selection with respect to the already-existing varieties of species. We should then be back with Patrick Matthew in 1831 and Edward Blyth in 1835 (Section l). Geological changes could release genetic potential in the sense explained in Section 4, but geology cannot create genetic potential.

4. Evolution by Gene-Addition

The concept of higher and lower animals, higher and lower plants, is widespread throughout classical biology, and it can be given objective definition in terms of greater or lesser degrees of complexity in the organisation and function of living forms. It is safe to say that if the biologists of the first half of the present century had been asked to guess the relative quantities of genetic material present in various forms general opinion would have favoured a strong positive correlation between

quantity and complexity of function, the higher the plant or animal the greater the amount of genetic material. Figure 2 shows the results of actual measurements, the one part of the figure for animals, the other for plants, with the various taxa ordered generally with respect to complexity of function (A.H. Sparrow, H.J. Price and A.G. Underbrink, in: 1972, *Brookhaven Symp. Biol.* **23**, 451). Except that procaryotes do have significantly fewer base-pairs than eucaryotes, and viruses have still less than procaryotes, the expectation is not borne out. The lungfish easily outclasses the human in the number of its base-pairs. Who would have guessed that the amoeba *chaos chaos* would have had five hundred times more genetic material than the primates?

It might seem odd that the ideas on evolution held by neo-Darwinians have managed to survive Figure 2. One might have expected this remarkable new data to have sparked at least one or two revolutionary ideas. The reason for this congealed state of affairs is simply that the usual evolutionary theory explains little or nothing anyway, so that a further mysterious set of facts scarcely makes an already unsatisfactory theory much worse. It is only good theories that can be upset by new facts. A dead horse can take any amount of beating.

Evidence that microorganisms are continuously incident from space has been discussed by us elsewhere, it being argued that such microorganisms are most readily detected through a component which is pathogenic to terrestrial organisms. Viruses and viroids were considered as well as bacteria, microfungi and protozoa. Some commentators (not professional virologists, at least not to our faces) have claimed that pathogenic viruses cannot be incident from space, for an imagined reason which they believe overrides the many facts which prove otherwise. The argument seems on minimal thought to have the attractive quality of a one-line disproof. Viruses are specific to the cells they attack it is said, as if to claim that human viruses are specific to human cells. While a minority of human viruses might be said to be specific to the cells of primates, most human viruses can actually be replicated in tissue cell cultures taken from a wide spectrum of animals, some indeed outside the mammals entirely. The proper statement therefore is that viruses are generally specific to the cells they attack to within about 150 million years of evolutionary history. Actual diseases tend to be specific to particular species it is true, but this is not the same question, which appears to be where confusion has arisen in the minds of some critics. The ability of a virus to produce a clinical attack of disease in a multicellular plant or animals involves the special physical structure* and the particular immunity system of the creature under attack, and

* The herpes virus can attack brain cells. Fortunately, this does not happen normally because the virus is not permitted physical access to the brain, otherwise its effects would be widely lethal.

possibly other factors also,† all of which are irrelevant to whether the virus can attack individual cells.

If we had knowledge that evolution was an entirely terrestrial affair then of course it would be hard to see how viruses from outside the Earth could interact in an intimate way with terrestrially-evolved cells, but we have no such knowledge, and in the absence of knowledge all one can say is that viruses and evolution must go together. If viruses are incident from space then evolution must also be driven from space. How can this happen? Viruses do not always attack the cells they enter. Instead of taking over the genetic apparatus of the cell in order to replicate themselves, a viral particle may add itself placidly to one or other of the chromosomes. If this should happen for the sex cells of a species, mating between similarly infected individuals leads to a new genotype in their offspring, since the genes derived from the virus are copied together with the other genes whenever there is cell division during the growth of the offspring. Viroids, consisting of naked DNA and perhaps representing only a single gene, penetrate easily into cells, and their augmentation of the genotype may well be still more important than the addition of viruses.

