CHAPTER 9

EVOLUTION: A VIRAL THEORY.
Evolution as Natural Genetic Engineering

The picture of the cell that emerges — perhaps rather hazily and tentatively — is that of a composite derived, through evolution, from elements which may have ‘come together’ via distinct evolutionary pathways.¹

THE THEORY

I put forward the theory essentially as initially conceived in 1977,² and later published in the general literature.³,⁴ I do this to show that the theory itself remains essentially unchanged. Updating of the previous chapter (Ch. 9,⁴) will be more related to evidence accumulated in the intervening years, largely under the sections dealing with: ‘What the present theory might explain,’ and ‘Evidence for a viral contribution to the prokaryotic/eukaryotic genomes.’ Those familiar with the general viral evolutionary theory might prefer to go straight to those sections.

I suggest that life has evolved largely by viruses moving within and between cells, individuals, populations, and even species, gradually integrating genetic material into their hosts. By this view, whilst any virus remained in an infective if not pathogenic relationship with its host species, new pieces of genetic information would continually be introduced into each host generation. Moreover, this genetic material would not always simply be of viral origin since occasionally, in ‘viral transduction,’ pieces of previous-host genes can be carried in as well due to imprecise excision of the virus from its previous-host integration site.⁵ As well as integrating into new host-sites, such host gene fragments can also recombine with similar segments of the new host to cause part host-gene or ‘exon’¹ conversion. Even at one time in evolution, repetitive cross-infection in this way could result in great genomic diversity among the species, derived not only from variation in the type of viral and host-gene fragment carried, but from differences in the point at which they were inserted into new host chromosomes. Different points of chromosomal insertion could result in very diverse effects, due to variation in both extent of viral gene expression and degree of influence of the new DNA on the function of adjacent host genes, in the same way as Britten and Davidson⁶ have illustrated the enormous potential for evolutionary diversity that lies in the dispersion of repetitive DNA sequences throughout the eukaryotic genome (see also chapter 10). Particular points of viral insertion might confer especially advantageous characteristics on the host, and this would provide a powerful driving force for the processes of Darwinian natural selection and evolution of the species. At other sites of insertion, advantage might also accrue from suppression of potentially lethal or otherwise undesirable viral effects, for example when the viral sequence was inserted beyond some host-gene terminator sequence or within a transcribed but not translated segment of host DNA, or when it was partially suppressed by an association with histones or other chromosomal proteins.⁶

Note that this view places as much emphasis on host-gene segment transfer and recombination via viruses as it does on viral transfer itself.

¹ See chapter 10 for more detailed discussion of the organisation of higher species ‘genes in pieces,’ as multiple part-genetic exons interspersed within vast non-translated introns.
Horizontal Evolution and Viral Exchange

Even at one point in evolutionary time, the potential of this mechanism to promote genomic change within the various species would be vast. First, a whole range of different viruses could become integrated into the host genome. In higher species, this would most obviously include the DNA viruses and the RNA retroviruses, but even ordinary RNA viruses that normally play little part in the build-up of the host (DNA) genomes of today might possibly have done so in the past, and could perhaps still do so on occasions by being reverse-transcribed into the host genome by the presence of cellular reverse transcriptase.\(^7\)\(^-\)\(^9\)

The term ‘virus’ is used here to include not just complete viruses readily recognizable as such, but the whole range of mobile and other transmissible and transposable genetic elements, including the small and relatively primitive plasmids, ‘insertion elements’, transposons, and retrotransposons, all of which are known to be capable of transmission between, and integration into, the genetic material of a wide variety of biological forms, including larger viruses.\(^10\) These small genetic elements may not be able to transfer independently of their hosts, but as we will see, they can often be carried by viruses to achieve this same end, and once within any new cell their influence on the genome can be just as profound.\(^7\),\(^11\),\(^12\)

By these means, the various host species could be exposed to a wide range of outside genetic ‘experience’ at any one time. Further, many viruses are able to infect a whole range of different host species and interchange freely between them, so adding further to the potential range of genetic variability and interchange. This is because imprecise excision of virus from host genome means that host DNA sequences can occasionally be transferred (transduced) between species normally having no genetic inter-communication at all. In addition, if one virus/host ecosphere system slightly overlapped another, viral co-infection of an individual within the area of overlap might lead to genetic recombination between the two viruses concerned, so extending the range of virus/host genetic experience even further. As we shall see, there is evidence that viruses can recombine in nature along these lines, so this is much more than just a theoretical possibility.

The potential for exchange of genetic material between species in nature by viruses and like elements is therefore very broad, because it would only require the introduction of occasional genetic novelty for this process to have profound consequences in evolutionary terms. Of course, it would have to have a limit, for otherwise either species would be totally eradicated by their detrimental effects or by disease, or at the other end of the spectrum, all life would finish up as one grand conglomerate. However, there are ‘restriction’ endonucleases within the cell that can act to limit uptake of foreign DNA,\(^13\) and other means to control its influence.\(^11\),\(^14\),\(^16\)

Variety at any one point in evolutionary time could also be bought about by variations in the type of viral genetic material inserted into host. We have noted how, in viral transduction, different host sequences can be incorporated into virus and carried with it into a new host at next infection. But there are similar possibilities for partial viral sequence insertion as well. A case in point is where a similar stretch of viral DNA exists in both host (as ‘pro-virus’) and virus, from previous rounds of viral integration. This not only provides a site of homology for the integration of viral, plasmid and other nuclide forms\(^17\) into host via ‘legitimate’ recombination\(^18\) but also sets the stage for recombinational part-exchange of homologous viral segments between the two – just as in host segment transduction and part exchange. There are even opportunities for the insertion of partial viral sequences into host in the absence of any apparent homology between them at all, through so-called ‘illegitimate’ recombination. This occurs quite widely within prokaryotes at least, where it is responsible for the bacterial integration of a whole range of transposons and other mobile genetic ‘insertion elements.’ In the current context, it is of note that some of the higher viruses, for example the eukaryotic transforming viruses, contain subunits of transposon-like character within their integrated molecular substructure,\(^10\) because these might well subserve a similar role in promoting partial viral sequence insertion higher up the evolutionary scale.
By such mechanisms, the breadth of genetic experience available to any species existing in an infective relationship with a number of different viruses at any one time could be very broad indeed. But add to this the variety of genetic sequences that might be inserted at different host genomic positions, with widely differing effects, together with the fact that each modular unit of genetic material so integrated would already have been 'processed' at some previous level by its own evolutionary background of Darwinian development, and the potential for significant evolutionary novelty by this means becomes vast – and certainly far greater than could ever be envisaged under any random DNA point-mutational process. As well as being important to molecular evolution, the modular nature of this process of genetic interchange might also be relevant to the very process of speciation itself.

**Vertical Evolution**

In addition to the scope this mechanism offers for genomic evolution at any one time, it is proposed that evolutionary build-up of the genome has been achieved largely in the same way since primitive times. By this view, mobile genetic elements would be seen as always having been capable of transferring genetic material between cells, with the effect of elevating life forms towards ever-increasing levels of complexity via the germ line. As primitive viruses gradually shared and exchanged genetic ‘experiences’ with their hosts in this way, both would be seen to evolve, with the initially primitive genetic material of the virus-host ecospheres of earliest times being gradually ‘processed’ towards higher and higher levels of biological organization, even to the most sophisticated levels now existing. This process is seen as including the insertion elements and DNA transposons of prokaryotes and, in higher species like the eukaryotes, the DNA viruses, RNA retroviruses and retrotransposons. There is some evidence that in most primitive times of all, the evolutionary history of the gene may have been written as RNA rather than DNA, and if so, the RNA viruses, viroids and other transposable RNA material might well have made a contribution to the past evolutionary build-up of host-species genomes just as important as the DNA transfer elements of today. Of course, as soon as DNA became established as the major genetic code, any contribution from such RNA viral material would have substantially declined. Even so, it could still make an occasional contribution, for example by being reverse-transcribed in the presence of cellular reverse transcriptase. This might not be common, but because reverse transcriptases seem to be relatively non-specific in their substrate requirements, once it did occur it could have huge consequences in evolutionary terms. Even simple RNA prokaryotic viruses might also occasionally contribute to such a process.

