CHAPTER 10

RECENT INFORMATION ON THE MOLECULAR BIOLOGY OF DNA

‘The aim of science is to acquire not just information, but understanding - of what things are and why they happen.’ Thomas Nagel

INTRODUCTION

No modern day theory of evolution could be complete without taking into account the data from the recent discoveries in molecular biology, namely:

1. Repetitive DNA.

It transpires that there is a most unusual organization to eukaryotic DNA. Not only is there a staggering large amount of apparently non-informational DNA present in eukaryotes, but much of it is now known to be highly repetitive, often consisting of extremely small sequences repeated up to millions of time. This repetitive DNA has been widely studied in the past few years, and certain aspects of its organization and distribution within the genome are gradually becoming clearer, but its function remains totally obscure. Some regard it as mere genetic ‘junk,’ but there is increasing evidence that at least part of it can be transcribed, and this may turn out to be important in relation to the regulation of gene expression. Its origins are even more obscure, and this is the major aspect we will address in the present chapter, particularly the possible relationship to viruses

2. Non-coding DNA.

Genes in higher organisms are divided into exons separated by many segments of intervening non-coding DNA. Both are transcribed, and the exons joined to form the final gene, with the intervening mRNA being clipped out. Translated genes, which code for functional cell proteins and enzymes, constitute a very small proportion of the total genome (less than 1.5% in humans). The emphasis here will be mostly in the intervening sequences (Introns), and their possible origin as viruses or other mobile genetic elements.


This has become an important area, but we will not discuss it to any extent, save where it might
relate to viral influences on evolution, and whether it can be seen in any way as being truly 'epigenetic.'

1. REPETITIVE DNA

A great deal of DNA in eukaryotes consists of repetitive nucleotide sequences. Some is transcribed into RNA, but very little is translated into protein. Repetitive DNA can contribute up to half of total genomic DNA. In general, is restricted to eukaryotes. Two broad categories have been defined, termed interspersed repetitive DNA and highly repetitive DNA respectively.

Interspersed Repetitive DNA

Background

This class of repetitive DNA consists of multiple copies of variable length nucleotide sequences is widely dispersed throughout the genome, largely outside genetic sequences, and constitutes up to as three quarters of total repetitive DNA. Its periodicity, sequence length, and pattern of dispersal vary between different species and with different repetitive sequence types, but the amount and copy number of any particular family usually remains fairly constant within any one species. It is organized largely as single copies of the repeat sequence spaced at fairly regular intervals between coding stretches of genetic DNA, but may occur in long tandem arrays of repeats clustered at particular genomic sites, similar to the organization of highly repetitive DNA discussed below. Different species, of drosophila for example, contain different copy numbers of repeat ‘families’ or consensus sequences, and even different strains among the same species can have different members of these repeats located in quite different chromosomal positions, as if they wander about in an almost ‘nomadic’ way. Nonetheless, if we take this ‘consensus’ view of these repeats, the number of different families within species reduces to the relative few, for example eight in drosophila, and one dominant family in both rodent and primate species viz., the Alu 1 family. It is also possible to discern similarities between the consensus sequence families of different species, as in the close homology between the dispersed Alu 1 repeats of rodents and primates.

Despite the above, it has to be said that this grouping of dispersed repetitive DNA into families of common consensus sequence often hides an enormous amount of sequence variability within any ‘family’ unit. This even includes a variability between different individuals of the same species. Moreover, much of this variation is of a non-random type, with some sections of the Alu 1 consensus sequence in man, for example, apparently being much freer to vary than others over its different genomic locations.

The origin of interspersed repetitive DNA
In the first edition of this book, I spent much time and space discussing this, and concluded that it could be explained as being of viral origin. This arose partly from the finding that differences between sequences, for example the Alu 1 consensus sequences of primates and rodents, seemed often related more to insertional events than random point-mutational change. It also became increasingly clear that a good deal this repeat DNA was very mobile within the genome, in keeping with a ‘viral’ view.