Genes newly obtained in this way may have no evolutionary significance for the plant or animal which acquires them, and for the majority of new genes this would quite likely be so, because each life-form will tend to pick-up a random sample of whatever happens to be incident upon it and in the main a gene acquired at random will probably find no useful genetic niche. It will simply replicate with the cells of the life-form in question without yielding a protein of relevance to the environmental adaptation of the species; indeed, if the gene remains unaddressed in the operation of the cell, it will not yield any protein at all. It will remain 'unexpressed' as one says. So we deduce that many of the genes present in the DNA of every plant and animal will be redundant, a deduction that is overwhelmingly true. Some 95% of the human DNA is redundant. Even higher percentages are redundant in lower animals, which goes some way towards an understanding of how it comes about that a lowly creature may nevertheless have an enormous amount of DNA (Figure 2).

A gene that happens to be useful to the adaptation of one life-form may be useless to another. Incidence from space knows nothing of such a difference, however, the gene being as likely to be added to the one form as the other. So genes that become functional in some species may exist only as nonsense genes in other species. This again is true. Genes that are useful to some species are found as redundant genes in other species. Suppose a new gene or genes to become added to the genotype (genome) of a number of members of some species. Suppose also that one

† It seems possible that attacks of disease are in some cases triggered by a space-borne viroid rather than by the fully-fledged virus. The fully-fledged virus is the output from diseased cells, and it is conceivable that the output from cells contains genes derived from the cell itself. The output would then be more specific to the cell than was the original trigger. There are indications that the special peculiarities of influenza may be due to this kind of process (in *Space Travellers*, 1981, University College, Cardiff Press, page 171).

or more of the genes could yield a protein or proteins that would be helpful to the adaptation of the species. The cells of those members of the species possessing the favourable new genes operate, however, in accordance with the previously existing genes, and since the previous mode of operation did not take account of the new genes, a problem remains as to how the new genes are to be switched into operation so as to become helpful to the species. This question is discussed in Section 6. Here we simply note that, because there is no immediate process for taking advantage of potentially-favourable new genes, such genes tend to accumulate unexpressed. As potentially-favourable genes pile up more and more, a species acquires a growing potential for large advantageous change, it acquires the potential for a major evolutionary leap, thereby punctuating its otherwise continuing state of little change – its 'equilibrium' (Section 3). This is why new species appear abruptly, a concept that will be developed further in Section 6.

5. Genetics In Open and Closed Systems

According to our point of view essentially all genetic information is of cosmic origin. The information does not have to be found by trial and error here on the Earth, so that mutations in the sense of the base-pair shufflings discussed in Section 2 do not have the positive relevance for us that they have in the neo-Darwinian theory. Indeed, just the reverse. Base-pair shufflings are disadvantageous because they tend to destroy cosmic genetic information rather than to improve it, and this is especially so during the interim period before advantageous new genes are switched into the 'program' of a species (Section 4), before they become protected from serious deterioration by natural selection. In neo-Darwinism on the other hand, systems are closed, they start with no information and seek somehow to find it, whereas open systems start with high-grade genetic information which it is important for them not to lose.

For this latter point of view the base-pair copying-error rate should be as low as possible, while for the neo-Darwinians it needs to be high if the requisite sophisticated information is ever to be found, just as the monkeys with their typewriters need to work exceedingly fast if they are to arrive within even a cosmic time-scale at the plays of Shakespeare. The copying-error rate is in fact very low, DNA is very stable, clearly supporting the position discussed in Section 4, not that of the neo-Darwinians.

Since many people think neo-Darwinism to be established beyond doubt, and the questioning of it an act of sacrilege, it is worth leading that theory to the knacker's yard yet again, which will be done in the present section. We shall now show that even within its own postulates neo-Darwinism is self-contradictory. At the end of Section 2 the neo-Darwinian theory was shown to require each important base-pair of every gene (initially not correct) to be held by natural selection when it eventually becomes miscopied to the correct form. The 'discovery' of genes has

to be a one-step-at-a-time process, otherwise there is no possibility worth speaking about of all the many base-pairs coming to their required forms simultaneously. If neo-Darwinism is to be consistent with the detailed structures of genes it is therefore essential that evolution proceeds in very many small steps.