The present theory, then, proposes that right from the very earliest evolutionary times, mobile genetic elements have provided the ‘melting pot’ sorting out random DNA point-mutational change, with life gradually evolving to increasing levels of biological sophistication ever since by ‘viral’ transfer mechanisms interchanging genetic material between individuals, populations, and even between species. According to this view, any mobile genetic element introduced into a host species from the most primitive to the most complex biological level would be tested by the natural selection processes appropriate to that level for its ‘usefulness’, and the new ‘virus’ (or virus-host genetic complex) accepted or rejected depending on the outcome of that test. If accepted, the DNA concerned could then be transferred, tested, and processed to some higher level, all the time with increasing variability being available from the type of sequence (virus or virus-host genetic complex) transferred, and the site of its integration into, or recombination with the new germ line host DNA.

With the passage of evolutionary time, an even further potential source of variation could be provided by evolution in the virus itself, sometimes to the extent that each new generation of host species might be exposed to importantly differing molecular viral forms. Moreover, some of that viral evolution might be due not just to random DNA mutational change within the virus, but to genetic...

---

This has potentially important consequences for autoimmune disease as well. (Boyd, GW. 1997. Immunol Cell Biol 75: 503-507)
recombination with the host genome, and in this way both virus and host could co-evolve in a
dynamic, dependent, and relatively orderly manner. Further, because each host generation seems to
be largely naive with respect to its species viral invaders (i.e. have no inherited antigenic memory for
similar viral infections in its forebears), this process of co-evolution could continue for as long as the
virus remained in any sort of infective relationship with its host species, pathogenic or not. Of course,
such a relationship would be unlikely to last indefinitely, because of the many factors tending to
militate against it. These include the occurrence of increasingly homologous sequences from
recurrent viral transduction making the virus look more and more like host, the emergence of host
mechanisms to suppress the virus, and the occurrence of mutations in the viral genome during
long periods of its latency or suppression within the host. But even where such factors resulted in the
complete loss of a capacity for viral transfer between individuals, enough mobility characteristics
might still persist for it, or one of its subunit mobile elements, to move between the chromosomes
within any one particular cell, or even between the cells of an individual. And provided this at least
occasionally included transfer to cells of the host germ line, such a mechanism could continue to
provide a source of important genetic variation long beyond the time overt infectivity had been lost. In
1979, Steele suggested that such germ line transfer might be common during evolution, and
although at the time this raised the ire of Medawar and his colleagues, it would appear to remain
quite feasible. Of course, there would have to be limits on such germ line mobile element transfer,
but this seems highly likely.

An important issue raised by the present discussion is whether viruses themselves might have
been largely built up during evolution by the gradual (symbiotic) accumulation of smaller mobile
genetic elements. The traditional view is that they are derived from their host species, rather than
having had a separate evolutionary history of their own right from primitive times. This view is based
largely on the observation that presently observable viruses lack the ability to survive independent of
their hosts. But this may not always have been the case, in the same way that the late Lynn
Margulis argued, correctly, that present-day mitochondria are derived from what were once free-
living aerobic organisms, now modified to a state of dependent existence by a long history of
symbiosis. In any event, viruses have long been capable of transferring between other living forms,
and this is the most important point as far as their role in genetic exchange and evolution is
concerned. What is at issue here is just when this process of genetic exchange began. The present
perspective is that viruses and other mobile genetic elements have existed from earliest times as
units capable of transferring freely between other living forms, and moreover that they have co-
evolved along with their hosts ever since by a continual process of genetic interchange with them.
According to this view the higher viruses should not now be seen merely as genetic splinters of their
hosts, split off in some recent evolutionary time. Rather, beginning as the simplest mobile genetic
elements interchanging between the most primitive life forms of primordial times, they would have
been gradually built up through continuing genetic interchange to higher and higher (though perhaps
always-dependent) genetic complexity as they co-evolved along with their hosts, even to the highest
branches of the evolutionary tree.

The structure of present-day integrated retroviruses, and even some of the DNA viruses, would
certainly fit well with this view, because we now know that the subunits of at least some of these are
closely similar to the mobile genetic elements and transposons of lower species. Of course, after
such a long history of evolutionary integration, much of the mobility characteristics of integrated
viruses and other genetic elements inserted so long ago might by now be very largely lost. But, even
so, enough could still remain to be brought back under appropriate conditions, such as those which
cause cell disruption or damage; and if not, vestiges of a viral background nature might still be
discernible at the molecular level. This point will be taken up further below.

**Diagonal Evolution**

I suggest that not only have the evolutionary exchange processes outlined been of major
importance in building up genetic material to increasing levels of biological sophistication in the past,
but that this whole process still continues, i.e. that there is 'horizontal' virus/host co-evolution continuing to broaden the whole range of vertical evolution of the past. By this I mean that each virus/host ecosphere now in existence still overlaps sufficiently with others to allow the whole biological chain to be in continuing genetic contact right the way up the present 'evolutionary scale'.

According to this view, the various virus/host ecospheres would not only have provided the basis for processing DNA mutations from the most elementary to the most complex biological levels of the past, but this would remain so. Genetic novelty would be introduced and fixed at higher levels largely by what we might call a process of 'boot-strapping' of increasingly complex and previously processed DNA higher and higher up the current 'evolutionary' scale through increasingly-developed virus/host ecospheres. I suggest that this process provides the basis for present-day evolutionary progress at all levels, with the 'horizontal' evolution of the present mimicking the ways of 'vertical' evolution of the past. In addition to this 'long-loop' process of genetic exchange through overlapping viral/host ecospheres, short cuts might sometimes be taken along the way as well, for example by genetic recombination occurring between two viruses in an area of overlap in such a way as to lead to a radical extension in the range of infectivity of the newly-emerged recombinant viral form. 31, 32

Such a genetic interchange mechanism would be important enough in single-celled organisms, but its potential would be even greater among multi-cellular species. In the first place, viral infection is often characterized by an early viremic phase, with the consequent involvement of many different organs allowing that interchange of genetic material could occur between cells. And as long as it occasionally involved the germ-cell line, this would provide the opportunity for genetic changes even in somatic cells to be transmitted to the germ line of descent,23 so extending the range of genetic 'experience' of the species as a whole. Second, multicellular organisms can be viewed as the site of cohabitation of a wide range of different organisms, as discussed in: "The Cell as a Habitat." 33 One example of this may be the human colon that is so heavily infested with a wide range of prokaryotic bacteria, because this provides a potential point of prokaryotic/eukaryotic interface across which important genetic exchange might well occur. 13

**Darwinism and the Viral Evolutionary Theory**

It might be thought that the current theory would be strongly at variance with Darwin's views of evolutionary mechanisms, 34 but this is not so at all. Darwin's contribution to evolutionary theory was to delineate how natural selection could act on variety already established within the biosphere and so promote the evolution of increasingly complex forms. The basis of that variety was unclear to Darwin, and indeed where expressed, his ideas on the subject were distinctly Lamarckian. In an important sense, the current view is not at all in disagreement with that, because it sees the inheritance of acquired characteristics as being perfectly feasible through mechanisms of viral genetic transfer. Any point of conflict perhaps arises more with neo-Darwinism, although there should be no serious disagreement even there. The neo-Darwinists, quite properly, placed the inheritance of variability on a genetic basis, and the present view would merely suggest that the source of this heritable material is not always just the genes which have characterized that particular host species throughout all past generations. It may be true that I place more emphasis on a symbiotic or co-operative biological inter-relationship than on the Darwinist struggle for survival, but I do not at all deny the importance of natural selection as a mechanism sorting out new DNA substitution. I therefore suggest that there should be no serious philosophical barrier at all to the present general theory.
WHAT THE THEORY MIGHT EXPLAIN

Population Genetics

Before considering in more detail the evidence supporting the view proposed, let us first consider whether the theory provides an improved explanation for the problems of population genetics outlined in the previous chapter. In my view it does this well.