It is now evident that most interspersed repeat DNA is indeed derived from viruses or viral elements. The most prominent among these is the retroviral group of viruses and their subunit mobile elements. The human integrated endogenous retrovirus group is the classic case, with integration being either of complete sequences as so-called HERVs (mostly replication-defective), or of their subunits - the LTR retrotransposons, the non-LTR retrotransposons (L1 elements), and the short Alu1 elements. Kazazian has summarized these in detail, not only in regard to their ability for internal retrotransposition and mobility around individual genomes, but also the prospect they hold for horizontal transfer as potential ‘drivers of genome evolution’. Accepting this latter point, others have taken the necessary further step of attempting to explain the means by which such horizontal genetic transfer from one individual or species to another might be realized. Gilbert and colleagues have proposed that parasites might be the vehicle of transmission, and Schaack and colleagues have extended this further by suggesting that transposon DNA could be directly transferred by parasites via exchanges of blood and saliva during feeding. However, it is one thing to transfer DNA in secretions, and quite another to have it taken up and dispersed around the entire cellular DNA of a new host, including the germ line. A common salt lick hardly seems adequate to such a task.

My own view, already discussed in chapter 9, is that viruses mediate such horizontal transfer. In the case of subunits of retroviral elements, this might well be done by the parent retrovirus itself, but when these eventually become replication-defective other viruses can carry out the task equally well – for example, the adenoviruses.

Highly Repetitive DNA

In contradistinction to interspersed repetitive DNA, this repeat DNA class is highly clustered at localized sites within the genome, in long tandem arrays. Highly repetitive DNA represents a large proportion of the total, sometimes over 50%. The basic repeating sequence is often extremely simple, sometimes only a few base pairs long. But within each clustered array there is usually a longer-range periodicity, sometimes extending over several thousand base pairs, and consisting of many different versions of the same fundamental short repeating ‘consensus’ unit. One surprising finding in this respect is that there is often a much closer degree of hybridization matching between these longer-period repeats than between the different versions of the short unit of which they are composed. This has turned out to be due to the fact that different versions of the short repeating
Satellite DNA

This consists of very large arrays of short base pair nucleotides arranged as thousands, even millions, of tandemly repeating, non-coding DNA. It is the main component of centromeres, and the main structural constituent of heterochromatin. Satellite DNA is extremely variable, not only in nucleotide sequence, but also in the repetition number and location of family members within the genome - even within individual species. If we accept a ‘consensus’ concept of the resulting tandem arrays as sequence variation on a common theme, the number of satellite ‘families’ within any one species can usually be reduced to a relative few. However, as with interspersed repeat DNA, this should not be allowed to obscure the fact that within each family there are characteristically many different sequence versions both within and between species. Moreover, as with other highly repetitive DNA, these different versions are not infrequently organized in the longer units in a rather precise and characteristically nonrandom way.

Satellite DNA distribution within and between species is complex. Even closely related species can show enormous differences in their satellite DNA. An extreme case exists in the kangaroo rat species, where D. deserti appears to carry none of the three satellites which make up more than 50% of the closely related D. ordii species. On the other hand, consensus sequences can be very similar in the satellites of distant species - even as far apart as (kangaroo rat and guinea pig, or drosophila and man. Importantly, many such sequences often seem to disappear at some stages of evolution only to reappear later at others as the species lines are followed over evolutionary time. Even more perplexing is that different versions of the same consensus sequence within the tandem arrays of any one satellite can show enormous variation in the repeat sequence, and with a variation that can be highly non-random. I will now look at the generation of these sequences a little more closely.

In an attempt to explain the variation of sequence type, as well as amount, within satellite DNA between species, put forward the suggestion that related species, whilst often appearing to carry entirely different satellites, each actually contain all of the different satellites characteristic of the species group, as if sharing them all from a common ‘library’ of sequences, but in vastly differing amounts. According to this view, differences in satellites between related species would be more quantitative than qualitative, and be due to particular sequences being ‘amplified’ to abundance in some species and ‘contacted’ to below the limit of detection in others. Dover and Coen have suggested a similar ‘spring-cleaning’ process for amplification and contraction, drawing parallels from dispersed sequences in ribosomal DNA. They considered that the initial step was one of gene
‘conversion’ where new sequences were exchanged for old. Drawing on parallels between interspersed repeats and transposable elements, they postulated further, that repeated transposition of new sequences led to its subsequent spread throughout the genome, in a process they termed ‘concerted evolution’.

