Modular Transposition and the Dynamical Structure of Eukaryote Regulatory Evolution

Chris C. King,
Department of Mathematics and Statistics,
University of Auckland.

Abstract : This paper examines a model in which transposable elements provide a modular architecture for the cellular genome, complemented by cellular recombinational transformations, arising in turn as a dynamical consequence of this modular structure. It is proposed that the ecology of transposable elements in a given organism is a function of recombinational protocols of the evolving cellular genome. In mammals this is proposed to involve coordinated meiosis-phased activation of LINEs, SINEs and retrogenes complemented by endogenous retroviral transfer between cells.

1 : Introduction

The theme of this paper is that two forms of feedback between mobile and cellular protocols lead to a complex ecological relationship between the forms of transposable element and recombination processes in the cellular genome. The first of these is the feedback between phenotypic and element survival, which may require for example negative regulation of element transposition. The second is the feedback cellular recombination processes have on the transposable element ecology, as a result of both element and organism adaption and survival.

It is proposed that the evolution of major phylla is accompanied by particular structural relationships between the resident mobile families, constituting a *stochastic dynamical system*, which is characteristic both of the individual protocols of the elements and of cellular recombinational and structural features. In particular in mammals, a global linkage is proposed between five types of transposable structure, LINEs, SINEs, retrogenes, endogenous LTR-retroelements and cellular conversion processes, which constitutes a stochastic process phased to the reproduction cycle, providing a higher-level event space structure, enhancing survival through adaptability.

2 : Phylogenetic Histories and Species Ecologies

The work of Xiong and Eickbush (1990) demonstrates that the phylogenetic history of reverse transcriptase-containing (RT) retro-elements and viruses has a common ancestry probably running back to the eukaryote-prokaryote divergence and the RNA era, just as has been proposed for introns.

At least three groups of retroelement have developed habitats spanning the major kingdoms of biota : The (+)-strand RNA viruses (group A) have already been recognized as representing a ubiquitous habitat spanning bacteria, vertebrates and higher plants, of likely ancient origin.

The retroposons (group B) similarly span the mammalian L1 family, Cin 4 of maize, several elements from protozoa to insects, the ms-RT from bacteria such as *Myxobacteria* and *E. coli* (Inouye & Inouye 1991, Temin 1989) and the RT from yeast mitochondrial group 2 introns and green algal plastids. The very wide distribution of this type, combined with its putative confinement within the organism lead to the conclusion that it is a ubiquitous component of major phylla which may fulfil some regulatory or evolutionary function, in addition to element protocols. Notably the retroposon gene structure is consistent with the putative ancestor of the entire tree.

The LTR-retroelements (group C) again span eukaryotes from the slime mould, through yeast, gymnosperms, mono and dicotyledons, insects, and vertebrates, however the radiation now appears to encompass two major groups. One of these, represented by the *Copia*-like elements, spans yeast, insects and flowering plants, the other comprising the *Gypsy* group of retrotransposons and the retroviruses spans the slime mold, yeast, plants, insects and vertebrates. From this point of view it is certainly true that retroviruses have evolved from transposable elements (Temin 1980), but the phylogenetic tree indicates development of the *env* gene may have been the result of particular selection for the higher vertebrate habitat, possibly involving species or tissue-specific expression, and suggests the lack of an *env* gene may not prevent at least occasional horizontal transfer in other capsid-forming LTR retrotransposons, particularly since (+)-strand RNA viruses require only a capsid which would naturally be coded by the *gag* gene. A possible *env*-like sequence in *Drosophila* 17.6 should be noted here.

The relations between host species in the case of LTR retrotransposons, and the occurrence of two distinct groups of horizontally infectious virus within group C, both support the idea that the group has been subject to extensive inter-species transfer, despite many elements predominantly being transmitted endogenously down the germ line. Thus although retroviruses evolved from transposable elements, capsid-forming retrotransposons may themselves have a capacity for occasional horizontal transfer. Development of the retroviral *env* gene may thus represent a specific adaption to the vertebrate host, rather than the emergence of horizontally infectious habitat.

The conservation of retroposon and LTR-retroelement form through major evolutionary phylla supports a mutual evolutionary relationship between phenotypic adaption of the host organism and *complementary modes of mutational adaptability* possessed by groups B and C.

While the history of RTs demonstrates conservation generic classes of element, each spread broadly across the eukaryote phylla, examination of the diversity of elements in various species illustrates organismic conservation of diverse transposable element ecologies.

Dictyostellium harbours an LTR-retrotransposon which involves developmental and heat shock regulation, with unusual inverted terminal repeats (Rosen et. al. 1983, Zuker et. al. 1983, Firtel 1989), in ~ 40 copies, as well as two retroposon-like elements. *Zea mays* contains two retroposons in ~10 copies, a *Copia*-like LTR-retrotransposon sharing homology with species as distant as bamboo, as well as a variety of DNA transposons Ac-Ds, Spm etc. (Berg & Howe 1989, Schwartz-Sommer 1987, Johns et. al. 1985). *Drosophila* harbours two forms of LTR retrotransposon typified by *Copia* and *Gypsy*, retroposons including F, G, *Doc*, *Jockey* and the I dysgenesis element (Berg & Howe 1989), and a variety DNA-transposons including the P hybrid dysgenesis element, and the foldback FB transposons. Significantly all of these elements come in a comparable copy number of between 50 and 200.

By contrast, the vertebrates and particularly *Mammalia* present a unique spectrum of novel retroelements, and appear to be deficient in the DNA transposons found in insects and plants. The LTR-retroelements have evolved into retroviruses specialized for cell to cell transfer through the *env* gene, with a diversified habitat of endogenous and exogenous forms. Although they exist in only about 10^2 copies, the many cryptic families and solo-LTRs may account for 1% of the genome. The retroposons have developed into very high copy number L1 LINEs. These are complemented by two new types, the SINEs, which appear to have distinct origins as pseudogenes of small nuclear RNAs and occur in uniquely high copy numbers, and processed retrogenes and pseudogenes, which may assume predominance in the cellular genome according to a possible *Vesuvian* scenario.

The lack of obvious evidence for DNA-transposons in mammals suggests that some of their functions in evolution have been replaced by a process in which the active retrotransposition of LINEs, possibly in the context of a cellular function such as gap repair in the reproductive cycle has given rise to new modular structures in the SINEs and pseudogenes. In particular the capacity of SINEs and LINES to form sites for conversion, translocation and inversion parallels that of FB and Ac-Ds elements, and provides new modes of coordinated regulatory interaction, such as pol III generated sense or anti-sense RNAs and alternative poly-A sites.

3 : Modularity and Transposable Elements in Eukaryote Mutations

Investigation of spontaneous sequence rearrangements in different species gives a window on classes of recombination occurring in different eukaryote genomes. Although this provides only a very restricted time view of a very long-term process, it does throw up several hints of the action of transpositional and recombinational processes.

The picture in mammals is illustrated by studies of several human an murine mutational pathologies (Meuth 1989). In some mammalian genes, point mutations outnumber the other most frequent feature, deletions, possibly associated with the nuclear scaffold and topoisomerase action. In others large rearrangements predominate. A significant class of rearrangements is crossing over involving repeated sequences, most frequently an Alu at one or both ends, or palindromic stem-loop structures. Others are retroviral insertions and anomalous immuno-recombinational events, involving switch regions. Germ-line mutations are generally more varied and complex than somatic ones, possibly because they include inter-chromosome events from crossing-over.

In *Drosophila melanogaster* , discounting the diverse modulated mutations caused by dysgenesis elements such as P and I, and FB and its tandem forms, a majority spontaneous mutations are insertions of transposable elements. Although *Copia* is more abundant in RNA transcripts, *Gypsy* predominates in spontaneous transposition (Green 1988). A variety of mutations including the dilute locus (Jenkins et. al. 1981, Copeland et. al. 1983a) and the *agouti* locus (Copeland et. al. 1983b) in the mouse display transpositional insertions with accompanying suppressor action by other genes and formation of solo LTRs. The yellow locus in *Drosophila* (Geyer et. al. 1988) illustrates selective developmental alterations regulation through insertion and consequent release by solo LTR formation.

One of the most significant studies of the potential of transpositional mutation comes from a *Gypsy* insert into the *ct* locus which resulted in *concerted transposition* (Tchurikov et. al. 1988) possibly involving both DNA-transposition and retrotransposition giving rise to *transposition explosions* - significant movements of *Gypsy*, other retrotransposons and possibly also P and FB elements, but without obvious chromosome aberrations. Because of the identical arrangement of these in all mutant offspring, they occur in a single germ cell at premeiotic stage. Repeated reversions of transposable elements back to deleted positions occurred, presumably by reinsertion into the remaining LTR. These included revertions to wild-type of deletions of single-copy gene sequences. There were also multiple insertions of other mobile elements such as *jockey*, *hercules*, *burdock* and *roo* into the 5' LTR of *Gypsy*. Several multiple transpositions and deletions were repeated in a significant proportion of the cases indicating a controlled explosion.