Haldane’s Dilemma Resolved.

In the first place, it offers a solution to Haldane’s dilemma of genetic load\(^{35}\) by altering substantially all three of the assumptions on which it is based (Ch. 8).

The first assumption, it will be recalled, is that population size remains relatively finite and estimable. The trouble is that this results in a calculated population size that is often far too small to account for the observed speed of evolution, particularly that occurring at the molecular level. But with the sort of viral/host interchange relationship envisaged under the present theory, the genetic pool with which any population or species was in contact at any one time would be enormously increased. Even a single cell infected with a number of integration viruses during its life-time might be exposed in this way to millions of viral genetic particles, each having at least some potential to insert at different points within the host genome, and each perhaps carrying differently evolved viral variants, as well as different host-gene fragments around its site of incorporation due to imprecise excision from some previous host. Add to this the broad range of viruses that may infect any one host species, as well as the potential for genetic contact between different species through the overlapping virus/host ecospheres discussed, and the effective genetic pool for even a single species could be enormous in comparison with traditional estimates of population size.

Even where there were few present-day viruses still interacting with particular host species, a long past history of genetic interchange from previous viral infections could still be responsible for extensive diversity within the evolutionary background. And whilst it might have been thought on general grounds previously that the largest populations would be found amongst the prokaryotic species, it is possible that similarly increased opportunities for genetic interchange under the present theory would give eukaryotes an effective population size just as large. Finally, the two levels of biological development, prokaryotic and eukaryotic, may be in much more genetic communication than previously supposed. In this respect, the traditional idea that ‘restriction’ endonucleases exist solely to prevent the entry of foreign genetic material into the host cell may need to be revised, because but there are now quite strong indications that these enzymes also facilitate the uptake of foreign DNA on occasions.\(^{13}\)

The viral view proposed here also bears strongly on the second assumption implicit in Haldane’s dilemma, namely that DNA substitution is a random or stochastic process throughout evolution. Although the present theory would see this as being essentially true at the most fundamental level of the primitive biological ‘melting pot’, it would become less and less so as the genes - or perhaps more correctly, genetic blocks or modules of them\(^{11}\) - were ‘processed’ higher and higher up the biological scale by viral /host interchange. Because this would provide a continuing selective testing of the genetic sequences at each stratum, by the time any gene or gene-fragment arising at the lowest evolutionary strata had become fixed by this means at some higher level, it would have had its own extensive selective evolutionary background history, so that although starting out as random mutational change, it would be far from random in the long run.

The third assumption underlying Haldane’s dilemma is that gene substitutions occur independently of one another during evolution, and as we have seen, one of the ways out of this is to postulate that they do not (chapter 8). Along these lines, the ‘threshold’ or ‘truncation’ models of selection propose that many gene substitutions are not independent of one another at all, in that they are linked together through subserving similar functions. This may provide a partial answer to the dilemma, but is not entirely satisfactory because it does not help us understand how such correlated
DNA substitutions could come about in the first place. The present view, on the other hand, would see modular genetic change as being the rule rather than the exception, with DNA substitution and integration occurring as whole blocks of DNA structurally linked together as viruses or virus-host genetic complexes. This provides a rational basis on which Haldane’s third assumption may realistically be challenged.

**Molecular Evolution**

When we can question substantially all three of the assumptions underlying Haldane’s dilemma in this way, much of what was previously puzzling about molecular evolution now becomes explicable.

1. **Neutral mutation.**
   This becomes much easier to understand. By the present view, DNA substitutions found in the higher reaches of evolution could now be seen as the result of genetic exchange and processing of initially primitive DNA through different viral/host ecospheres to higher and higher levels of biological complexity. Initial substitutions occurring at the lowest levels would indeed be seen as random, and therefore mostly detrimental as expected from any stochastic process. As such, they would largely be eliminated by negative selection, leaving the few more ‘neutral’ and ‘favourable’ ones to be subsequently passaged to higher levels of the biological scale. Along the way these, in turn, would be ‘examined’ by the processes of natural selection on each occasion, and accepted or eliminated depending on the outcome of that test. Of course, the terms ‘neutral’, ‘favourable’, and ‘detrimental’ are relative in this context, because their precise effects would depend both on the level being examined, and the biological circumstances prevailing at the time. Nonetheless it would remain substantially true that the bulk of mutations generated at the most basic level by random mutational change would indeed be detrimental, and would represent in effect the bottom of a very large iceberg of molecular DNA alteration mostly eliminated at the lowest levels of the biological melting pot. This would leave the rest, the relatively ‘neutral’ and ‘favourable’, with the chance of being subsequently refined and integrated (perhaps initially in non-coding regions of DNA – see Ch.10) according to selective circumstances, and processed to higher levels of biological sophistication.

2. **Non-randomness of neutral DNA substitutions.**
   The theory could also explain why, at higher levels, much of even neutral DNA substitution in nature does not appear to be altogether random because such substitutions would now be seen simply as those melting pot mutations to survive after others had been eliminated by processing through ever higher virus/host ecosphere levels. Although neutral in their effects, therefore, and perhaps random at origin, these substitutions would have become far from random in their evolutionary background development. Equally, the occurrence of a greater rate of mutation in some areas of DNA than others (so-called mutational hotspots) would be seen as being the product of preferred sites of viral integration.

3. **Molecular mutation rates.**
   The observation that mutation rates within many higher species show an apparent constancy per unit time rather than per generation span might also be explained. According to the present view, the rate of molecular change in various genetic codons should be determined not so much by the life-span of the particular host species, but more by the extent and duration of past virus/host genetic interaction. On this basis, a constancy of DNA substitution per unit time might be expected for as long as the viruses concerned continued to exist in any sort of infective relationship with the host species. Indeed, such constancy might even continue beyond that time, if remnants of viral function continued to offer a selective advantage via internal mobility of genetic material between the soma and germ line of individual cells.
Eventually, of course, there would come a time when the genome was totally stable with respect to the virus concerned, and any capacity for even internal genomic mobility entirely lost. This, in turn, could provide an explanation of why many codons appear to undergo a high rate of molecular substitution early during their evolutionary history when, for example, multiple gene duplications arise. By the present view, such duplications would be much more likely during the phase the virus affecting the codon had an infective relationship with the host species. Then, one might also sometimes only see the footprints of recurrent evolutionary rounds of viral insertion and excision. This could account for the highly repetitive DNA found within the eukaryotic genome. Such repetitive or ‘satellite’ sequences are important to the present general theory, and will be discussed in more detail in the next chapter.

4. Non-informational DNA.
The large amount of apparently non-informational DNA present in eukaryotic genomes might also be explained by a viral view, because this would now be seen as deriving from a suppression of potentially harmful viral genes incorporated into the host during successive bouts of viral infection over evolutionary time. The selection pressure to inhibit any such detrimental viral effects would certainly be very strong, and when taken in conjunction with the large number of different possible types of viral infection, the tendency for each to be repeated during successive generations, and the large number of sites at which viral insertion into the genome could be disadvantageous, the amount of non-functional DNA generated by such means could become very great, even within a relatively short space of evolutionary time. Of course, some apparently non-informational DNA might have other origins. It might well be, for example, that much of it was DNA useful in the evolutionary past of the species concerned, and even perhaps still important during the embryological development of each of its individual members, but now suppressed. Yet other DNA non-functional from the standpoint of protein coding might contribute important functions to cell architecture and/or regulation of gene expression, regardless of how it arose.