This need to proceed in small steps was already guessed by mathematical geneticists in the first quarter of the present century (e.g. R.A. Fisher, *The Genetical Theory of Natural Selection*, Oxford, 1930). Looking back at this old work it is surprising to find advantageous results for the neo-Darwinian theory being claimed, when even quite easy mathematics shows otherwise, especially as the claimed results were an affront to commonsense. When a mutation is small, its effect on the performance of an individual is so marginal that it scarcely affects the number of offspring born to the individual. Is natural selection really so powerful that in such marginal situations it can stamp-out the flood of slightly negative mutations while preserving the trickle of slightly positive ones? Commonsense says no, and commonsense is correct, as we shall shortly demonstrate.

The remedy of R.A. Fisher was to postulate that small negative mutations are not more frequent than small positive ones, but this supposition also defies commonsense, because it is a matter of experience that complex organisations are much more likely to develop faults than they are to find improvements, a view well-supported by modern microbiology. If the identities of only a hundred base-pairs per gene are important for an animal with 100 000 genes, there are ten million ways at each copying of going wrong. With an error probability of $\sim 10^{-8}$ per copying per base-pair, the chance Q of a significant deleterious mutation occurring per generation per individual is $Q \cong 10^{-1}$. For a breeding group with N members, the number of deleterious mutations injected into each generation is $2QN$, which for a typical breeding group, say $N = 10\,000$, gives two thousand deleterious mutations per generation, quite a burden to be carried every few years. The number of advantageous mutations must surely be much less than this.

An example will make the situation clearer. Suppose a printer sets up a page of 400 words with a dozen spelling mistakes among them. A single letter somewhere on the page is changed at random, thereby introducing a small 'mutation'. The chance that such a mutation will make the spelling worse, giving thirteen mistakes, is evidently overwhelmingly greater than that the mutation will just happen to correct one of the initial dozen errors. Except that genetically there are only four letters for a base-pair (A-T, T-A, G-C, C-G) instead of the twenty-six letters of the English alphabet, the cases are not unfairly compared, especially as the greater number of letters in the literary case is more than offset by the far greater number of genetic 'words', 100 000 genes, any one of which can go wrong.

Since we have analysed the mathematical problem elsewhere (*Why Neo-Darwinism Doesn't Work*, University College Cardiff Press, 1982) it will be sufficient to quote the main results here. In the case of an individual with an advantageous dominant mutation present on either set of chromosomes write $1 + x$ for the ratio of the average number of offspring produced to the average number of offspring

for others without the mutation. Then the fraction of such mutations which natural selection spreads through the entire species is about $2x$. Thus for $x = 0.001$, a fairly considerable advantage of 0.1 percent, the chance of a mutation spreading through the species is no more than 1 in 500. It therefore needs some five hundred fairly considerable mutations, each of them likely to be a rare event, before just one is retained by the species. Hence for mutations with x small, natural selection adds up very little that is good.

The trouble lies in stochastics, an effect that was inadequately considered by the early mathematical geneticists. For a heterozygote with respect to a gene of small x there is already nearly a 25 percent chance that the mutation in question will be lost in the first generation, simply from the random way in which the heterozygote allots one or other of its duplicate set of genes to each of its offspring. In the second generation there is again a chance of about 3/16 that the mutation is lost. Stochastics consists in adding up and allowing for these extinction possibilities, which greatly dominate the effects of natural selection when small mutations first arise.

For the same reason natural selection by no means removes all that is bad, as classical biologists supposed. For deleterious mutations it is the recessive case that matters most. If for simplicity of argument one takes all recessive deleterious mutations to be equally bad* an elegant result can be proved. Subject to the disadvantage factor x being sufficiently small, the rate at which deleterious mutations spread through a whole species is equal to the rate Q of the mutations per individual,† just the same result as was proved about a decade ago for neutral mutations (M. Kimura and T. Ohta: 1969, *Genetics* **61**, 763).