This leads to the hypothesis that under conditions of genomic shock (McClintock 1978, 1982), or cellular processes connected for example with meiosis, concerted transposition of diverse types of element spanning both the retroelements and DNA transposons may occur, resulting in significant changes in genomic organization. These

may precipitate species discontinuities, along with processes such as *hybrid dysgenesis.*

Central to these ideas is the contrast between *microevolution* involving selective optimization of phenotype, maintaining species optimality and graduated variety, and *macroevolution* arising from genomic catastrophe, resulting in discontinuous changes in species, caused by major changes to the environment or niche of the organism. Existing organisms have survived, both because they have been optimized through phenotypic selection, and because they have sustained extensive transformations of structure in times of evolutionary disruption. Negative regulation of transposable elements reduces mutational load in times of optimization but provides structurally significant modes of mutation through coordinated transposition under situations where phenotypic optimization is breaking down. Since the principal challenges to the long-term survival of a given line occur during the disruptive phase, traditional models of Darwinian selective advantage or neutral evolution may be a relatively secondary feature of the evolutionary development of complexity, providing optimizing stability between catastrophic discontinuities - Waddington's *chreode*. Such discontinuities may be responsible for the great radiative adaptions such as the Cambrian and mammalian radiations.

4 : Evolutionary Features of Retrogenes and Retropseudogenes.

Discounting the pseudogenes generated by cellular gene duplication, which generally have introns and are chromosomally linked to the original, three major classes of genes appear to be derived all or in part through reverse transcription. The first of these consists of pseudogenes of small nuclear U RNAs participating in intron splicing (Bernstein et. al. 1983). tRNAs and 7sRNA s also have multiple pseudogenes, modified forms of which constitute SINEs (Weiner et. al. 1986, Sharp 1983).

The second is typified by the well-known classes of retroviral oncogenes. These can arise either by the insertion of a retroviral element into a cellular gene or by incorporation of a cellular gene into a retroviral element through recombination. Several functional genes display regulation by an upstream retrovirus. The human amylase family comprises five genes, two expressed in pancreas and three in salivary gland (Samuelson et. al. 1990). Evolution of the family shows a complex history including interaction between an inserted retroviral LTR enhancer and a cryptic promoter in an inserted actin pseudogene, as well as deletions and duplications. Mouse complement C4 has a homologous partner, androgen-sensitive sex-limited protein *Slp*, (Stavenhagen & Robins 1988) containing an upstream insertion of an ancient endogenous provirus.

The third type of structure the *processed retrogene* is very common in mammals but uncommon in other phylla (Vanin 1984), consistent with evolutionary studies suggesting their formation since the mammalian radiation. Most lack introns, and have homology to the cellular gene coincident with the transcribed mRNA. Generally there is an oligo-A tract at the 3' end from 11 to 38 nucleotides, or in variant cases $(NA_m)_n$ tract, and they are often flanked by short 9-14 bp target-site repeats consistent with insertion at a nick or staggered break. They often lack the original pol II promoters 5' to the transcription start, although some appear to be transcribed from upstream promoters including pol III sites. Frequently, but not always, these have interruptions in their open reading frame caused by base changes and multiple deletions and insertions.

Many examples occur of processed retrogenes with new or altered cellular function. The human tropomyosin processed gene (MacLeod & Talbot 1983) and mouse zinc-finger Zfa retrogene (Ashworth et. al. 1990) both show functional changes consistent with novel function. Notably the Zfa retrogene has arisen from the X-linked Zfx rather than the Y-linked Zfy 1,2 although its loss of the third finger shared by Zfy2 may signal testis-determining function.

Several gene families consist of a great number of retropseudogenes, suggesting that mammalian genomes may be literally flooded with retropseudogenes in various states of divergence. This *Vesuvian* model (Graur et. al. 1989) combined with the slow rate of loss is consistent with a majority of mammalian genomes consisting of processed pseudogenes in various stages of *compositional assimilation.*

An indication of the scope of large retrogene families is given by the following four examples. There are 300 glyceraldehyde-3-phosphate pseudogenes in rat a heavily transcribed housekeeping gene. The human non-histone chromosomal protein HMG-17 (Srikantha et. al. 1987) and β-tubulin families both have several tens of genes most of which are retrogenes. Mammals contain 20-30 mitochondrial cytochrome-c sequences, most of which appear to be pseudogenes (Limbach & Wu 1985) while other animals have only 1 or 2. The mouse has two functional genes, differing by mutations in the heme crevice, one of which is general and the other testis-specific. Surprisingly, it is the general gene which has given rise to the three pseudogenes studied, which appear to represent three different transcripts, again consistent with oogenic rather than spermatogenic expression.

It is still not resolved to what extent retrogenes may be inserted into the germ line by retroviral transfer. The mouse $\psi\alpha_3$, $\psi\alpha_4$ globin (Vanin 1984) and human $\lambda\psi_1$ (Hollis et. al. 1982) and Cε pseudogenes (Ueda et. al. 1982), which have common $(NA_m)_n$ ends (Rogers 1985), are not normally expressed in the germ line. These could be generated either by aberrant transcription e.g. by an Alu pol III promoter, or by the RNAs being carried to the germ-line from somatic cells packaged along with retroviral RNAs as is known to be the case for mouse VL30 genes.

The $\lambda\psi_1$ gene has J and C regions joined which occurs during splicing, but does not have a V region to the left, which is normally generated by somatic recombination during development. Similarly the $C_{\varepsilon 3}$ gene has four C region exons but none of the V, D, or J regions joined as in mature mRNAs, consistent with aberrant germ-line transcription. On the other hand $\psi\alpha_3$ and $C_{\varepsilon 3}$ are flanked by retroviral LTR-like sequences although the former appears to be an accidental association, as the intracisternal A particle LTRs are in the wrong orientation and may be older (Lueders et. al. 1982). The $C_{\varepsilon 3}$ pseudogene is particularly interesting because it is neatly flanked by two functional LTRs and may encode a distinct novel protein. However Alu-like pol III promoters have also been found at the 5' end of these genes.

The many K^+ channel and G protein-linked receptor genes have no introns suggesting they may arise from retrogene families. Large families of genes with differential variation of structure would be particularly good candidates for diversification through retrotransposition.
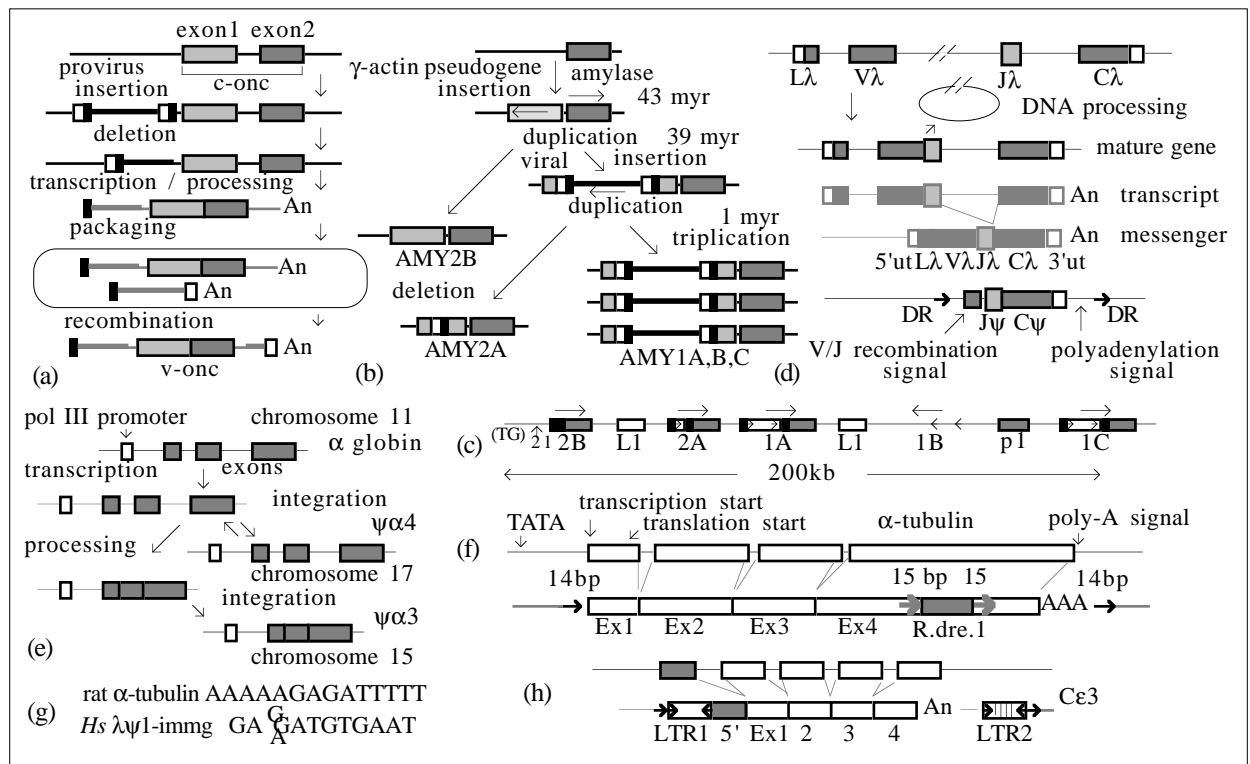