5. Genetic polymorphism.
This could also be explained by the present viral view. Indeed, there is increasing evidence that at least some polymorphism in nature is due to just such a viral mechanism. Thus, Blumberg’s classic work on ‘Australia antigen’ began as a study of molecular polymorphism, but ended up with the discovery that this genetic marker is in fact related in an evolutionary sense to infection with the hepatitis B virus, now known to be quite capable of integrating into the host genome under conditions of persistent viral infection. I suggest that this kind of mechanism is not at all the exception, and indeed provides the basis for most genetic polymorphism. Both DNA viruses and RNA retroviruses that had settled down after the acute phase of any infection to a more latent relationship with the host might well have opportunities to integrate some DNA into the host genome in this general way. And again, if this process involved the germ line as well as the organs more clinically affected by the infection, it too could result in the transfer of viral genetic material through ‘vertical transmission,’ with important evolutionary consequences for the species as a whole.

6. Heterozygosity.
The finding of widespread differences in the gene forms on each of the two members of any one chromosome pair within an individual is puzzling in traditional evolutionary terms (chapter 8), yet might even be predicted by a viral evolutionary view. Not only could a piece of host chromosome be transferred from one locus to a similar one in a new host by viral transduction, but reciprocal or double cross-over recombination between homologous segments on virus and host could lead to the part exchange of comparable DNA sequences between the two without any intervening viral integration at all. Either of these processes could result in a different version of a particular gene arising within one of the two members of any chromosome pair (gene ‘conversion’).
Other evolutionary data.

The surprisingly high genetic variability shown by some species which reproduce asexually, the peculiarities in others of chromosome structure and function (e.g. the formation of segmental loops and ‘Lampbrush’ chromosomes, and the modular arrangement of chromosomes during certain phases of the cell cycle, might all bear re-interpretation in the light of the present view. Another problem which might be better explained is the dilemma Darwin himself faced, namely the continued existence of very primitive life forms from earliest times until now, despite the emergence of much more sophisticated biological species which should readily have replaced them by selective competition. By the present perspective, these primitive forms would continue now, as in the past, to make an important contribution to the biological chain of evolution, by maintaining a melting pot for the early sorting and ‘processing’ of new molecular DNA mutational change. Indeed, the evolutionary stability of the whole range of biological forms intermediate on the evolutionary scale might well have a similar explanation, because of their importance to the ‘boot-strapping’ system of biological ‘improvement’ postulated here. The long periods of evolutionary stability of many biological forms is certainly very striking, and requires better explanation than merely being dubbed as ‘punctuated equilibrium.’

Others have recognized that viruses might make some contribution to biological evolution, especially that occurring at the prokaryotic level. In particular, Anderson put forward views and analogies not dissimilar to some expressed here, and in addition suggested that other puzzles of evolution might also be explained by a viral view. Thus, he regarded the constancy of the genetic code as strong evidence in favour of a continuing process of genetic interchange throughout the biosphere. The most usual explanation for this is that the genetic code did in fact vary early in evolution, but subsequently became fixed as a sort of ‘frozen accident’ once a dominant organism, or group of interbreeding organisms, emerged to displace all other forms. But Anderson’s point should be taken as being at least as good as the more traditional. Another of his arguments was that with the development of highly sophisticated biological forms, one might have expected total annihilation of disease-producing agents as simple as viruses. The fact that this has not occurred suggests that susceptibility to viral infection might confer a selective advantage on the species as a whole. This view is now enhanced by the finding that viruses actually have specific cell receptor sites, sometimes multiple, to facilitate their cell uptake and entry. Anderson also suggested that ‘parallel’ evolution (i.e., the emergence of similar biological forms in areas separated by wide geographic distances) - hard to explain in traditional evolutionary terms - is readily accountable on a viral basis as a “continuous flow and interchange of gene parts ‘on approval’.” (See also Jacob)

Therefore, although my own views of the evolutionary importance of viruses were developed independently of others, being stimulated by considering the possible relationship of viruses to cancer, their potential evolutionary significance has not gone unrecognized by others. Indeed, in some ways all I have done is to take the argument a stage further.

I emphasise one point. Some puzzling aspects of evolution explicable by the present viral perspective were not part of its theory generation process. Popper, I think, would have approved, because he held one of the tenets for a good theory to be that it should explain more than it was initially put forward to do.

**PREDICTIONS OF THE VIRAL THEORY**

Before going on to consider in detail the evidence in favour of the present view, it might be useful to look at the sort of predictions such a viral theory would make. If it is correct, we should expect to find recognizable evidence of a viral make-up to higher genomes, either as viruses themselves, viral
genetic components, or a molecular sub-structure and genomic organization consistent with an ancestral capacity for genetic mobility. There should also be evidence in keeping with repetitive viral infection of the type proposed. Of course, recognizing any virus integrated into higher species might now be difficult, for the following reasons:

First, viruses integrated in the evolutionary past might by now have entirely faded out of existence as viruses per se, so as to be no longer recognizable by present-day standards as viruses at all. Second, even where past-integrated viruses might have borne some resemblance to their present-day counterparts, neutral evolution in viral segments long suppressed in the host might again have largely obscured it. On the same basis, it could be exceedingly difficult to recognize any ‘host’ DNA segments with an important evolutionary background of genetic exchange, and knowing whether such segments had originated in host or virus might be well nigh impossible. Nonetheless, if viruses have been as important in the build-up of higher genomes as I suggest here, then discernible traces of those origins should remain, particularly where internal genomic mobility continued to offer a selective advantage to the species long after any capacity for cell to cell viral transfer had been lost. And even if not overtly expressed, a basically viral nature might still become apparent under appropriate conditions, for example when the cell was damaged in some way (see Ch. 7), or when investigation was carried to the molecular biological level of DNA sequence type and organization.

EVIDENCE OF A VIRAL CONTRIBUTION TO THE PROKARYOTIC GENOME

Virus and other mobile element integration into host

Integration and exchange of viral genetic material between bacteria is turning out to be widespread, and has been increasingly recognized for its potential contribution to the genomic evolution of both host and viral species. In addition to complete viral sequence integration, there is an opportunity for partial viral sequence exchange, including the uptake and interchange of the small sub-unit mobile genetic elements that viruses may carry.

Plasmids are small DNA particles that are not usually directly transferable between bacteria, but may achieve this effect by being carried as part of the sub-unit structure of conjugative and other phages. And once transferred, they may take up residence within the bacterium as an epizome, covalently integrated with the host cell genome. Such plasmid transfers have become widely recognized because of their importance to the exchange of antibiotic resistance factors among bacteria. This was first described in the gram negative bacteria, but is now known to involve gram positive species as well. Moreover, the process is not necessarily confined merely to genetic exchange between bacteria of the same species. Indeed, one of the first examples of resistance to antibiotics arose in the Shigella bacillus in Japan following the widespread use of antibiotics there for bacillary dysentery, and once the plasmid carrying the antibiotic resistance factors had become established, it rapidly spread to other gram negative species as different as salmonella and non-pathogenic E. coli. Mobile plasmids are now known to carry a whole variety of different antibiotic resistance factors, but their importance may go well beyond this. Representing as they do one of the few markers of genetic exchange in nature, they may be reflecting a much more widespread process of natural genetic engineering among the prokaryotic species. The usual integration mechanism of plasmids and other circular extra- chromosomal elements into host is by classical ‘legitimate’ (‘rolling circle’) genetic recombinational crossover at areas of homology between the two; actually, much of this homology may now owe its existence to previous rounds of genetic exchange between virus and host.

In addition to the plasmids and insertion elements, there is another important group of small mobile transmissible genetic elements, namely insertion elements and transposons. But whereas
plasmid integration occurs by ‘legitimate’ recombination at areas of DNA homology to yield an epizome, integration of insertion elements is much less dependent on such homology and results in the generation of flanking repeat sequences in the host. This lack of dependence on homology for integration means that transposons and like insertion elements can become integrated at a wide variety of host chromosomal sites. Because of this, they are looked upon as being highly ‘promiscuous,’ and the process of their integration is called ‘illegitimate’ recombination. Actually, some insertion elements have preferred integration sites (possibly related to the existence of some ‘microhomology’), and this could be important to our consideration of the generation of mutational hotspots within the prokaryotic genome.