If natural selection fails for moderate mutations to add-up more than a small fraction of what is good, and if natural selection fails to exclude a damaging fraction of the much more frequent disadvantageous mutations, how can species ever become better adapted to their environment? For small-step mutations they cannot, which is why neo-Darwinism fails genetically, why positively-evolving systems must be in receipt of genetic information from outside themselves, as was discussed in Section 4. The best a closed system can do is to minimise in *disadaption* to the environment, a topic that is discussed in Section 7.

Natural selection works excellently for open systems, since with high-grade genetic information coming from outside a system, advantageous changes have large values of x, with $2x$ of order unity, so that if such a change occurs for only

* For a deleterious mutation write $1 - x$ for the ratio of the average number of offspring produced by an individual with the mutation on both chromosome sets to the average number of offspring produced by individuals without the mutation. The disadvantage factor x (> 0) is taken the same for all deleterious mutations.

† The condition on x is that the product of x and the number N of individuals which constitute a breeding group be not greater than ~ 1. This leads to a *disadaptation* factor $\exp(-QG/N)$ arising in G generations. For values of N appropriate to mammalian species this disadaptation factor becomes an embarrassment to neo-Darwinian theory as G increases above a million generations.

one or two individuals of a species, natural selection operates to fix the change throughout the entire species. Such major advantageous steps have to occur with a sufficient frequency to more than offset the numerous small deleterious mutations which still produce disadaptation at the rate discussed above. In effect, the situation is a race between uphill jumps produced by externally incident genetic information and the downhill slide of the already-existing genes, which natural selection can only moderate but not remove entirely. This produces a highly fluid situation, with species either advancing rapidly or sliding backward towards extinction as is observed to have happened for the higher plants and animals.

When one looks back at the mathematical geneticists of the first half of the present century, it is clear they approached their work in the complete conviction that the neo-Darwinian theory was correct. As the majority of them saw it, their duty was to explain why a theory known to be correct was indeed correct, a mode of argument not unlike a chemist attempting to work backwards through an irreversible reaction, or like an inept student in an examination trying to work backwards from the answer to a problem to its mode of solution. This wrong-headed approach led somewhat naturally to a prostitution of logic which was mercifully concealed from the public in a haze of mathematical symbols. The irony is that the correct answer was easy to find if only the mathematical geneticists had troubled to look for it in the right direction.

6. Favourable Mutations in Open Systems

Open systems do not have to find genetic information *de novo*, because they are in receipt of genes from outside themselves. However, newly-acquired genes must lie fallow for a while, since the mode of operation of the cells of the species in question cannot 'know' in advance of their arrival. The sequence of events whereby genes are used may usefully be described as the cell program. What needs to be done therefore to promote evolution in an open system is to alter the cell program to take into its operation new genes which it did not use before. The problem to be considered here is the logic of this situation.

A cell program may be thought of as analogous to a computer program. With computers, the program is something different from data and from the closed subroutines which constitute the backing storage. Computers can be operated on many different programs using the same physical hardware and the same backing facilities – examples of the latter are routines for taking logarithms and integrating differential equations. Something of the same kind almost surely exists in biological systems. Genes for the production of enzymes, haemoglobin, the cytochromes, are examples of subroutines that run across all of biology. It is even the case that genes capable of producing some of these standard products, haemoglobin for instance, exist in life-forms which normally make no use of them, just as stand-

ard computer languages like FORTRAN or BASIC contain more facilities than are used in any particular individual program.

In days long ago, before sophisticated computer languages were available, when it was necessary to remain closer to the electronic nature of the computer itself, one was perhaps more keenly aware of the distinction between the logical instructions which constitute a program and the numbers or words on which the program operates, even though both were stored in the computer in exactly the same way, as sequences of bits. Although numbers and logical instructions were similar electronically, you could not use numbers for logical purposes or process your logical instructions arithmetically (a few very slick fellows tried and were sometimes successful, but the tricks of this particular trade were too subtle to have survived into current practice). As well as numbers constituting data and logical instructions making up the program, something else was needed, a starting point and an end point, birth and death.