Fig 1 : Diversity of retrogene structures and composite evolutionary events: (a) Generation of viral genes through retroviral recombination (Varmus 1982). (b) Evolution of the human amylase family showing inserted γ-actin retropseudogene, retroviral and LTR regulators (Samuelson et. al. 1990). (c) The amylase gene region showing the compound genes, a pseudogene and L1 elements. (d) λ-immunoglobulin processing and its relation to the pseudogene λψ1. (e) A possible tree for the rat α-globin pseudogenes (Vanin 1984). (f) A typical processed retropseudogene, rat α-tubulin, showing bounding direct repeats, 3' poly-A, removed introns and an inserted SINE (Lemischka & Sharp 1982). (g) Target site repeats. (h) The $C_{\varepsilon 3}$ pseudogene with LTRs, 5' region and exons flanked by short direct repeats (Ueda et. al. 1982).

5 : Coordinated Mobilization of LINEs, SINEs and Processed Retrogenes.

Several key features of LINEs, SINEs and processed retrogenes suggest they have a common mode of activation. A few key features of each group are summarized to highlight their potential common modes of activation as well as individual differences.

LINEs, although defined originally to be long interspersed elements of high copy number, particularly in mammalian genomes, are now recognized as *active retroposons* in contrast with retro*trans*posons (with LTRs) and *passive* SINEs and processed retrogenes (Carrol et. al. 1989, Finnegan 1985, 1989, Hutchinson 1989, Rogers 1985). Examples are L1Hs (Kpn1) in primates 6.5kb, and L1Md (MIF-1) in rodents 7 kb, with extensive sequence homology, forming the L1 super-family. There are 90,000 partial L1Hs, but only about 1000 complete elements, and possibly at most 1-10 active elements.

Elements of the L1 type are widespread in eukaryotes including *Xenopus* Tx1, *Drosophila* I, F, G, *maize* Cin4, *Trypanosoma* Ingi. The copy number in mammals is notably higher 10^3- 10^5 (10% of mouse DNA) compared with 10^1-10^2 in the other groups. The elements in mammals show *concerted evolution*. L1Md sequences are only 4% divergent within *Mus domesticus*, but 33% between mouse and human. The most conserved parts are only 30%

divergent between marsupials and mammals. Concerted evolution is explained either by gene conversion or a few complete elements acting as *molecular drivers*, (Dover 1982) leaving many defective copies.

Terminal 5-15bp direct repeats arise from duplications of a single target sequence. The variable size of these, and the lack of an endonuclease *int* domain in the RT supports non-specific integration at a nick or staggered break. Integration may be associated with repair of staggered DNA breaks (Voliva et. al. 1984). L1s are terminated by a polyadenylation signal followed by A_n, $(NA_m)_n$ or both. This pattern is paralleled in both Alu-like SINEs and processed pseudogenes. Of the two open reading frames, ORF1 codes a hydrophilic protein with similarities to retroviral *gag* while ORF2 encodes reverse transcriptase. Complete functional L1 elements remain elusive. They frequently have a partial deletion of the 5' end, or an inversion, or both.

The nature of I-R hybrid dysgenesis in *Drosophila* shows that these elements code for their own regulatory repressors and that activation of I elements is dependent on cellular control phased in relation to meiosis, which is also capable of selective effects on maternal and paternal chromosomes in a single cell and is sensitive enough to allow recovery of fertility during a breeding season (Finnegan 1989). A 6.5kb transcript is detected in *early* teratoma cells suggesting a possible transposition in oocyte diplotene or cleavage embryos (Weiner et. al. 1986).

SINEs or short interspersed repeated DNA (Deininger 1989) occur in very high copy number (10^4-10^6) and are similarly confined to vertebrates and particularly *mammalia*. They are transcribed by pol III which uses internal promoters, permitting the transcribed sequence to contain all the information necessary for self-induction. The 300bp primate Alu element is a dimer, with pol III promoters remaining only in the left half. The 900,000 human Alu elements constitute about 5% of the genome, involving many structural genes. The Alu monomer is derived from the 7sRNA of the signal recognition particle involved in inserting nascent proteins into the endoplasmic reticulum (Ullu & Tushudi 1984). Other families appear to have come from other nuclear RNAs such as tRNAs. A t-RNA-like structure appears common to pol III activated SINEs including Alu (Okada 1990). Similar elements occur also in the tortoise, newt and salmonid fishes (Koishi & Okada 1991). The structure of SINE families is thus consistent with *primitive reentry* - the reassertion of primitive RNA structures, such as tRNAs, as transposable sequences in advanced organisms under a regime of cell-modulated reverse transcriptase activity, providing new classes of short non-translated element with distinct origins, but common secondary structure and mutational and regulatory avenues.

SINE families vary from 2-16% divergence in sequence, greater than the L1 superfamily. A near-neutral rate is consistent with their non-coding nature. Sequence variation in human Alu indicates either repeated bursts of replication followed by quiescent periods during which the families have diverged, often by neutral mutational variation, or that only a small sub-population are able to transpose, leaving behind inert variants over time (Bains 1986). Detailed sequence analysis suggests three to four bursts coinciding roughly with mammalian, primate, and great ape radiations (Willard et. al., 1987, Quentin 1987).

Transcription continues through an A-rich section at their end into adjacent cellular sequences where it terminates in a T-rich region. This region then assists the reverse transcription of the element by snapping back to the poly A section to form a primer for transcription. In a similar manner to the L1 superfamily, the A-rich section is variable in length and sequence again having A_n, $(NA_m)_n$ or a composite. Integration is often at A-rich sites, but is not sequence-specific and results in varying target repeat lengths of 9-21bp, again consistent with no nicking integrase. SINE transposition can result in tandem insertions occur into other SINEs or LINEs to form composite elements. Both SINE and LINE transcription can be elevated in undifferentiated cells (Weiner et. al. 1986). These similarities to LINEs suggest SINEs are reverse transcribed by LINEs.

Integration of processed pseudogenes also has structural similarities to LINEs and SINEs, including A_n or $(NA_m)_n$ ends, insertion into the right end of an A-rich region and variable length target site duplications, suggesting all three are carried by the same process. The single or very small group incidence of I-R hybrid dysgenesis is consistent with retroposon activation in or just before meiosis (Finnegan 1989). A 6.5kb putative L1 transcript is found in early phase teratomas, suggesting a similar mode of expression in mammals (Weiner et. al. 1986).

6 : Recombinational Ecologies in Retrotransposons and Retroviruses

The LTR-retroelements are typified by retrotransposons such as Ty of yeast (Boeke 1989), *Copia* and *Gypsy* of *Drosophila*(Bingham & Zachar 1989), and mammalian retroviruses (Varmus & Brown 1989, Varmus 1982), which display both horizontally infectious *exogenous* habits and germ-line transmitted *endogenous* forms such as murine intracisternal A-particles. IAP-like sequences represent about 0.3% of chromosomal DNA in mice or 1000 copies (Lueders et. al., 1977, Hawley et. al. 1982). Retrovirus-related sequences may thus constitute up to 1% of mammalian DNA.

In addition to the core of capsid *gag* and polymerase *pol* genes, retroviruses have an additional envelope *env* gene responsible for membrane budding. Reverse transcription occurs within the capsid. The encapsulated genome is diploid, consisting of two RNAs plus priming tRNA and polymerase. The RNAs may include other viral or cellular RNAs with a size up to ~9 kb. There are three mechanisms for mixing functional and non-functional elements, reciprocal recombination, gene conversion, and copy choice (Boeke et. al. 1988) in which a jump from one RNA to

another occurs during reverse transcription at a frequency of 10-30%. These processes allow complementation of genotype between distinct RNAs, forming recombinant retroviral genomes including cellular genes.