The detailed mechanism of transposon integration has not yet been fully determined, and may well differ with different sorts of transposon. It seems most likely that staggered nicks are first made in the host DNA site of intended insertion, followed by the formation of co-integrate intermediate molecules. The whole process requires a ‘transposase,’ probably encoded by the transposon itself. The inverted repeat sequences at the ends of insertion elements may hold the key to their integration, and perhaps do so by coming together before insertion to form a sort of sticky double-stranded DNA ‘panhandle’ to provide the transposon with a ‘homologous stem’ for its attachment to the chromosome. Whatever the case, it seems that the transposon then just burrows its way into the host genome regardless of the niceties of any homology, and interleaves itself either way around the now looped-out insertion element to become either a direct or inverted integrated transposon sequence. Once integrated, these elements have the potential to exert a powerful effect, not only from the type of genes they carry (e.g. antibiotic resistance factors) but through their effects on adjacent host DNA sequence expression, regulation, and arrangement. Gene re-arrangements include gene deletions and transpositions, but the inverted terminal repeat sequences mean that such transposons can also promote chromosomal inversions, and we know from the work of Dobzhansky that this can have an effect to stabilise parts of the genome against evolutionary change. Being transposons, they are able to move around the host genome after initial insertion, and this provides a continuing potential for the re-arrangement and transfer of host as well as transposon gene sequences within the chromosomal network of individual prokaryotic cells. Finally, although most transposons seem incapable of independent transfer between cells they may, like the plasmids, incorporate themselves into higher mobile genetic units such as bacteriophages to achieve the same end. In this way, transposons may be just as important as plasmids in transferring genetic material like antibiotic resistance factors among, and even between, bacterial species. Indeed, because they do not require host DNA homology for insertion, they may be even more important than the plasmids in this respect.

Whether mediated by plasmids, transposons, or other mobile genetic elements, such ‘subunit’ genetic exchange between bacteria is now known to be very widespread, involving gram positive as well as gram negative organisms, and as we have seen, crossing species barriers as different as Shigella and E. coli. Its significance could be even greater if overlapping virus/prokaryotic host ecospheres do exist in nature in the way postulated above, because this could facilitate genetic recombination between the viruses (with or without attached host DNA) within the area of overlap, and significantly extend the range of host/virus exchange. There is evidence that some virus/host ecospheres do overlap in this way. Thus, Olsen and Wright have shown that some bacterial species are able to carry a number of different plasmids at the same time, each with different host/virus ecosphere distributions, and they recognized the important contribution this could make to bacterial evolution through genetic recombination between the plasmids concerned. Wherever it may originate, there is no doubt that viruses do recombine with each other in nature at times, although so far this has been best documented with the eukaryotic viruses.
Host sequence transfer by prokaryotic viruses

Viral Transduction

Additional host sequences may be transferred along with the virus, by viral ‘transduction.’ This is best described with the temperate bacteriophages of E. coli, and is a process whereby the integrated phage is excised imprecisely from its host genomic site, with the result that a piece of the host chromosome becomes covalently attached to the newly-formed viral genome and carried with it into the new host at next infection. One important consequence is that this extra host fragment can then undergo recombination with a homologous piece of new-host chromosome, so replacing it completely (gene-segment ‘conversion’). The net effect is that the virus has merely acted as a vehicle for the exchange of cellular sequences between two hosts without becoming permanently integrated into the new host cell at all. This whole process may not only lead to the multiplication and amplification of host gene sequences within any host species, but in reverse, could result in the deletion of specific fragments of DNA from host.

Specialized viral transduction in prokaryotes (at mutational hot-spots) has other important consequence for our theory. First, it can promote a region of increasing homology between the DNA on virus and host, and so provide an ever-stronger site for the integration of (circular) viruses via the classical mechanism of genetic recombination. Second, the repeated involvement of any segment in the process of viral transduction may eventually make it virtually impossible to tell whether that segment had originated in virus or host. Whatever the case, whilst they were being shuffled both into and out of host and viral genomes, such virally-transduced host sequences could have profound effects upon the function of each along the way, and so contribute importantly to an eventual dynamic co-evolution of both.

The Extent of Viral Contribution to the Prokaryotic Genome

Allowing that viruses may integrate their DNA into prokaryotic species, the question now is, “How much of the prokaryotic genome at any one time is actually composed of material exchanged by viruses in the evolutionary background?” Even if we restrict our discussion to the exchange of genetic material by still-recognizable insertion elements, there are at least some who believe the answer is: “A great deal”. Thus, Szybalski and Szybalski calculated that as much as 30% of the E. coli genome might be composed of integrated phage particles. Reanney went even further, and suggested that the E. coli genome may be able to take on two alternative states of existence, the first and more usual one where the phage modules are co-integrated in linear form as host chromosome, and the second emerging when bacterial cell integrity was threatened or disrupted in some way, so as to yield a series of separate viral forms. Reanney argued, further, that such a system could even provide an important mechanism allowing specific ecosystems as a whole to survive under conditions of threat.

Of course, even if prokaryotic genomes were built up entirely from a background of genetic interchange, we would not expect them to be just a collection of insertions elements. First, in viral transduction, the genetic material exchanged is of host origin, and moreover can be integrated into the new host without producing any lasting traces of the viral vehicle that carried it there. Second, the viral part of the integrated genetic material can be something less than a complete phage or viral genome, or even less than the mobile genetic sub-unit elements they carry, because once homology had been built up between viral segments in host and the original virus, similar and repeated viral recombinational part-exchange between them might blur their actual background viral origins in the same way. How much this actually occurs within the prokaryotic species in nature is unknown, but we will see below that it certainly applies to the eukaryotic transforming viruses.

Recognizing any particular sequence as stemming from viral or other mobile genetic element origins can therefore be difficult, doubly so if there were an evolutionary effect of ‘neutral mutation’ to
substantially alter sequences from their originally integrated viral form. Such factors pose serious difficulties in defining which particular sections of the prokaryotic genome have had any background of genetic exchange, and we therefore need to look elsewhere for evidence supporting this postulated mechanism.

**Indirect Evidence of Viral Make-Up to the Prokaryotic Genome**

**Similarity of bacterial and viral functions.**

There are certainly close similarities between some aspects of normal prokaryotic host cell function and viral behaviour. In particular, the various mechanisms of gene exchange between different strains within bacterial species are closely akin to the way viruses may transfer and integrate genetic material into their bacterial hosts. Thus, “sexual conjugation,” which results in the transfer of a small “sex factor” plasmid between two bacteria of the same strain at the time of mating, is very similar to the way antibiotic resistance factors can be transferred between bacteria on “conjugative phages.” Another means of genetic exchange between similar strains is bacterial ‘transformation’, classically observed with the gram positive pneumococcal species. There, naked DNA is transferred from one bacterium to another in a process that has exceedingly close parallels with the way naked viral DNA can be transferred from virus to host during phage infection. Of course, when this genetic exchange is limited to the transfer of material between different strains within bacterial species, it tends to be regarded as a perfectly normal part of bacterial function (e.g. in the case of sexual conjugation, as part of normal mating behaviour), but two points must be made. First, the mechanisms of gene transfer between bacteria are so close to those of viral genetic transfer that they are entirely consistent with having originated from an acquired viral characteristic. True, many regard the reverse as being more likely, namely that viral functions have been derived from bacterial ones, but as we shall see later, this view is on much less certain grounds now than previously, particularly as evidence accumulates to suggest a sub-unit mobile genetic element sub-structure to many viruses. Second, no matter how these processes of bacterial DNA interchange have been derived, their presence emphasizes the wide extent of genetic exchange in the bacterial species, and suggests that natural genetic engineering in the broader context is both important and widespread at the prokaryotic level.