Dover and colleagues’ concept of the mechanism of change in repetitive DNA between species has also been useful, particularly as it incorporates the (retro-) transposon nature of many interspersed repeats. However, it is inadequate alone to account for all aspects of variation within interspersed repeat DNA for several reasons. First, though it may give an explanation of how new sequences might be amplified and spread throughout the genome, it does not explain at all how pre-existing sequences are removed in the way they imply. Second, transposition would normally be expected to produce identical copies of the transposed units at each of the different genomic sites or, at the very most, randomly varying ones. And though such uniformity may be seen among, say, the non-transcriber repeats (NTS) of ribosomal spacer DNA, it is not the case for most satellite DNA. Third, the NTS sequences themselves are interrupted by multiple insertion elements that seem to be totally impervious to the proposed conversion process.

Despite all this, there is no doubt that the analogy Dover and Coen have drawn between the ‘concerted’ nature of the evolutionary processes which lead to the replacement of interspersed repetitive DNA, and ‘spring-cleaning’, does capture the essence of the dramatic genomic change underlying the variation in this repetitive DNA class. In addition, something akin to transposition would undoubtedly be a good explanation for how the pattern of dispersal of any given consensus sequence family may vary markedly both between related species, and between different strains among the same species — almost, in the case of some long-period interspersed repetitive DNA, as if they were indeed ‘nomadic’. What is needed is to find a process that would yield similar transposition events between individuals and species, so that library sequences could be borrowed across a larger scale. If we could understand that, it might even be possible to develop a unifying concept for repetitive DNA in general, because the two repetitive DNA classes are often closely intermingled.

A Viral Theory of repetitive DNA Origin and Change

I suggest here that most of the puzzles of satellite and other repetitive DNA could be explained if they arose and evolved by viral DNA sequence uptake and interchange between viruses and host. Larger partial sequences in repetitive DNA could arise either from viral transduction — where imprecise excision allows sequences from the viral integration site to be exchanged between virus and host — or from internal viral sequence exchange, either by direct transfer of sub-unit mobile elements from viruses, or by double crossover recombinational exchange between corresponding segments on viral mobile element and host. This could explain how longer repetitive units such as LTR/non-LTR retrotransposons and Alu1 segments are can be found incorporated into highly
It would account also for the non-random ‘patchwork’ nature of much satellite and other DNA build-up. Sometimes such accumulation appears in ‘transposition bursts’ as an extremely rapid process in evolutionary terms.

The heterochromatin location of satellites may mean that any mobile element insertion there could be relatively innocuous. This has two consequences. First is could be a convenient launching ground for genetic experimentation, in that transferred sequences could subsequently be moved elsewhere in the genome ‘on approval’ and even amplified in the process. Second, mobile element insertion and excision generates short base pair repeats and, in recurrent bouts, this could account for much of the short highly repetitive base pair sequences that constitute most satellite DNA. Different mobile elements can have different sites of preferred DNA insertion and excision, as in ‘transposition memory,’ and this could lead to variations in short base pair repeats with each new episode and hence to the apparently non-random consensus sequence build-up of much satellite and other repetitive DNA.

Such a theory could account for a great deal of the behavior of repetitive DNA in general. The existence of the various sequences of repetitive DNA in multiple copy number could now be related to the recurrence of viral infection with each new host generation over the long evolutionary time-period the virus remained in an infective relationship with the host species. Differences in repetitive DNA location from one individual to another among a species could be explained by the occurrence of different viral integration sites with each new host-infection. In addition, variability of repetitive DNA, including satellite consensus sequence type, might well be due to molecular evolution within the virus itself, and as such it would not be surprising if, being determined by selective mechanisms, it were of a non-random nature. On the other hand, the absence of particular repeat sequences from one of a group of related species could now be explained by the failure of a particular virus to interact with the species concerned, either because of the lack of an appropriate receptor for the virus, or because that species was not exposed to the virus by virtue of its being located in some different geographic ecosphere distribution. Occasionally, the absent sequences might have been deleted by homologous recombination via direct repeats at the ends of some satellite sequence block. This might occur, for example, when two insertion sequences, known to exist in at least some satellites, happened to become incorporated into satellite DNA in such a way as to flank a stretch of its oligonucleotide tandem arrays. The whole block of intervening satellite could then be removed, either by its transposition into some virus resident at the time, or by the formation of (circular) extrachromosomal DNA with subsequent gradual deletion from the progeny line of descent.