The long terminal repeats (LTRs) 200-1400bp contain transcription promoters and terminators and multiple (6-8) enhancer sites (Speck & Baltimore 1987), which enable regulation specific to species, tissue type, and developmental phase, as evidenced in HTLV2 (Chen et. al. 1984), and C-type retrovirus from embryonic tissues (Sawyer et. al. 1978, Moroni & Schuman 1977, Staber & Schläfli 1978, Deinhardt 1980). Post-transcriptional regulation also occurs (McDonald et. al. 1988). Cellular inhibition of some retroviruses has been noted, as well as converse activation of retroelements under a variety of forms of genomic stress including heat-shock (Firtel 1989), anoxia (Anderson et. al. 1988)· hypomethylation (Kuff 1988) and radiation (McDonald et. al. 1988).

Integration results in short target site duplications of 4-6bp, and although it is not sequence-specific, some retroviral insertions appear to be chromosomally selective. The LTR of an amphotropic retrovirus with homology to ecotropic AKR hybridized in a manner consistent with ~50 copies in *males only* of four inbred strains of *M. musculus* (Phillips et. al. 1982). These represent 3% of the Y chromosome if all are on the Y chromosome.

Although retroelements do not excise from the cellular genome as part of their life-cycle, recombination between the LTRs leaves a *solo-LTR* as the end result, often accompanied by reversion of a null mutation. Some retroelements have a typical complement of ~20-50 full copies, but ~1000 LTRs. The solo-LTR may continue to function as an enhancer of an adjacent cellular gene. Recombination may provide for occasional reversion of a solo LTR back to a full element. Different mammalian species display a variety of *endogenous* retroviral ecologies with differing characteristics, recombinational events between different viral strains, and modifications of both genes such as *env* and particularly the LTRs (Callahan 1988, Stoye & Coffin 1987, 1988) including a variety of cryptic elements. Similar interactions occur between *exogenous* and *endogenous* habits.

Retrotransposons such as *Copia* and *Gypsy* of *Drosophila*, and Ty of yeast have no *env* gene for specific cell-to-cell transfer, but have similar structure in most other ways to retroviruses, including capsid formation. The 2.3kb THE1 (transposon-like human elt 1) occurring in 10^4 copies, with 350 bp LTRs occurring in $3x10^4$ copies may represent a particularly prolific defective retrovirus. It has only a 129 codon open reading frame with no homology to retroviral *pol* (Weiner et. al. 1986, Finnegan 1989).

7 : Mutator Processes, Enhanced Adaptability and Event Space Dimensions

Bacterial transposons Tn5 and Tn10 have been shown to have similar advantage to mutator genes against otherwise isogenic strains under chemostat competition (Chao et. al. 1983). Successful strains show transposition and in particular insertion of IS10 at determining sites. Hence possessing the transposon is a strongly favoured phenotype, in contradiction to the selfish DNA hypothesis (Doolittle & Sapienza 1980, Orgel & Crick 1980). The same conclusion holds for phages λ and μ and IS50 (Weiner et. al. 1986). The conservation of the major classes of eukaryote transposable element may also be a result of enhanced adaptability, particularly in times of genomic crisis. Mutator genes cannot function in eukaryotes because sexual recombination separates unlinked genes.

One of the principal roles of transposable elements in evolution may be to introduce new classes of probability structure which are also phased with the higher level informational structure of the genome and its regulation, rather than the lower level primary structure of base sequence mutation and random deletion/insertion. The larger scale spontaneous mutations in mammals are dominated by deletions unrelated to gene structure, possibly associated with topoisomerase and nuclear scaffold effects. By contrast, although often lacking promoter sequences, retrogenes consist of a complete coding gene unit, an insertional mutation of relatively high probability of functional advantage if coupled with suitable promoters. Their capacity to associate genes with *new* regulatory elements may complement the more restricted action of cellular gene duplication (Brosius 1991) making the lack of promoter a positive feature. Various forms of gene conversion may provide an avenue for assimilating information contained in inactive pseudogenes or regaining new function.

The very much smaller information content of a typical response element by comparison with a typical coding sequence makes it possible that retrogenes constitute an optimal form of higher level mutation, given the upstream promoter structure of pol II transcription. Calculation of the probability of finding a cryptic response element consisting of 8-10 key bases with 20%-25% divergence, flanked by a 6 bp inverted repeat (20bp) with possible 1 base inserts in either side of the stem or loop can be approximated as follows :

Probability of 8-10bp loop ~ .0004 - .004,   Probability of stem ~.01,

Thus probability of a response element ~ $4x10^{-6}$ - $4x10^{-5}$

Hence probability (response elt /1 kb pseudogene) ~ $4x10^{-3}$ - $4x10^{-2}$ ~ 10^{-2}

The very short sequences in the MuLV LTR enhancer (5-15bp) show that even this construct may overestimate complexity. By contrast, if a 450bp open reading frame has 70 amino acids with 25/64 possible codon changes, 40 with 8/64 and 40 with 4/64 we have a probability of ~ $1.3 x 10^{-113}$. Allowing 550 positions per kb we have ~ 7 x

10^{-111}.

It is thus possible that although only a minority of retrogenes carry 5' regulatory signals, the probability of neighbouring sequences to an insert harbouring cryptic promoters is high enough to provide useful retrogenes such as the case noted with human amylase. Trends such as increasing AT content in spacer regions (Moreau et. al. 1982) and similar increase in pseudogene mutations (Graur et. al. 1989) may further increase the likelihood of specific signals such as TATA and AATAAA and contribute to enhancer structures, through weaker base-pairing.

LTR-retroelements provide transposition of coupled LTR regulatory sequences and recombinant coding genes, and support cell to cell transfer of information through budding. The formation of solo-LTRs in high copy number also provides an important independent mechanism for the dissemination of regulatory sequences, complementary to the spread of coding sequences by retrogenes (McDonald 1990). LTRs are both highly variable and include generic promoters and multiple enhancers capable of constitutive, tissue-specific and hormonal regulation.

SINEs such as Alu may contribute further types of transformation through their ubiquitous repeated nature as illustrated in fig 3. The existence of repeated sequences makes further provision for gene duplication through unequal crossing over and subsequent conversion events. Limitation of the effects of inversions and translocations caused by repeated sequences to tolerable levels is likely. The apparent absence of DNA transposons and compound foldback structures in mammals may reflect the replacement of the TE type translocations by more effectively modulated Alu-based events. LINEs provide for modulated expression of transposition in a manner possibly linked to cellular recombination in gametogenesis fig 2, thus forming a central catalyst for modular transposition.

If each of these modes of transpositional transformation is limited to a tolerable load per generation, probabilities of deleterious mutation remain fixed while probabilities of functional mutation are elevated by a factor of up to 10^{100}. The variation of global family distributions between for example *Drosophila* and mammals may thus reflect distinct stochastic dynamical systems linking element and organism survival.

8 : Coordinated Models of Recombinational and Element-Mediated Transposition

Cellular recombination, including gene duplication and divergence are pivotal to the structure of the genome. The very existence of repeated sequences in the genome can lead to unequal meiotic crossing over through recombination between staggered repeats, and gene duplication or the spread of one sequence copy through a tandemly repeated family. Gene conversion, resulting from repair of mismatch between partially homologous alleles of a gene, is a more general process which can occur between sequences on different chromosomes. The free single-stranded feeler appears to be dominant in conversion.

Both these processes constitute important cellularly-regulated mechanisms of sequence transposition, which play a role in processes from the cassettes of the yeast MAT mating locus, through to the maintenance of sequence uniformity repeated genes such as those for rRNA and sequence relationships in immunoglobulin families (Edelman & Gally 1970). The particular form of cellular recombination mechanisms varies from predominantly homologous form in yeast through to non-homologous forms in the majority of eukaryotes (Fink 1987). The exact form of recombination is likely to be specific to each major phylum, and apply differently to coding gene families, transposable elements and simple sequence DNA.
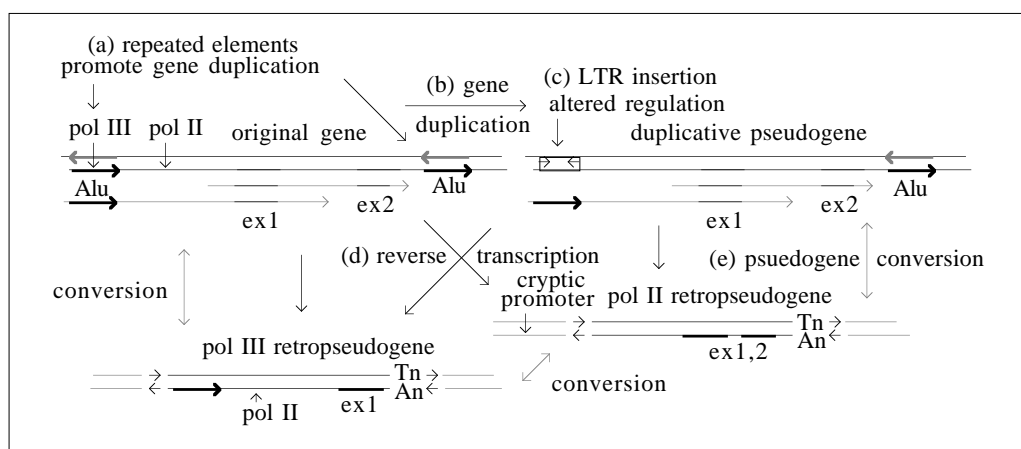

Fig 1 : Coordinated family recombination : (a) Repeated elements such as Alu and L1 facilitate gene duplication (b). This is complemented by pol II and pol III (promoter-containing) retrogene formation (d). LTR insertion (c) and cryptic promoters recombine further regulatory signals. Conversion (e) results in intra-family recombination.