Gene exchange between members of any one bacterial species may occur by means that are more obscure, but even then it often has a modular character highly reminiscent of viral genetic exchange. An example emphasizing the potential extent and rapidity of such genetic exchange is seen in an experiment where different strains of the organism Bacillus subtilis, each containing several different antibiotic resistance markers, were grown together in soil culture. After a period of only a few days, there was extensive interchange of genetic resistance factors between the different bacterial strains, and moreover one that lead to a radical re-organization of the whole genome. Now, Bacillus subtilis does not normally partake of sexual conjugation, so that some other process of genetic exchange must have been involved. Regardless of how the bacterium achieved it, the way the whole process resulted in genes being transferred between the bacteria as linked blocks or genetic modules is highly reminiscent of the process of viral genetic transfer, and entirely consistent with its having derived from viral sub-unit origins.

It is even possible that some structure and function of bacteria have derived from viral origins. We know for example that insertion elements can subserve a variety of genetic regulatory roles within the E. coli genome. Moreover, it is beginning to look as if these are not isolated examples, and that mobile genetic elements may make an extensive contribution to normal bacterial genomic structure, function, and organization. E. coli, for example, contains the sequences of at least two insertion elements present in multiple copies throughout their genomes, and this could explain the observation that many bacterial replicons appear to be assembled on a modular principle. Indeed,
some now believe that many of these insertion sequences provide essential mobile joint sites connecting sub-modules of the prokaryotic genome together. Insertion elements, like transposons, are often relatively ‘promiscuous’, but they also tend to have preferred sites of insertion, and this could help explain mutational the existence of hotspots. Certainly, if this process occurred at all widely, such intra-genomic mobility would add greatly to the evolutionary role of insertion elements, especially as they so frequently excise imprecisely from their chromosomal integration sites, so allowing that they could transfer contiguous segments of the bacterial chromosomal DNA as well. In the case of the so-called mutator phage of E. coli, for example, imprecise excision is at least ten times as common as precise excision, so that ‘internal transduction’ of this type could be very important.

In concluding this section, I emphasize that genetic exchange within prokaryotes is widespread, and that no matter what its origins, the genes so exchanged are often integrated covalently into the host genome. Because of this, they may subsequently be spread into the host-line progeny in what is essentially a process of “infectious heredity”. The potential importance of the latter, long recognized by microbiologists, is yet to be widely appreciated, but is highly relevant to our present general analysis.

The case for an important role of viruses and other mobile genetic units in the make-up of the prokaryotic genome, though incomplete and circumstantial, is persuasive, and the more time has elapsed since I originally proposed this view, the more the evidence has accumulated in its favour. For example, insertion sequence elements are now known to be present in almost all bacterial genomes.

EVIDENCE FOR A VIRAL CONTRIBUTION TO EUKARYOTIC EVOLUTION

There is evidence consistent with an important role for viral interchange in eukaryotes as well, even if the general idea of an essentially viral nature to the build-up of the genome has been less well explored at this level. The most suggestive of the data come from the microbiological and molecular biological fields, so we will look at this first, and start by analysing the evidence for gene transfer between eukaryotic hosts by viral or other genetic mobile element means.

Virus Infection in Eukaryotes

Viral infection is widespread among eukaryotes, and is usually regarded as being entirely detrimental to the host. But such a view may merely reflect the fact that the study of viruses has arisen largely from the investigation of disease, with the result that microbiologists have been much more interested in the harmful effects viruses may have than with the possibility of their ever conferring any advantageous characteristics on their hosts — other, of course, than those which might accrue from the evolutionary selection of those host individuals most capable of surviving viral onslaught. But as Anderson pointed out, if agents as simple as viruses have served no useful evolutionary purpose at all, how is it that they can still so readily evade all the powerful and sophisticated armamentaria of their eukaryotic hosts and produce such devastating disease? This alone would seem to argue in favour of their having some positive role to play in eukaryotic evolution, especially since eukaryotic cells actually have receptors for viral uptake.
Integrated viruses and eukaryotes

1. Complete virus integration

a). DNA viruses.

Single stranded DNA (ssDNA) viruses can integrate into host genomes. So, too, can some dsDNA viruses such as simian virus SV40, adenoviruses, polyoma virus, and hepatitis B. Their potential contribution to evolutionary cellular genetic change is large, as discussed below.

b). RNA retroviruses

These are the best studied of the ‘integration viruses’. They can reverse transcribe a DNA copy of the RNA viral template and integrate it co-linearly with the host genome as a ‘provirus’. Baltimore and Temin and Mizutani first discovered these as ‘type C’ retroviruses, and we now know that a variety of retroviruses are widely integrated into the genomes of eukaryotes. Many are capable of causing disease, but some are appear quite harmless to their hosts.

Retroviruses have important potential in evolutionary terms because, once integrated, they not only remain capable of being transferred to other cells by conventional ‘horizontal’ routes of infection, but they can also be passaged ‘vertically’ through the germ line of the host to become an integral part of the inherited genome of that species. This is reminiscent of the process of “infectious heredity” already discussed with prokaryotes, and could exert evolutionary effects just as profound. Such integrated ‘endogenous’ viruses exist in multiple copies throughout eukaryotic genomes, including those of primates and man. Todaro and co-workers have implied that this multiplicity of endogenous virus number has arisen from an amplification of some ancestral viral insertional event, but in my view it is equally likely to have been derived from the repeated integration of exogenous viral sequences following repetitive viral infections of many generations of hosts, each immunologically naive with respect to the virus concerned. Moreover, the evidence now appears to favour this view. Thus, many strains of mice contain multiple copies of the provirus sequence of mouse mammary tumour virus (MMTV) inserted at a number of different points within their genomes. One study that followed a single such laboratory mouse strain over a period of about sixty years found a widespread DNA heterogeneity between the various retrovirus copies, more in favour of their having arisen by recurrent rounds of independent ‘horizontal’ infection than by amplification of some single initial viral insertion event.

Since the first description, many other retroviruses have been found integrated into eukaryotic genomes including man, doing so either alone, or with the help of unrelated viruses,, even replication defective adenoviruses.

c). Non-retroviral RNA viruses.

These can also integrate into eukaryotic host genomes. The key to this is probably their latent persistence within their hosts, and the availability there of reverse transcriptase enzymes from already-integrated host retroviruses, or from similar host retrotransposable elements, such as the so-called LINE-1 element repeat. Even non-viral cellular RNA can be reverse transcribed on occasions. This is the case with some eukaryotic gene mRNA when the primary mRNA transcript has already been processed to remove the mRNA intervening between exons. The resultant conjoined exon sequences can sometimes then be reverse transcribed back into the DNA genome, to form processed genes or ‘pseudogenes’ (see chapter also 10).