Do biological systems operate in a similar way? Are the logical instructions constituting the cell program stored as genes, but used quite differently from the genes which code for working polypeptides such as the enzymes? Is everything stored as base-pairs in the DNA, just as everything in a computer is stored in sequences of electronic bits? It is tempting to suppose so, but there are indications that it may not be so. The DNA of a chimpanzee is extremely similar to that of a human. Therefore the scope for producing working polypeptides is essentially the same in the chimpanzee as it is in ourselves. Thus the chimpanzee and the human look like two different programs operating on the same physical hardware, on the same backing storage as one might say. If the different programs were on the DNA we might expect to see less close similarity, less homology, between the base-pair sequencing of the two species, unless program storage occupies very little of the DNA, unless the logical ordering which makes us specifically human and a chimpanzee specifically chimp is in each case rather trite and short. Perhaps the logic of being human is rather trivial, but one prefers not to think so.

A less subjective objection is that DNA seems far too stable to be the source of the cell program. If the cell program were so contained, body cells could be replicated a very large number of times without the program being much impaired, permitting animals to have exceedingly long lives, whereas the evidence shows that the program becomes seriously muddled after only a handful of replications. Recognizing this discrepancy some biologists have argued that senesence is itself a deliberate part of the program, deliberate in the sense that natural selection has prevented us from living long by explicitly stopping the coding of essential working polypeptides. This opinion is to be doubted, however, because wild animals commonly die violent deaths before their time is run, so there is no cause in nature for natural selection to prevent lives from being too long. Yet all animals do show senesence, if artificially protected against violent death most of them even more markedly than we do, indicating that senesence is not artificially contrived. The implication is that storage of the cell program must be ephemeral. It is preserved

with reasonable fidelity in gametes, but soon runs down and becomes forgotten, leading to grey hair and the like, as soon as the somatic cells are required to replicate more than about a hundred times.

If a person tells you that the telephone number of a mutual acquaintance is 752146 and you immediately commit the number to paper you have it in stable storage, like base-pairs on DNA. But if you seek to remember the number aurally in your head, it will be gone at the first distraction, a knock on the door or a pan of milk boiling over on the stove. This seems to be the way of it with our cell program. Once we have lost it, the thing never comes back, although if it really is retained in our gametes somebody may succeed someday in copying it back into our somatic cells, with interesting sociological consequences.

In spite of these difficulties, suppose for a moment that those who think the cell program is written on the DNA are correct. How would the program actually do something? Not by merely remaining on the DNA, because DNA by itself is inert. The program would need to be translated into polypeptides and it would be the polypeptides that really did something. So why not let the program be polypeptides in the first place? Or if not the whole program, suppose an essential part of it is in polypeptide form, without there being any reference genes on the DNA form from which the initial polypeptides can be recopied if they become lost. One might conceive for instance that the initial polypeptides comprise a catalogue of what in computer terminology would be referred to as calling sequences, which is to say some means of determining so-called introns for finding important genes on the DNA. Senescence looks very much like the progressive garbling of the entries in such a catalogue, so that we end in old-age by not being able to find more than a small fraction of the genes necessary for vigorous life. All this is relevant to the evolutionary problem set out at the beginning, since the less rigidly fixed the cell program the more readily one can conceive of it being changed. The change needed for an evolutionary step must involve some means of addressing new genes added to the DNA, the genes which supply the potential for an evolutionary leap. This means actually doing something, not just adding DNA blueprints for doing something at some stage in the future. Actually doing something means polypeptides, and doing something new means new polypeptides, which implies a working addendum to the old cell program. Where one now asks is such a working addendum to come from? Only it seems from a virus.