Generally speaking, we would expect that related species, particularly those evolving within the same geographic area, would interact with similar viruses, so it would not be surprising if the build-up of their repetitive DNA were analogous to a process of borrowing sequences from some common
library,’ as Fry and Salser proposed. But, according to the present viral view, it would no longer be necessary to postulate that all members of any related group of species contained all of the library sequences because interactions with different viral forms might well produce different types of repeatedly-integrated sequences. In addition, whilst the sharing of similar sequences between distantly related species might sometimes indeed be due to a long evolutionary history of sequence conservation through traditional evolutionary mechanisms, on the present basis it might also reflect the two species merely having belonged to the same general host/virus ecosphere at some stage of their evolution, and not necessarily a distant one in evolutionary time. Even in the case where the original sequence did arise remotely in time, subsequent and very real conservation over long evolutionary periods need not reflect any importance that sequence had for the host species itself, because it could relate to the continuance of some important viral function, or to a function in some other host species with which the virus was in separate genetic contact. Because this would mean that sequence conservation might often be due to an existence that sequence had entirely outside the species under consideration, it could also explain the paradox of how some repeated DNA sequences come to be deleted after long evolutionary conservation with relative impunity.

Most of the longer sequences incriminated here in the development and evolution of repetitive DNA are segments of retroviruses. Apart from HIV viruses, these are no longer as active as they once were. “So”, it might be asked, “Where are the vehicles for continuing repeat sequence mobile element transfer and interchange between individuals and species in the manner proposed?” In this respect, a number of DNA viruses have been shown capable of taking up into host repetitive DNA elements. This includes the Epstein-Barr virus, the papovavirus, a mutant adenovirus SV40 virus, and the SV40 virus. Indeed, Rosenberg and colleagues have observed that when SV40 is grown on monkey kidney cells, there is repeated incorporation of primate alpha satellite DNA sequences into the virus such that viral variants eventually arise consisting mostly of satellite DNA. Of course, these are examples of incorporation of host sequences into virus rather than the reverse, but there is evidence for the opposite process as well. Thus, primate repetitive DNA has now been found that contains sequences of close homology with the replication origins of the SV40 virus and the paparvoviruses. Even replication-defective adenoviruses can be a vehicle for the integration of retroviruses into human cell lines, so that the process for continuing foreign mobile element uptake and integration as repetitive DNA might sometimes hinge on something as simple as a common cold. Come to that, it seems that simple RNA viruses can occasionally be reverse transcribed by cellular reverse transcriptases, so broadening the potential range of intercommunication between viruses and higher life forms.

2. INTERVENING SEQUENCES IN THE EUKARYOTIC GENOME
The second surprise in the field of molecular biology during modern times has been the finding that individual genes are broken up into pieces called ‘exons’, each of which codes for a particular domain or segment of the total protein molecule, and each one of which is separated from the next by non-translated intervening sequences called introns. The process of intron removal is referred to as ‘gene-splicing,’ where the sub-unit gene ‘exons’ are joined together to form the final messenger RNA. Introns are located in a wide range of genes, including those generating proteins, ribosomal RNA (rRNA), and transfer RNA (tRNA), and moreover even in species as primitive as bacteriophages. At the time of their discovery, none of this made any sense. Many aspects have since become clearer.

The mechanism of gene-splicing differs with different introns, and interestingly RNA enzyme catalysis plays a role in them all.

The most primitive are the group I self-splicing ribozymal introns - i.e. with RNA enzymes rather than protein - that catalyze their own excision (as simple circular RNA) from mRNA, tRNA and rRNA of relatively simple organisms such as bacteria, protozoa, and lower eukaryotes. The intron ends are brought together preparatory to splicing by base pairing of internal guide RNA sequences with the downstream exon.

Group II introns occur in somewhat higher life forms. The yeast mitochondrion provides the classic example. Here, the intron is excised in two stages, the first being formation of an internal loop, and the second the excision of the whole intron, but now as a lariat structure. Catalysis is again directed by RNA ribozymes. A conserved sequence UACAACC box within the intron specifies the branch point for lariat formation via U2 complementary base pairing binding. (Reviewed by Robertson)