All of the above viruses can transfer between hosts within their various host/virus ecospheres, and integrate there into new-host DNA. Different positions of insertion into can produce evolutionary novelty by a variety of mechanisms, including exon/gene duplication, inversion, transposition, rearrangement and deletion. Over evolutionary time, the integrated viral sequences can eventually
lose their more aggressive and potentially oncogenic characteristics and come to be regarded more as host than virus, sometimes to the extent of fulfilling perfectly normal cell functions.\textsuperscript{108}

Even simple RNA viruses might have played a role in the eukaryotic genomic build-up of the evolutionary past if the genetic code were ever written early during evolution as RNA rather than DNA in the way some suspect.\textsuperscript{109} There are signs that they may still do,\textsuperscript{30} including by occasionally being ‘accidently’ reverse-transcribed by reverse transcriptase activity within the cell.\textsuperscript{9} Because there are many sequences within mammalian genomes highly complementary to parts of primitive RNA viruses and viroids,\textsuperscript{110} such a mechanism is certainly possible.\textsuperscript{30} Its evolutionary potential is made even greater by the high rate of mutation of RNA genomes.\textsuperscript{111}

2. Integration of partial viral sequences in eukaryotic host genomes

a). DNA mobile elements and retrotransponsons

As well as insertion of complete viral/retroviral sequences, there is now good evidence of multiple DNA transposons and partial retroviral sequences in eukaryotes. Indeed, the genomic contribution of complete viruses is vastly outnumbered by repetitive ‘mobile elements.’\textsuperscript{11} These are the DNA transposons and retrotransposon-like elements.\textsuperscript{7,11,30,82,106,112-115} Most are highly reminiscent of DNA virus or retroviral subunits,\textsuperscript{11,106} and have a correspondingly high capacity for movement around the host genome,\textsuperscript{7,11} with all the implications that holds for alteration of gene expression,\textsuperscript{11,12,116} chromosomal duplication, ‘exon shuffling’, oncogenesis, etc.\textsuperscript{11,12,93,117} Of course, host cells must have mechanisms to minimise any initial threats to genomic instability such transposons/retrotransposons undoubtedly pose.\textsuperscript{11} Nonetheless, these elements have come to represent about half of the human genome,\textsuperscript{118} and 85% of the maize genome,\textsuperscript{119} and continue to expand. This is entirely consistent with the thesis proposed here that retroviruses and even the eukaryotes themselves may have evolved by the gradual build-up of smaller mobile genetic subunits (see also Flavell,\textsuperscript{120} Lankenau et al.\textsuperscript{30}).

The whole process of reverse transcription and ‘retro-transposition’ is becoming increasingly recognised as an important force shaping the eukaryotic genome.\textsuperscript{30,112,121} It certainly has real potential to boost host evolution whenever it involves the germ line, inhibitory limits on such a process notwithstanding.\textsuperscript{12,15,16,122,123}

b). Retroelements

Some retroelements within eukaryotic genomes are incapable of de novo retrotransposition, but many can do so with the aid of better-endowed elements nearby, as in the way repetitive LINE-1 elements can help smaller Alu SINE elements transpose within the human genome.\textsuperscript{11,124}

\textit{Big fleas have little fleas, Upon their backs to bite 'em,}
\textit{And little fleas have lesser fleas, and so, ad infinitum.}\textsuperscript{125}

3. The role of transposons and Retrotransponsons per se in horizontal evolution

Actually, most genomic retrotransposons, whether LTR\textsuperscript{3} or non-LTR, are highly defective, i.e.

\textsuperscript{3} LTR: Long Terminal Repeat Elements that characterise the ends of most retroviruses
unable to exit the host genome unaided. But in my view, just as LINE elements can help Alus transpose, so too can LTR elements help non-LTR ones, and whole retroviruses lend a hand to mobilizing defective LTR retrotransposons, all essentially by a process of ‘cooperative hijacking,’ which forms, I suggest, a sound basis for retroviral driven horizontal evolution. So whilst widespread transposition and retrotransposition within individual genomes can lead to substantial intragenomic change, this may be quite overshadowed by the potential these elements have to escape their genomes on the backs of other viruses and pass to other individuals and even other species. Understanding this process is currently seen as being the ‘holy grail’ of inter-individual horizontal mobile element transfer. My view is that it is relatively common, and is achieved by such mobile elements being transduced whenever they are happen to lie adjacent to integrated retroviruses at the time the latter exit cell DNA - by imprecise excision to infect other hosts within its host/virus ecosphere. In the case of DNA transposons, and even with integrated retroviruses/retrotransposons, this same end might sometimes be accomplished by helper DNA viruses, such as the adenoviral group, perhaps during something as simple as a common cold or ‘flu. In such cases, both the retroviruses and mobile elements concerned would have access to populations and even to species quite outside their normal host-residing range, with all the potential that holds for genuine evolutionary novelty.

. . . And the great fleas, themselves, in turn, have greater fleas to go on;
While these again have greater still, and greater still, and so on.

**Host gene segment transduction by viruses and their sub-unit mobile elements**

Important though the above is, it would have even more potential if inter-individual transduction of host sequences could occur along with the virus, i.e. host-sequence uptake and transfer by viruses to other hosts, as happens with bacterial viruses. Then, host gene sequences could be shuffled not just around any one individual genome, but between host genomes, between viruses, and even between quite different host species within the realm of infectivity of the virus concerned. This is a mechanism of enormous potential. Actually, uptake of host segments into viruses is implicit in the process of viral transduction of retrotransposons already described. It’s just that here, we are discussing the consequences for actual host segment rather than virus segment transduction.

**1. Transduction of host segments within genomes by DNA transposons and RNA retrotransposons.**

Mobile genetic elements have the capacity to transduce host gene fragments to a different genomic site within the host, such as in ‘exon shuffling.’ As discussed with the prokaryotes, it results from imprecise excision of the mobile element together with an adjacent gene fragment, and insertion of both into some new chromosomal site either randomly or, in ‘specialised transduction’, into a more specific site or ‘mutational hot spot.’

**a). Transduction by DNA transposons**

There is good evidence that several types of several types of DNA transposons have transduced up to thousands of gene fragments in grass species. The DNA transposon, MULE, has been long suspected of capturing and carrying host gene fragments. Jiang and colleagues have identified over 3000 so-called PACK-MULEs containing fragments from more than 1000 cellular genes, with some MULE elements capturing fragments from multiple host gene/exon
sites.\textsuperscript{131} This is a truly impressive feat, but seems to me unlikely to have been so widely achieved by \textit{intra}-genomic transduction in the way some have implied,\textsuperscript{7,132} and would very likely have needed help from more transmissible viral forms.

\textbf{b). Retrotransposons:}

There are many examples of \textit{intra}-genomic host gene transduction,\textsuperscript{7,133,134} including retrotranspositional intron/exon shuffling.\textsuperscript{129,135} As with DNA transposons, the mechanism by which this could then be propagated \textit{across} populations and species has so far been less clear.\textsuperscript{12}

\section*{2. Transduction of host segments \textit{between} genomes by DNA viruses and retroviruses}

\textbf{a). Host sequence insertion into DNA viruses}

Eukaryotic viral transduction of host segments has been well demonstrated with DNA viruses,\textsuperscript{19,136} and might also contribute importantly to horizontal evolution.

One of the best examples is where the DNA simian virus SV40 is grown experimentally in monkey kidney cell culture. There, an increasing DNA part-exchange takes place between virus and host, probably on the basis of an initially-slight but ever-increasing homology building up between the two.\textsuperscript{85} As a result, host sequences are repeatedly incorporated into the SV40 virus.\textsuperscript{86} This observation has now been amply confirmed\textsuperscript{137} and extended to show that, whilst tending to favour some sites of viral insertion over others, the host DNA transferred may be integrated at many different points within the viral genome.\textsuperscript{138} With continued passage of the virus on cell culture, a viral genome may eventually emerge that contains \textit{mostly} host cell DNA, and only a fraction of the original SV40 parental viral strain.\textsuperscript{139} Different cellular sequences may become incorporated into the virus in this way,\textsuperscript{138,140} but the most frequent is derived from a relatively restricted host region.\textsuperscript{137} To put things in perspective, however, it should be said that direct examples of DNA virus transduction in nature have so far been relatively sparse.

\textbf{b). Retroviruses.}

The capture of host genes by mobile elements was first discovered in the context of cellular oncogenes transduced by retroviruses.\textsuperscript{7,108,141,142} These transforming sequences were once regarded as being an integral part of the retroviral genome, but now that they are known to exist so extensively within eukaryotic host chromosomes - even quite separate from the virus itself - it has become difficult to know whether they should be viewed as of originally viral or host origin.

In the present context, the important conclusion is this. Given the capacity retroviruses have to infect a variety of species, and their innately mobile-element sub-unit structure,\textsuperscript{30,112,143} it seems highly likely that they are the agents primarily responsible for genetic interchange of host (and viral subunit) sequences \textit{between} individuals and populations. Internal mobile-element mediated transduction of host-gene fragments may underlie much \textit{intra}-genomic host segment evolution,\textsuperscript{11,134} but it seems highly likely that retroviruses \textit{per se} are the real \textit{‘drivers’} of inter-individual \textit{horizontal} evolution (pace Kazazian\textsuperscript{11}).