When a virus invades a cell it mostly happens that the virus multiplies itself at the expense of the invaded cell, which it does by stopping the old cell program and inserting its own program, both necessary but not sufficient properties for what we are seeking. The several viral particles thus produced then emerge from their host in search of still more cells to invade, and so, on apparently *ad infinitum*. This behaviour is usually viewed as a permissible oddity of biology, permissible because the virus survives, and survival is all according to the opinions of neo-Darwinians. Yet mere survival leaves the virus as a disconnected organism without logical relationship to anything else. Once one admits, however, that logical rela-

tionship is at least as valid a concept as survival, indeed that survival is impossible for any organism without logical relationship, the situation becomes different. The virus becomes a program insertion with the essential capability of forcing cells to take notice. Many such program insertions are needed to cope with many stages of evolution for many creatures, both on the Earth and elsewhere. Hence many viruses are needed and even if the entry of a particular virus into cells is restricted to situations in which the cell program and the viral program match together in a general way, it will not usually happen that a virus on entering a cell has precisely the appropriate program insertion to suit the life-form in question exactly at its current stage of evolution. There will have to be many trials before precisely the current program insertion is found. So what is the virus to do in the majority of cases where the situation is not quite right? Give up the ghost and expire? If it did so, what about the other creatures somewhere in the Universe that may be in dire need of its particular evolutionary contribution?

Viruses seek cells, not *vice versa*. Speaking anthropomorphically, they have the job of driving evolution. They cannot give up the ghost and expire, otherwise nothing would happen, the situation would be as dead as mutton. So they augment themselves by increasing their number and then they press on, forever seeking to find the cells where they are needed. As soon as one looks for logical design, the situation immediately makes sense. Besides which, the infective ability of viruses also plays a crucial logical role. For species with a sexual mode of propagation there is a big question mark as to how an evolutionary leap could ever be possible, because the same leap must occur in at least one male and one female, otherwise the male and female gametes will not match properly, and there will be reproductive trouble in the second generation, if not indeed immediately. Since the probability of an evolutionary leap occurring is small, requiring first a building of a potential for the leap and then finding the correct addendum for the cell program, it would be a poor result if the individual for whom all this happened were then to be sterile. Yet if we need the same improbable sequence for the opposite sex also, the small probability is squared, and moreover the changed male living in London would then have the problem of finding the changed female living in New York, making such an uncorrelated situation quite hopeless. The solution to this last problem is infectivity. The same changes, all being virus induced, can be infective between individuals in close contact at the same geographical location, and in this case the small probability is not squared, and moreover the similarly affected individuals are automatically together and so cannot avoid finding each other. The logic of an evolutionary leap demands infectivity. Infectivity also explains why after an evolutionary leap the previous line does not persist, since with an evolutionary improvement sweeping through a species like a disease, a negative disease as one might say, the previous line is overwhelmed by the superior adaptation to the environment of the drastically changed creatures. Only in this dramatic way can evolution counter the degenerative effect of the small but steadily-occurring mis-

copying of genes, the downward drag that was mentioned above and is considered in more detail in Section 7.

The above discussion also makes it clear why viruses have to be generally specific to the cells they invade.

7. The Survival and the Extinction of Closed Systems

Here we accept the conclusion of Section 5, that natural selection is not able to fix in a species more than a small fraction of the infrequent advantageous mutations which arise through the shufflings of base-pairs on the DNA, and hence that internal processes cannot improve the adaptation of a species sufficiently to be significant. Only by importing genetic information from without can adaptation be improved in an important degree, and this we consider in the present section to be absent.

Although natural selection (together with stochastic processes) remove a large fraction of the numerous deleterious mutations, sufficient of them necessarily remain to degrade the adaptation of a species quite seriously. The most troublesome deleterious mutations are the recessives, which arise because initially useful polypeptides change gradually into nonsense proteins as random shufflings of the base-pairs alter their amino-acid sequences to less useful arrangements. A deleterious recessive on the same gene of both chromosome sets of a diploid cell has a disadvantage expressed by the average of the ratio of the number of offspring produced by such individuals to the number of offspring produced by individuals without the mutation (but who are otherwise similar). Write this disadvantage factor as $1 - x$, so that x is a positive number between zero and unity.