The universality of recombinational enzymes is hinted at by the common use of the χ sequence in both prokaryotic λ-phage and immunoglobulin somatic recombination signals (Kenter & Birshtein 1981). The relation between the mini-satellite consensus GGAGGTGG GCAGGAXG and χ (Jeffreys 1985) appears to provide a mechanism for its

transposition throughout the genome. The observed rates of recombination are 10 times higher than average consistent with being recombinational hot spots.

The high incidence of transposable families in the cellular genome is a source of frequent recombinational events including inversion excision or insertion. Merely by existing, a repeated family can promote duplication, both of itself and of any gene flanked by tandem repeats. Gene conversion can also act on transposable families both to create new variants and to standardize a family through concerted evolution. Cellular recombination processes between simple sequence DNA may also assist in disseminating promoter/enhancer sequences.

One of the most interesting possibilities is that gene duplication and retrogene formation can act as complementary transpositional pathways promoting recombinational diversity. Duplicative and retro-pseudogenes to form a family gene pool in the organism capable of generating diversity through conversion events. Novel regulatory or functional genes existing in a divergent family would enable mutational divergence to induce differential changes in coordinated regulation, with recombinant pseudogenes providing either a type of class switching or more graduated change than that possible by a single mutating gene.

Coordinated family recombination : Duplicative gene families and their processed retrogenes constitute a mutually interactive recombinational pool of variants optimizing modular gene diversity through independent routes which both preserve and recombine regulatory sequences, fig 1.

These mechanisms taken together lead to a model which generates a diversity of recombinational mechanisms to provide as general a class of genetic transformations as possible. High copy repeat SINEs and LINEs provide local nuclei for unequal crossing over, which increases the probability of duplication of cellular genes, complete with regulatory signals. Three further processes recombine coding genes with new regulatory signals. Pol II retrogenes provide coding sequence, which if it is preferentially inserted into an AT-rich region already has enhanced probability of possessing a cryptic promoter. Pol III transcripts provide a further class of sequence sometimes including promoter regions. Complementary to this, LTR insertion provides an independent route for generating new control regions. Finally the interaction of duplicative gene families with their collective retrogenes provides a pool of successive mutual rearrangements through cellular conversion.

9 : Integral Evolution

After the chiasmata of crossing over become apparent, the diplotene phase of meiosis occurs. Lampbrush chromosomes with 100kb loops producing massive transcripts proceeding far past their normal 3' ends into various repetitive sequences are notable in many species. Although part of the function of these, particularly in large oocytes of rapidly dividing embryos such as *Xenopus*, includes maternal mRNA precursors, they have also been cited as a recombinational phase providing for gene conversion processes. Rates of putated spooling motion of DNA around the loops allows time for transient sequential expression of the entire genome during oogenesis in *Xenopus* (Callan 1963, 1969, Wolfe 1972). The diplotene lampbrush phase in mammals is very prolonged. In humans it lasts from the fifth month of gestation to menopause 40 years later. Transcription over this period cannot be necessary for oocyte mRNA production. The relation of these transcripts to Davidson & Posakony's (1982) embryonic dsRNA remains to be determined. The female plays a predominant role, both because of the length of diplotene and the fact that the lampbrush phase occurs in both autosomes and the X chromosomes (King 1978). By contrast lampbrush expression occurs only transiently and on the Y-chromosome in spermatogenesis (Watson et. al. 1987).

The prospective dependence of SINEs such as Alu and processed retrogenes on the L1 LINE reverse transcriptase suggests that the three structures of LINEs, SINEs and processed retrogenes may all be mobilized in a form of concerted transposition phased with meiosis and driven by the L1 elements in a manner similar to I-R dysgenesis (Weiner et. al. 1986). It is noticeable that in several genes with testis-specific variants including the mouse zinc-finger Zfx and cytochrome c, the retrogenes formed arise from the non-testis version of the gene, supporting *oogenic* transposition. Although retrogenes are predominantly spliced, suggesting a cytoplasmic origin, the exceptions such as $\psi\alpha4$ (both introns), rat preproinsulin 1 (one intron), partially processed U2 and the predominantly nuclear expression of the Alu family allow for nuclear retrogenes. The lampbrush chromosome phase is bypassed in *Drosophila* supporting the different mode of action of its mobile element ecology.

Meiotic transcript expression could also explain how genes other than the most ubiquitous housekeeping genes become retrogenes. If transcription is hyper-regulated in the prolonged diplotene e.g. by high polymerase content to produce transcripts reflecting genomic information rather than metabolic enzymes, the transcript population may reflect a more balanced distribution, in which rare genes are much more heavily expressed than usual. Variation may also occur over time, during the extended diplotene, and may involve unusual patterns of transcription, for example involving long pol III transcripts.

The wide differences of ~30% divergence in single-copy DNA between rat and mouse, compared with only 2% between man and chimp, is a function of the time of the divergence, the longer generation time (~15yrs) and hence lower real-time tolerable mutational load in higher apes, and more stringent DNA repair. However the more than comparable differences in phenotype, particularly in brain structure, of the higher primates suggests a more elaborate explanation is required. The long generation times, have to be explained by increasingly improbable mutations of a

few key genes, or by a recombinational mechanism to allow the germ line to keep pace with the changing somatic phenotype.
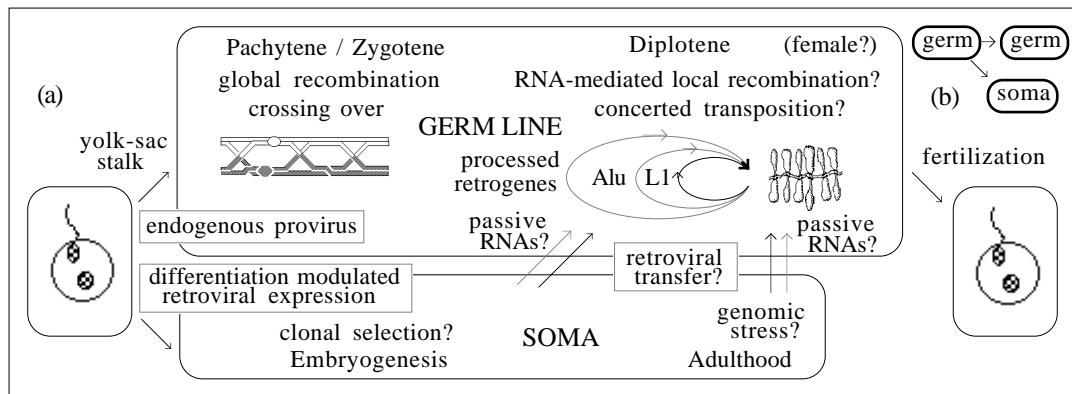

Fig 2: Possible linked structures in the integral model of transposition-based evolution (a), include violations of Weismann's doctrine (b).

Mammalian transpositional evolution is thus dominated by the complementary action of two reverse transcriptases - the retroviral and LINE RTs. While the LINE RT has catalysed the emergence of the LINEs themselves, the ubiquitous SINEs and the processed retrogenes, the retroviral RT acts as a vector for passive and recombinant cell to cell transfer of genes and mutational retroviral and solo LTR inserts. The *integral evolution* model combines complementary action of the these two retroelement RTs into a complementary action between the meiotic expression of LINEs, SINEs and retrogenes and the developmental expression of retroviruses :

(1) Meiosis includes a local recombinational editing process, involving coordinated mobilization of SINEs and retrogenes by LINEs, and possibly cellular conversion processes. This process is mediated principally through the female extended diplotene and is pivotal to the evolution of mammalian species with long generation times, particularly the great apes and *Homo sapiens* (King 1978).

(2) Retroviral transfer to the germ-line both in early development and adult life contribute an additional source of mutations containing effective somatic information derived by clonal selection of somatic cells during development.