\textquote{‘But best of all are host gene fleas, that ride on backs of viruses’}

\textbf{Defective Viruses and Eukaryotic Genome Evolution}
It might be thought that for viruses to have a maximal evolutionary impact, it would be important that their various functions, in particular their capacity for host to host transfer, should not be affected by any attached host DNA segment. Yet such transduction is not infrequent. Indeed, some transforming viruses can be so modified by DNA sequence exchange with host that they become totally 'replication-defective.' The question therefore arises whether, by preventing transfer between different hosts, this mechanism could nullify the contribution such viruses might otherwise make to horizontal evolution. However, any limitation of their role in this respect is much less might appear. First, whilst it is true that present-day interaction between virus and host may produce 'replication-defective' viruses, this may not always have been the case in the evolutionary past; nor today is it by any means likely to affect all virus particles involved in any single viral infection. Second, where any replication-defective virus happened to arise during infection of the host germ line, it could still contribute to host integration and intra-host genomic shuffling. It might even minimise the chances of disrupting the cell by overt viral disease, either directly or through the formation of 'defective-interfering' particles that can limit the replication of remaining viruses with a full genomic complement. Thirdly, as things turn out, the replication-defective nature of many of these viral variants may be more apparent than real, because they often remain perfectly capable of being transferred between different host individuals with the aid of 'helper' wild-type viruses to supplement their defective viral functions. A fascinating variant of this is in the way adenoviruses, even replication defective ones, can help retroviruses extend their infectivity beyond their normal host range. The adenovirus does this by facilitating retroviral entry into cells that have no receptors for the retrovirus at all.

Integrated Viruses and Normal Eukaryotic Host Cell Function

The foregoing has demonstrated a number of different mechanisms by which viruses become integrated into host genomes. Moreover, there is increasing evidence that integrated endogenous retroviruses and/or their sub-unit sequences are importantly involved in many quite normal eukaryotic cell functions. This includes a role for endogenous viruses in coat colour mutations of certain strains of mice, in the normal functioning of the immune system, in the expression of retroviral antigens on leukocytes, and on haemopoietic stem cells including those of man, and in viral antigen expression during eukaryotic foetal and placental development. Cellular 'oncogenes' are also becoming increasingly recognised for their importance as hormone receptors, and, perhaps not surprisingly, as mediators of normal cell growth and proliferation. Furthermore, sequence similarities between the viral and cellular versions of these genes are often very close. Apart from oncogenes, other mobile subunit viral genetic elements such as transposons/retro-transposons have been shown to make an important contribution to the normal structure and function of the eukaryotic genome.

Not that the case for a role of viruses and transposition elements in normal eukaryotic cell function hinges entirely on the expression of viral genes within functional or structural cell proteins. To begin with, even where they are not now expressed in adult members of present-day species, they might well have played a significant role in the evolutionary past of the species concerned, and could continue to do so during the embryonic development of each of its present-day members. In addition, they could make a crucial contribution to the organisation, structure and expression of eukaryotic DNA without ever being expressed in protein at all. Indeed, long after the first recognition of eukaryotic mobile genetic elements by McClintock in 1951 as mobile genetic controlling elements in maize, the evidence now accumulating is that transposition/retrotansposition and viral sub-unit mobile genetic elements can exert a powerful influence over both the structure and mode of expression of a wide range of eukaryotic genomes.
which underlies the generation of antibody diversity in mammals is strikingly akin to the behaviour of transposable insertion (IS) elements, and may well turn out to have stemmed from such origins. Transposition may also be an important alternative to gene repression as a means of controlling the expression of a whole variety of genes.\textsuperscript{163}

From the structural viewpoint, the more the evidence accumulates, the more the eukaryotic genome looks to be organised on a modular or mosaic basis,\textsuperscript{30,166} with particular sequences being responsible for the coding of broad domains of function,\textsuperscript{167} and with genomic sub-modules being linked together by insertion elements serving as joints and/or connectors,\textsuperscript{160} all highly consistent with an important role for sub-unit mobile elements in shaping the eukaryotic genome.\textsuperscript{112} Just how much integrated RNA retroviruses and related sub-unit mobile DNA elements are involved in eukaryotic gene structure and function in this way\textsuperscript{112} is uncertain, but it would be surprising indeed if this and other sometimes apparently ‘non-functioning’ eukaryotic DNA turned out to be mere genetic ‘junk.’\textsuperscript{168}

**SPECIATION**

As well as having the potential to subserve gradual or phyletic evolution, the general mechanism proposed here might also explain some of the puzzles of speciation itself, particularly its not-infrequent sudden or saltation nature.\textsuperscript{45,46} Such dramatic steps in evolution are difficult to explain on any random mutational (see chapter 8) or natural selection basis,\textsuperscript{45} but might well be accounted for by the sudden introduction of a new (viral/host) gene from elsewhere. We certainly know that chromosomal translocations provide at least one important mechanism for speciation, namely the Robertsonian translocations which constitute the genetic difference between sheep and goats.\textsuperscript{169} We will look at other aspects of this in more detail again in chapter 10.

**CONCLUSION**

The viral interchange mechanism discussed has obvious potential importance in evolutionary terms. The evidence that it is involved to the extent I suggest is far from complete, but that which has accumulated since I first put forward my views\textsuperscript{2,3} is increasingly in its favour.\textsuperscript{7,11} The theory sees genetic novelty in evolution as stemming largely from the acquisition of foreign genetic material by viruses, rather than from random DNA point-mutational change, and in this respect, the current hypothesis has an obvious neo-Lamarckian flavour (cf. Steele\textsuperscript{23}). Others have highlighted the evolutionary importance of mobile genetic elements within individuals, but the emphasis here is critically on *inter-individual* genetic exchange via the medium of viruses, to give this general mechanism its full evolutionary potential.

Finally, it might well be that for inter-host gene flow to occur in any important way, the cell must be damaged to release the viral elements from their normal existence as part of the stable collection of ‘host’ chromosomal loci,\textsuperscript{19} making the general proposal difficult to test fully until such aspects have been taken into account. So perhaps this brings us full circle back to the proposed mechanism of cell damage by physiological or psychological stress. Whatever the case, we certainly have to remain open in our approach to understanding the mechanisms of evolution, particularly given the evidence for a nonrandom, ‘directed’, and even anticipatory nature to some genetic DNA mutational change.\textsuperscript{170,171}

There is ease and there is disease; in between those extremes, much remains to be known.
GLOSSARY OF TERMS

**Prokaryotes:** A group of organisms lack a cell nucleus or any other membrane-bound organelles. Prokaryotes belong to two taxonomic domains: the bacteria and the archae. We are concerned here mostly with the bacteria.

**Eukaryotes:** Organisms other than the bacteria, such as the plants, insects, and animals. They have a cell nucleus and other membrane-bound organelles.

**LINE elements:** Long internal repeat elements interspersed within the genome

**SINE elements:** Short internal repeat elements interspersed within the genome

**Alu repeats:** Short repetitive stretches of DNA originally characterized by the action of the Alu restriction endonuclease. Abundant in the primate genomes, including man. Derived from the small cytoplasmic 7SL RNA, a component of the signal recognition particle.

**Exons:** The relatively short interspersed stretches of the gene derived from transcribed DNA that will constitute the RNA for translation of the gene.

**Introns:** Those long interspersed stretches of the gene derived from transcribed DNA that are transcribed but not translated.

**LTR:** Long Terminal Repeats at the ends of retroviruses

References


86. Rao GR, Singer MF. Studies on a defective variant of simian virus 40 that is substituted with DNA sequences derived from monkey. II. Structure of DNA. *J Biol Chem.* 1977; **252**:5124-34.