The extent to which the combination of stochastic effects (Section 5) and natural selection permits such a mutation to penetrate a species depends on $4xN$, where N is the number of diploid individuals making up the breeding group, taken to mate within itself at random. Write Q_1 for the average rate of occurrence of deleterious recessive mutations with $4xN > 1$ per individual per generation, and Q_2 for the average rate per individual per generation for mutations with $4xN \leq 1$. Starting from a pure line state of affairs in which all chromosome sets are identical throughout a species, the situation which transpires is the following. Over a very long time-scale the mutation rate Q_2 degrades the quality of the pure line while on a shorter time-scale the rate Q_1 degrades the species relative to the slowly changing pure line by the factor $\exp - Q_1$. We discuss these two distinct effects separately, after noting that both Q_1 and Q_2 are generally of order unity. Taking mammals as an example, each diploid has $\sim 6.10^9$ base-pairs, so that with a copying error rate of $\sim 10^{-8}$ per base-pair per generation there are 60 miscopyings per individual per generation. However, only those miscopyings of pairs belonging to expressed genes are relevant in the present connection, say 5 per cent of the total, giving 3 relevant miscopyings per generation, i.e. $Q_1 + Q_2 \cong 3$, taking most of the miscopyings to be deleterious and most of them to be of a recessive nature. In the absence

of information as to how these ~ 3 miscopyings should be divided between Q_1 and Q_2 we assign them equally, $Q_1 \cong 1.5$, $Q_2 = 1.5$, both per individual per generation.

Suppose for the moment that all mutations contributing to Q_1 have the same value of x. Stochastic effects give a chance $\sim \sqrt{(x/N)}$ of each such mutation spreading into $\sim \sqrt{(N/x)}$ members of the species. Thereafter natural selection operates to prevent further spreading. Indeed natural selection works to reduce the number of distinct mutations which become spread by stochastics, while the injection of new mutations works to increase the spreading of distinct mutations. An equilibrium between these opposing effects becomes established in $\sim \sqrt{(N/x)}$ generations, an equilibrium in which $\sim 2Q_1N$ distinct mutations are each spread at random in $\sim \sqrt{(N/x)}$ members of the species, giving an average of $\sim 2Q_1\sqrt{(N/x)}$ mutations per diploid.

Because of the randomness with which the distinct kinds of mutations are distributed, the mutations on the two chromosome sets of a diploid cell are uncorrelated, so that the same gene is affected on both chromosome sets only by chance, the chance of a coincidence being $\sim 0.5\sqrt{(N/x)}$ for each of the $\sim Q_1\sqrt{(N/x)}$ kinds of mutation that on the average are present on every chromosome set. For a diploid there are thus $Q_1/2x$ deleterious recessive coincidences. Each of the N individuals forming the breeding group therefore encounters a reproductive penalty relative to the initial pure line expressed by the factor $(1-x)^{Q_1/2x}$ which for x appreciably less than unity in general is $\exp -Q_1/2$. Hence at a typical mating of a male and female, each with the degradation $\exp -Q_1/2$, the combined penalty is $\exp -Q_1$, as already stated above.

The value of x does not affect the penalty, only the rate Q_1 is relevant at which mutations arise per individual per generation. This remarkable result permits the assumption that all mutations contributing to Q_1 have the same x to be dropped. (If one had mutations with either x' or x'', $x' > x''$, the greater deleterious effect of an x' mutation would be compensated by the greater number of x'' mutations that penetrated the species.) Thus the penalty per mating pair relative to the original pure line is $\exp -Q_1$ with Q_1 now interpreted as the total rate of occurrence of mutations with $4xN > 1$ per individual per generation, a result which leads to the deduction that no closed species can have appreciably more than 10^8 expressed base-pairs on its DNA. Otherwise with a miscopying rate of $\sim 10^{-8}$ per base-pair per generation we should have Q_1 much larger than unity and the penalty $\exp -Q_1$ would be exceedingly severe, likely enough leading to an extinction of the species.*