Pre-messenger nuclear RNA excision extends this mechanism. Here, the intron is again first folded into a lariat structure using sequence complementarity, but the actual splicing is done by a more complex ribonucleoprotein or ‘spliceosome.’ Lariat removal involves U1, U2, U4, U5 and U6 short RNA sequences. U1 binds at the U5 junction and U2 binds at a consensus sequence of about 30 nucleotides from the 3’ junction where ‘U2AF’ binds. In the second step, U1 and U2 come together with U4, 5 and 6 and the free lariat is generated, with U6 being particularly important for final intron excision. This has all been reviewed well elsewhere

A Viral Theory of Intron Origin

I suggest that the whole process of gene-splicing and intron origin could be explained if it reflects a long history of life evolving as the gradual build up of viruses, as I envisaged at the outset, even before introns were described. In an extension of this initial view, such introns would be seen as being established on the basis of RNA viroid/viral inserts into host in earliest RNA world, followed by the much later reverse-transcription of similar inserts into DNA with the emergence of the reverse
transcription and the retroviruses. On this basis, the finding of ‘genes in pieces’\textsuperscript{68} would reflect the essence of the way life has evolved over the eons - by the ad hoc ‘experimental’ shuffling of genetic modules, but a shuffling now based on non-random genetic change due to prior evolution within the viral modules themselves. And the increasing evidence is that introns exist even in the most primitive life forms, including bacteria,\textsuperscript{69-71} viruses,\textsuperscript{59} and even simple phages,\textsuperscript{58,72} so this kind of basis for evolutionary progress is well established.

In respect to early life forms, some of the nucleotide sequences in simple viroids bear a striking resemblance to primitive group I introns,\textsuperscript{73,74} particularly at intron ends.\textsuperscript{75,76} These group I introns are removed as circular structures very reminiscent of viroids, and often behave as mobile genetic elements.\textsuperscript{77,78} They encode sequence specific DNA endonucleases, and so are quite capable of transposing to other genomic sites via direct DNA transfer.\textsuperscript{77,79}

Higher level (group II) introns are able to reverse RNA-splice themselves into other RNA,\textsuperscript{80,81} and they display a remarkable similarity to retroviral reverse transcriptases,\textsuperscript{82,83} even to the extent that they can be regarded as ‘infectious.’\textsuperscript{84}

Things get rather more complicated higher up the evolutionary scale, but there are strong indications here, too, that introns retain a capacity for mobility. It is clear, for example, that many nuclear messenger RNA introns contain sequences of close homology with DNA repetitive transposable elements\textsuperscript{85} such as Alu\textsuperscript{1,86,87} and even retroviruses.\textsuperscript{88}

**Evolutionary origin of introns**

There are two opposing views on this: ‘Introns early’, and ‘introns late.’(See Sharp\textsuperscript{89}) Needless to say, the present thesis is more closely aligned with the ‘introns early’ view. And there is increasing evidence this is the case.\textsuperscript{70,72,78,90}

 Particularly because they can be so mobile, introns might be seen as being potentially disruptive to the genome, and therefore to the species. However there are probably ways of limiting this,\textsuperscript{14,91} and on the other hand they offer important evolutionary advantages because of being ‘silent’ host genetic areas into which foreign DNA can harmlessly be inserted via transposition and retrotransposition. And after such incorporation, that DNA could itself subsequently be transposed, almost ‘on approval,’ to other sites on the host genome without overwhelming it randomly with new DNA information at any one time.

Such a process could also facilitate the entry and incorporation of sophisticated genetic material from outside in a non-random way, as discussed in Ch. 9. Of course, in some individuals, such
experimental inserts, and the genetic shuffling arising from them, might well be disadvantageous, and because of this many would no doubt be deleted, or suppressed, over evolutionary time. But as I see it, this would be more than offset in evolutionary terms by the continuingly offering an evolutionary advantage to both virus and host species.

Nature seems to me to be essentially opportunistic, and as such, it would be surprising indeed if the various species did not take full advantage of any genetic information available to them from other individuals and species, particularly that which had already been ‘processed’ in some host/virus ecosphere elsewhere. Genetic ‘shuffling’ along these lines could offer enormous mutual benefit to both host and viral species. In recent years, it has become increasingly recognized that DNA within the eukaryotic genome has a high degree of mobility, and in some ways all I am suggesting here is that this mobility may extend far beyond the individual.