The possibility that retroviral transfer from the soma to the germ-line can stochastically violate Weismann's doctrine has been a controversial theme (Gorczynski & Steele 1980, Steele 1984) which nevertheless has received repeated interest and comment (Reanney 1974, 75, 76, Bernstein et. al. 1983, Vanin 1984, King 1985, Pollard 1987). Somatic information occasionally penetrating the germ line in the form of transcribed RNAs could subsequently recombine with germ-line genes in the diplotene phase. Certain aspects of the retrovirus life cycle could form a basis for such *soma-germ transposition.*

Differentiation-specific expression is common to endogenous retroviruses and is consistent with selective inhibition of transcription in germ line cells, but capacity for activation of at least some retroviruses in other cell types. Antisera to simian sarcoma virus demonstrated viral antigens in 24/24 human placentas examined (Deinhardt 1980). Such somatic expression and would allow proviral germ integration since the block is at transcription and the integrase is packaged within the viral capsid. The

Recombinant retroviruses carrying cellular genes would be particularly effective at both having a specific regulatory effect and carrying the *same* pattern of regulation subsequently back into the germ line. Such somatic information would already be selected as fit by the regulatory competition and controlled cell death that is a prominent feature of mammalian neurogenesis (Blakemore 1991) and immunogenesis. The existence of recombinant oncoviruses, and recombination rates per cycle of 10-30% demonstrate that this process is able to act generally enough to apply to a variety of cellular regulatory genes of length $< \sim 10$ kb, as illustrated by $C_{\varepsilon 3}$. A selectively male retrovirus has been noted (Phillips et. al. 1982). A variety of retroviruses carry sexually selective steroid enhancers. This raises the possibility of sex-specific activation of retroviruses in gametogenesis and development.

Such a stochastic process would thus only permit fixation of acquired characteristics on an occasional mutational basis, however it would resolve the conceptual problems of describing an organism as complex as *Homo sapiens* as a mere causal offshoot, whose only function in evolution is to facilitate fertilization of a single-celled germ line.

Modulation of the process would limit the occurrence of deleterious mutations to a tolerable load per generation. An estimate of 1% of retrovirus-related sequences in human DNA based on an average size of 1.5kb (5 solo LTRs per 5kb retroviral unit) would constitute 20,000 copies, however this number is accounted for by the ~40,000 human THE1 LTR-retroelement family inserts alone. 20 smaller families could lift this figure to 60,000-80,000, similar to the ~90,000 L1 copy number and ~50,000 cellular gene number, but less than the 18/1 ratio of retrogenes to cellular

genes predicted by Zuckerkandyl et. al. (1989). Given 35 myrs at 20 yrs/generation we have 1.7×10^6 generations representing a germ-line entry rate of .045/generation, somewhat less than tolerable loads. Several opposing factors may make this figure inaccurate, including mutational assimilation of solo LTRs, and post-insertional family amplification. Small numbers of mutational events may be consistent with quantum non-convergence (King 1989).

The above processes could be linked to a second set of transpositional processes occurring in embryogenic development and adult life. A particular strain of recombinant virus generated during developmental activation, including for example a mutant gene effecting neurogenesis could trigger a regulatory change in much the same way the better-known leukaemia viruses cause tumorigenic transformation. Expression of such elements could result in mutations to new regulatory schemes, such as a repeated cycle of mitotic development, or a new pattern of cytotactic growth. The higher levels of growth factor transcripts in early embryogenesis would promote the formation of retrogenes for these genes at this time.

A further factor which is likely to intervene in this process is genomic stress. If a given cell type is under a form of stress, this may stimulate concerted transposition, activating germ-line transfer. The elongated diplotene could be phased to pick up retrotranscribed stress information through recombination. A variety of circumstances could cause stress, including repeated neuronal stimulation in learning.

The development of the 10^{15} synapses in the human brain requires the action of only 30,000 genes, about 60% of the total human complement, representing a particularly challenging problem in parallel regulation. Recent work indicates that development of the cerebral cortex may proceed on general principles of coordinated growth, tissue layer organization and cell migration in which specific sensory structures are only later established, partly through stimulation from the developing senses (Blakemore 1991), including chaotic excitation (King 1991). Regulatory competition and controlled cell death is a prominent feature. Diversity of both developmental proteins and interactive molecules such as neurotransmitter receptor proteins can play a significant role in brain organization. The receptors involved in long-term potentiation in the hippocampus also appear to be derived from genes involved in embryonic differentiation making a further link between these. Notably the G-protein linked receptors form an extensive intronless gene family consistent with the coordinated family recombination model, fig 1.
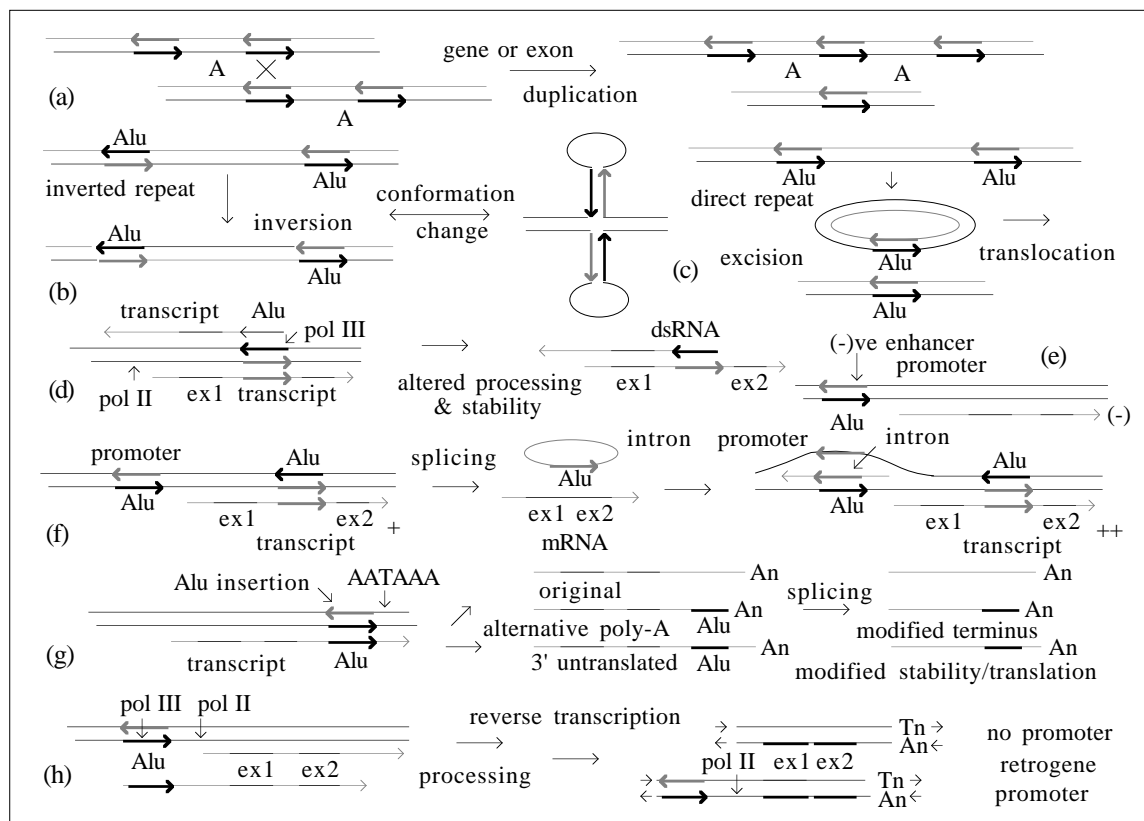

Fig 3 : Diverse mutational and regulatory potential of SINEs: (a) Gene or exon duplication, (b,c) Inverted and direct Alu repeats permit chromosomal reaarangements, (d) anti-sense pol III transcript forming dsRNA, (e) (-)-ve enhancer (f) spliced Alu intron interacts with Alu in the promoter, (g) altered poly-A sites, translation or stability by 3' Alu insertion, (h) generation of a retrogene with 5' pol II promoter by upstream transcription from a pol III Alu site.

Appendix : SINEs as Cellular Regulatory Elements

Although copy numbers vary widely between species, arguing against a fundamental regulatory role, a variety of

cellular functions have been suggested for Alu elements, fig 3, including origins of replication, transcription control, RNA processing, promoting recombination, transposition or inversion of sections of the cellular genome bounded by repeated elements, and limiting gene conversion by insertion into one of a family.