Although all base-pairs are subject to much the same miscopying rate, only a fraction of the mutations which occur ever penetrate a species significantly. The majority of mutations are removed by stochastic effects. There is no means of determining which mutations happen to penetrate and which are eliminated almost

* This limitation on the number of expressed base-pairs assumes no gene duplication. The number could be increased by multiple polyploidy, for example.

immediately, the issue is a matter of chance. Thus if we imagine an initially pure line separated into two breeding groups, after a suitable number of generations have elapsed both groups will be afflicted by the same degradation factor $\exp -Q_1$. However, the recessive mutations causing this same degradation factor will mostly be different from one group to another.

Suppose in such a situation that the two groups are artificially mated together, as for instance two varieties of wheat might be crossed by a plant breeder. The factor $\exp -Q_1$, afflicting both groups separately, evidently disappears almost entirely from the first generation of hybrids, because genes affected by recessive mutations on the chromosome set derived from the one group do not in general match the mutations on the chromosome set from the other group. In other words, the mistakes of the one are shielded by the other, and with the degradation factor $\exp -Q_1$ thus disappearing from the hybrids the vitality of the original pure line is restored. However, coincidences of recessive gene mutations begin to appear again already in the second mixed generation, and random matings with chromosome crossovers occurring degrades the situation in only a few generations about half-way back to what it was before. This is the phenomenon of hybrid-vigour well-known to plant breeders.

Of the total of NQ_2 mutations with $4xN \leq 1$ that arises in each generation, a fraction $\sim 1/2N$ penetrates a species entirely due to stochastic effects, thereby slowly changing the original pure line that provided the standard relative to which the degeneration factor $\exp -Q_1$ was measured in the above discussion. Hence the standard of reference itself deteriorates relative to the original pure line by a factor

$$(1-\bar{x})^{Q_2/2}(1-\bar{x})^{Q_2/2} \cong \exp(-\bar{x}Q_2)$$

per generation for each mating pair, $\bar{x}$ being the mean of x ($x \leq 1/4N$). This further source of deterioration is cumulative from generation to generation; after G generations it becomes $\exp(-\bar{x}Q_2G)$. Taking $Q_2 \cong 1.5$ as indicated above, there is a decline by $1/e$ in $\frac{2}{3}\bar{x}$ generations, which for $\bar{x}$, say, equal to $1/(6N)$ is $\sim 4N$ generations. A number of interesting conclusions can be drawn from this result.

Under the condition assumed in this section, namely zero input of genetic information from outside itself, a species with N no larger than 10^5 is exposed to an overwhelming threat of extinction. Thus in a geological period of $\sim 10^8$ years with G upwards of 10^7, $\exp(-G/4N) = \exp -25$, surely a disastrous decline. Curiously for closed species with breeding groups no larger than 10^5, it would be better if there were no small mutations, better if all deleterious mutations had $4xN > 1$, because natural selection could then prevent mutations from becoming fixed, and so could prevent the reference standard from deteriorating. The maximum penalty from deleterious recessives would then be $\exp -Q_1$, which is not cumulative from generation to generation. Continuing, however, with $Q_2 \cong 1.5$, for a species to survive over a geological time-scale one at least of the following conditions must be satisfied:

(i) The breeding group N is very large, say $\sim 10^8$ or more.

(ii) The species is open to the receipt of genetic information from outside itself, and this external impulse is sufficient to upgrade the species at least as fast as it is being downgraded by internal mistakes.

For the larger mammals in the wild (i) is not satisfied, so that (ii) is necessary for long-term survival, as well as for the evolutionary development of mammals (Section 6). If plants and invertebrates are considered to be closed systems, then (i) must be satisfied. Any closed species for which N falls appreciably below 10^8 is doomed to extinction on a geological time-scale, and this no doubt is the reason why so many species have in fact become extinct throughout the geological record.