It will be noted that the present view puts a marked emphasis on a co-operative aspect to Nature during evolution. This is not to say, however, that the struggle for survival does not occur at the level of the individual in the way Darwin envisaged. Indeed, it clearly does, but that does not deny, either, the importance the present mechanism offers to the evolution of species as a whole, by the interchange of pre-processed genetic material ‘on approval,’ as the basis for non-random mutational change. In this context, co-operation between organisms in evolution seems to me to have been given far too little credence in evolutionary mechanisms till now.

3. NON-CODING RNA REGULATION OF GENE EXPRESSION

The final aspect of recent interest concerning non-coding DNA is its importance in the regulation of gene expression.

Although some introns are now known to be translated into enzymes important in the processing of other introns within the same or different genes, the function of the vast bulk of intervening sequences and repetitive DNA remains obscure. In this respect, it is interesting that Britten and Davidson suggested many years ago that dispersed repetitive DNA sequences within the eukaryotic genome might well be important in gene regulation and, compatible with their view, they subsequently showed that the initial mRNA transcripts arising from these DNA repeats were often expressed in different ways, not only at different times during evolution, but during different stages of individual development, even within different tissues from the same individual. Now, that finding has turned out to be most perceptive. Since that time, many discoveries about non-coding RNA have been made, and is now clear that it has a vital role in regulation of gene expression, even within different tissues in just the way Davidson and Britten envisaged.
Epigenetics

Finally, a word on epigenetics. Because non-coding DNA is now known to have important functions in regulating DNA expression, there is a tendency refer to it as epigenetic. Now, the influence of non-coding DNA/RNA is certainly outside the exome (exon component of the genome), but it is still part of the DNA we inherit. The problem is that term epigenetic means different things to different investigators, and it sometimes carries more than a hint of neo-Lamarckism. Now, that may sometimes be the case, in the way Steele originally envisaged. But there is no reason to think, de novo, that the environment is any more likely to influence non-coding RNA than the exomic coding DNA. Even when it does, we should not assume that such influence will be transmitted through the germ line. Despite clear examples of epigenetic inheritance, we need to be careful not to extend this concept without due rigor. The same holds true for the effects of gene and histone methylation.
Our discussion on evolutionary mechanisms in the latter chapters of this book has emphasized the importance of viral infection, virus/host genetic exchange, and genomic DNA mobility in general, as a driving force for evolution. However, I have alluded several times to the fact that the genomes of most species, includes their integrated viral material, are normally fairly stable, unless and until the cell DNA is damaged or disrupted in some way. Then, these viral elements can become activated to promote genetic immobility and disseminate overt viral disease. In lower organisms, the stress leading to this activated state is normally thought to be in the nature of physical damage. However, in higher organisms, endowed with a vascular system, there is also the possibility that cell damage may be induced by ischemia, and from our earlier theory we can recognize that this could, in turn, be readily brought about by stress-induced activation of the sympathetic nervous system.

This extended view therefore allows us to complete the circle to a unified hypothesis of the importance of stress-induced ischemia, namely that it may not only cause disease but, by activating (germ-line) endogenous viral material, may provide a strong driving force underlying evolution itself.

So although individuals may be the worse for endogenous viral activation, the species to which they belong may be the better for it in terms of their evolution. The situation may perhaps be likened to battle, where the invaders, under the provocation of stress, may conquer and defeat individuals. Then, though initially hostile as prisoners, they may eventually become reconciled with, and integrated into, the new community to add their own strengths and talents. In the same way, it may well be that despite their propensity to maim and kill individuals, endogenous viral materials, activated from their genomic reservoirs by stress of whatever sort, may eventually offer long term evolutionary advantage to species concerned. Perhaps, therefore, immunologists may not be the only ones to have their Generators Of Diversity, and that just as no man is thoroughly bad, even the Damaging Effects of Viral Infection on Life can on occasions be Generators Of Orderly Diversity.

Finally, if viruses are part of our being, we cannot, and perhaps for the sake of our evolution, should not, hope to prevent all viral infection and viral-related disease. But we should at least be able to limit their adverse effects by preventing that cell damage from stress and other causes which facilitates the re-emergence of their normally suppressed and potentially lethal viral effects.
REFERENCES


2. Britten RJ, Kohne DE. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science*. 1968; **161**:529-40.


56. McCutchan TF, Singer MF. DNA sequences similar to those around the simian virus 40 origin of replication are present in the monkey genome. Proc Natl Acad Sci U S A. 1981; 78:95-9.