Some SINE transcripts show evidence of tissue-specific expression. The rat ID sequences (Sutcliffe et. al. 1984a,b), and a primate 200bp RNA homologous to the left Alu monomer are expressed selectively in brain tissue (Watson & Sutcliffe 1987). Although the dissimilarity of the rat and primate sequences suggests they may not perform a conserved function, their pattern of expression is conserved. The possibility that pol III transcripts, possibly including anti-sense RNA function in gene regulation has further support (Lassar et. al. 1983, Carlson & Ross 1983, Weiner et. al. 1986). Manley & Colozzo (1982) have also proposed an Alu pol III transcriptional control model based on pol III transcription from Alu. The frequency of Alu could allow internal binding of their sequences in introns or binding between pol III transcripts and hnRNA to play a role in post-transcriptional regulation similar to Davidson and Britten's (1979) model, or Alu in the 3' end at translation (Yamamoto et. al. 1984). It has been noted that differing sections of Alu bind C-factor, T-antigen and other proteins. *Cis* acting regulatory function would allow differential divergence from consensus to provide differential regulation. Saffer and Thurston (1989) have discovered a monkey Alu element containing a 2-5 fold modulating 38 bp negative enhancer - the *protein reducing sequence* which is followed by $(GT)_n$ instead of A_n, which acts on a variety of promoters of both pol II and pol III and may inhibit Alu itself. The motif could be common to a $(GT)_n$-containing subfamily, providing for coordinated regulation. The concept of functionality and selection may have more varied constraints, arising from RNA secondary structure than those on coding sequences. Zuckerkandyl et. al. (1989) have coined the term *cheap genes* for near-neutrally evolving forms of function.

REFERENCES

Anderson G.R., Stoler D.L., Scott J.P., Farkas B.K. (1988)
    Induction of VL30 element expression as a response to anoxic stress in Lambert et. al. 265-74.

Ashworth A., Skene B., Swift S., Lovell-Badge R. (1990) *Zfa is an expressed retroposon derived from an alternative transcript of the Zfx gene.* EMBO J. **9** 1529-34.

Bains W. J. (1986) *The multiple origins of human Alu sequences* J. Mol. Evol. **23** 189 - 99.

Berg Douglas E., Howe Martha (ed.) (1989) <u>Mobile DNA</u> American Society for Microbiology, Washington D.C.

Bernstein L., Mount S., Weiner A. (1983) *Pseudogenes for human small nuclear RNA U3 appear to arise by integration of self-primed transcripts of the RNA into new chromosomal sites* Cell **32** 461-72.

Bingham P.M., Zachar Z. (1989) *Retrotransposons and the FB transposon from Drosophila melanogaster*
    in Berg & Howe 485-502.

Blakemore C. (1991) Sir Douglas Robb Lectures, University of Auckland.

Boeke J.D. (1989) *Transposable elements in Saccharomyces cerevisae* in Berg & Howe 335-74.

Boeke J., Eichinger D., Fink G. (1988) *Regulation of yeast Ty element transposition*
    in Lambert et. al. 169-82.

Brosius J. (1991) *Retroposons - seeds of evolution* Science **251** 753.

Callan H.G. (1963) *The nature of lampbrush chromosomes* Int. Rev. Cyt. **15** 1-34.

Callan H.G. (1969) *Biochanmical activities of chromosomes during the prophase of meiosis*
    in Lima de Faria A. (ed) <u>Handbook of molecular cytology</u> North Holland Amsterdam 540-552

Callahan R. (1988) *Two families of human endogenous retroviral genomes* in Lambert et. al. 91-100.

Carlson D., Ross J. (1983) *Human β-globin promoter and coding sequences transcribed by RNA polymerase III*
    Cell **34** 857-64.

Carrol D., Knutzon D., Garrett J. *Transposable elements in Xenopus species* in Berg & Howe 567-74.

Chao L., Vargas C., Spear B., Cox E. (1983) *Transposable elements as mutator genes in evolution*
    Nature **303** 633-5.

Chen I., McLauglin Golde D (1984) *Long terminal repeats of human T-cell leukemia virus II genome determine target specificity* Nature **309** 276-9.

Copeland N., Hutchinson K., Jenkins N. (1983a) *Excision of the DBA ecotropic provirus in dilute coat-color revertants of mice occurs by homologous recombination involving the viral LTRs* Cell **33** 379-87.

Copeland N., Jenkins N., Lee B. (1983b) *Association of the lethal yellow (A^y) coat color mutation with an ecotropic murine leukemia virus genome.* Proc. Nat. Acad. Sci. **80** 247-9.

Davidson E., Britten R. (1979) *Regulation of gene expression : possible role of repetitive sequences*
    Science **204** 1052-9.

Davidson E., Posakony J. (1982) *Repetitive sequence transcripts in development* Nature **297** 633.

Deinhardt F. (1980) *Biology of Primate Retroviruses* in Klein 357-97.

Deininger P.L. (1989) *SINEs : Short interspersed repeated DNA elements in higher eukaryotes*
    in Berg & Howe 619-36.

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

Dover G. (1982) *Molecular drive : A cohesive mode of species evolution* Nature **299**, 111-7.

Edelman G.M., Gally J.A. (1970) *Arrangement and Evolution of Eukaryotic genes*
    in *The neurosciences* Schmit F.O. (ed.) Rockefeller University Press.

Engels W. (1989) *P elements in Drosophila melanogaster* in Berg & Howe 437-84.

Fink G.R. (1987) *Pseudogenes in yeast?* Cell **49** 5-6.

Finnegan D.J. (1983)*Retroviruses and transposable elements - which came first?*
Nature **302** 105-6.

Finnegan D.J. (1985) *Transposable elements in euckaryotes* Int. Rev. Cyt. **93** 281-325.

Finnegan D.J. (1989a) *F and related elements in Drosophila melanogaster*
in Berg & Howe 519-21.

Finnegan D.J. (1989b)*The I factor and I-R hybrid dysgenesis in Drosophila melanogaster*
in Berg & Howe 503-18.

Firtel R.A. (1989) *Mobile elements in the cellular slime mold Dictyostellium discoidium*
in Berg & Howe 557-66.

Geyer P., Green M., Corces V. (1988) *Molecular basis of transposable element-induced mutations in Drosophila melanogaster* in Lambert et. al. 123-30.

Graur D., Shuali Y., Li W. (1989) *Deletions in processed pseudogenes accumulate faster in rodents than in humans* J. Mol. Evol. **28** 279-85.

Green M.M. (1988) *Mobile DNA elements and spontaneous gene mutation* in Lambert et. al. 41-50.

Gorczynski R., Steele E. (1980) *Inheritance of acquired immunological tolerance to foreign histocompatibility antigens in mice* Proc. Nat. Acad. Sci. **77** 2871-5.

Hawley R., Shulman M., Murialdo H., Gibson D., Hozumi N. (1982) *Mutant immunoglobulin genes have repetitive DNA elements inserted into thier intervening sequences* Proc. Nat. Acad. Sci. **79** , 7425 - 7429.

Hollis G.F., Hieter P.A., McBride O., Swam D., Leder P. (1982), *Processed genes : a dispersed human immunoglobulin gene bearing evidence of RNA-type processing* Nature **296** 321-3.

Hutchinson C., Hardies S., Loeb D., Shehee W., Edgell M. *LINEs and related retrotransposons : Long interspersed repeated sequences in the eucaryotic genome* in Berg & Howe 593-618.

Inouye M., Inouye S. (1991) *Retroelements in bacteria* TIBS **Jan** 18-21. - structure of msDNA

Jeffreys A., Wilson V., Thein S.L., (1985) *Hypervariable 'minisatellite' regions in human DNA* Nature **314** 67.

Jenkins N., Copeland N., Taylor B., Lee B. (1981) *Dilute (d) coat colour mutation of DBA/2J mice is associated with the site of integration of an ecotropic MuLV genome* Nature **293** 370-4.

Johns M., Mottinger J., Freeling M. (1985) *A low copy number Copia-like transposon in maize*
EMBO J. **4** 1093-102.

Kenter A., Birshtein B. (1981) *Chi, a promoter of generalized recombination in λ phage is present in immunoglobulin genes* Nature **293** 402-4.

King C. (1978) *Unified field theories and the origins of life* Math. Dept. Rep. Series. **134** Univ. Akld.

King C. (1985) *A model for transposon-based eucaryote regulatory evolution.* J. Theor. Biol. **114** 447-462.

King C. (1989) *Dual-Time Supercausality* Physics Essays **2/2** 128-151.

King C. (1991) *Fractal and Chaotic Dynamics in Nervous Systems* Progress in Neurobiology **36** 279-308.

Klein G. (1980) <u>Viral Oncology</u> Raven Press New York.

Koishi R., Okada N. (1991) *Distribution of the salmonid Hpa1 family demonstrated by in vitrio runoff transcription assay of total DNA :* J. Mol. Evol. **32** 43-52.

Kuff E.L. (1988) *Factors affecting retrotransposition of intracisternal A-particle proviral elements*
in Lambert et. al. 79-90.

Lambert M.E., McDonald J.F., Weinstein I.B. (ed.) (1988) <u>Eukaryotic transposable elements as mutagenic agents</u> Banbury Report 30, Cold Spring Harbor Laboratory.

Lassar A., Martin M., Roeder R. (1983) *Transcription of class III genes : formation of preinitiati9on complexes* Science **222** 740-8.

Limbach K., Wu R. (1985) *Characterization of a mouse somatic cytochrome c gene and three cytochrome c pseudogenes.* Nucleic Acid Res. **13** 617-30.

Limborska S., Korneev S., Maleeva N., Slominsky Jinsharadze A., Ivanov P., Ryskov A. (1987) *Cloning of Alu-containing cDNAs from human fibroblasts and identification of smal Alu polyA RNAs in a variety of human normal and tumor cells.* FEBS Lett. **212** 208-12.

Lueders K., Kuff E. (1977) Cell **12** 963-972.

Leuders K., Leder A., Leder P., Kuff E. (1982) *Association between a transposed α-globin pseudogene and retrovirus-like elements in the BALB/c mouse genome* Nature **295** 426-8.

MacLeod A., Talbot K. (1983) *A processed gene defining a gene family encoding a human non-muscle tropomyosin* J. Mol. Biol. **167** 523-37.

Manley J., Colozzo M. (1982) *Synthesis in vitro of an exceptionally long RNA transcript promoted by an Alu sequence* Nature **300** 376-9.

McClintock B. (1978) *Mechanisms that rapidly reorganize the genome* Stadler Genet. Symp. **10** 25-48.

McClintock B. (1984) *The significance of the responses of the genome to challenge* Science **226** 792-801.

McDonald J.F. (1990) *Macroevolution and retroviral elements* BioScience **40** 183-91.

McDonald J.F., Strand D., Brown M., Paskewitz S., Csink A., Voss S. (1988) *Evidence of host-mediated regulation of retroviral element expression at the post-transcriptional level* in Lambert et. al. 219-34.

Meuth M. (1989) *Illegitimate recombination in mammalian cells* in Berg & Howe 833-60.

Moreau J., Marcaud L., Maschat F., Kejzlarova-Lepesant J. Lepesant J., Scherrer K. (1982) *A+T-rich linkers define functional domains in eukaryotic DNA* Nature **295** 260-2.

Moroni C., Schuman G. (1977) Nature **269** 600-1.

Okada N (1990) *Transfer RNA-like structure of the human Alu family :*

Implications of its generation mechanism and possible functions J. Mol. Evol. **31**, 500-510.

Orgel L., Crick F. (1980) *Selfish DNA : the ultimate parasite* Nature **284** 604-7.

Phillips S., Birkenmeier E., Callahan R., Eicher E. (1982) *Male and female mouse DNAs can be discriminated using retroviral probes* Nature **297** 241-243.

Pollard J.W. (1987) *The moveable genome* Rivista di Biologia - Biol. Forum **80** 11-54.

Quentin Y. (1988) *The Alu family developed through successive waves of fixation closely connected with primate lineage history* J. Mol. Evol. **27** 194-202.

Reanney D.C. (1974) *Viruses and evolution* Int. Rev. Cytol. **37** 21-55.

Reanney D.C. (1975) *A regulatory role for viral RNA in eukaryotes* J. Theor. Biol. **49** 461-92.

Reanney D.C. (1976) *Extrachromosomal elements as possible agents of adaption and development* Bact. Rev. **40** 552-90.

Ridley Mark (1985) *Selfish DNA comes of age* New Scientist **16 May** 34-7.

Rogers J. (1985) *The origin and evolution of Retroposons* Int. Rev. Cytol. **93**, 281-326.

Rosen E., Siversten A., Firtel H. (1983) *An unusual transposon encoding heat shock inducible and developmentally regulated transcripts in dictyostellium* Cell **35** 243-51.

Saffer J., Thurston S. (1989) *A negative regulatory element with properties similar to thos of enhancers is contained within an Alu sequence* Mol. Cell. Biol. **9** 355-364.

Samuelson Wiebauer K., Snow C., Meisler M. (1990) *Retroviral & pseudogene insertion sites reveal lineage of human salivary & pancreatic amylase genes from a single gene during primate evolution* Mol. Cell. Biol. 2513-20.

Sawyer M., Nachlas N., Panem S. (1978) *C-type viral antigen expression in human placenta* Nature **275** 62-4.

Sharp P. (1983) *Conversion of RNA into DNA in mammals : Alu-like elements and pseudogenes* Nature **301** 471-2.

Schwartz-Sommer Z. (1987) *The significance of plant transposable elements in biological processes* in Structure and Function of Eukaryotic Chromosomes Hennig H. Springer-Verlag Berlin Heidelberg 213-21.

Speck N., Baltimore D. (1987) *Six distinct nuclear factors interact with the 75-base pair repeat of the Moloney murine leukemia virus enhancer* Mol. Cell Biol. **7** 1101-10.

Srikantha T., Landsman D., Bustin M. (1987) *Retropseudogenes for human chromosomal protein HMG-17* J. Mol. Biol. **197** 405-13.

Staber FG., Schläfli E. (1978) *Expression of endogenous C-type virtal antigen on mormal mouse haemopoietic stem cells* Nature **275** 669-71.

Stavenhagen J., Robins D. (1988) *An ancient provirus has imposed androgen regulation on the adjacent mouse sex-limited protein gene* Cell **55** 247-54.

Steele E. (1984) Somatic Selection and Adaptive Evolution : On the Inheritance of Acquired Characters 2nd ed. Univ. Chicago Press, Chicago.

Stoye J.P., Coffin J.M. (1987) *The four classes of endogenous murine leukemia virus : Structural relationships and potential for recombination* J. Virol. **61** 2659-69.

Stoye J.P., Coffin J.M. (1988) *Polymorphism of murine endogenous proviruses revealed by using virus class specific oligonucleotide probes* J. Virol. **62** 168-75.

Sutcliffe J., Milner R., Gottesfeld J., Lerner R. (1984a) *Identifier sequences are transcribed specifically in brain* Nature **308** 237-241.

Sutcliffe J., Milner R., Gottesfeld J., Reynolds W. (1984b) *Control of neuronal gene expression* Science **225** 1308-15.

Tchurikov N., Gerasimova T., Georgieva S., Mizrokhi L., Georgiev P. Ilyin Y. (1988) *Concerted transposition in Drosophila melanogaster* in Lambert et. al. 103-11.

Temin H.M. (1980) *Origin of retroviruses from cellular moveable genetic elements* Cell **21** 599-600.

Temin H.M. (1989) *Retrons in bacteria* Nature **339** 254-5.

Ueda S., Hakai S., Nishida Y., Hisajima H., Honjo T. (1982) *Long terminal repeat-lke elements flank a human immunoglobulin epsilon pseudogene that lacks introns* EMBO J. **1** 1539-44.

Ullu E., Tschudi C. (1984) *Alu sequences are processed 7sL RNA genes* Nautre **312** 171-172.

Vanin E.F. (1984) *Processed pseudogenes : Characteristics and evolution.* Biochem. Biophys. Acta **782**, 231-241.

Varmus H. (1982) *Form and function of retroviral proviruses* Science **216** 812-20.

Varmus H., Brown P. (1989) *Retroviruses* in Berg & Howe 53-108.

Voliva C., Martin S., Hutchinson C., Edgell M. (1984) *dispersal processes associated with the L1 family of interspersed repetitive DNA sequences.* J. Mol. Biol. **178** 795-813.

Waddington T.H. The Strategy of the Gene

Watson J.,. Sutcliffe J. (1987) *Primate brain-specific cytoplasmic transcript of the Alu repeat family* Mol. Cell. Biol. **7** 3324-7.

Watson J.D., Hopkins N.H., Roberts J.W., Stietz J.A., Weiner A.M. Molecular Biology of the Gene Benjamin-Cummings, 1987, Menlo Park California.

Weiner A., Deininger P., Efstratiadis A. (1986) *Nonviral retroposons : genes, pseudogenes and transposable elements generated by the reverse flow of genetic information* Ann. Rev. Biochem. **55** 631-61.

Willard C., Nguyen H., Schmid C. (1987) *Existence of three distince Alu subfamilies* J. Mol. Evol. **26**   180-86.

Wolfe S.L. (1972) Biology of the Cell Wadsworth Belmont CA. 418-9.

Yamamoto T., Davis C., Brown M., Schneider W., Casey M., Goldstein J., Russel D. (1984) *The human LDL receptor : A cysteine-rich protein with multiple Alu sequences in its mRNA* Cell **39** 27-38.

Xiong Y., Eickbush T.H. (1990) *Origin and evolution of retroelements based upon their reverse transcriptase sequences* EMBO J. **9/10** 3353-62.

Zuckerkandl E., Latter G., Jurka J (1989) *Maintenance of function without selection :*
    Alu sequences as "cheap genes" J. Mol. Evol. **29**, 504-12.
Zuker C., Cappello J., Chisholm R., Lodish H. (1983) *A repetitive gene family that is induced during*
    differentiation and by heat shock Cell **34** 997-1005.