Journal and Proceedings of The Royal Society of New South Wales

Volume 120 Parts 1 and 2 [*Issued September, 1987*]

pp.39-48 **Return to CONTENTS**

Molecular Facts and Evolutionary Theory

George L. Gabor Miklos

Evolutionary biology has had a fascinating recent history. It was realized more than a century ago that the way of approaching many evolutionary problems lay in studies of morphology. However, as pointed out by Bateson in 1922, "*discussions of evolution came to an end primarily because no progress was being made. Morphology having been explored in the minutest corners, we turned elsewhere. We became geneticists in the conviction that there at least must evolutionary wisdom be found.*" At the same time, while it was clear that morphology must have its bases in embryology, it was instead the mathematically oriented theory of neo-Darwinism that rose to prominence over the next half century. This theory is essentially an amalgam of Mendelian genetics and classical Darwinian selection, firmly based on changes in gene frequencies at particular loci. In the late 1960s, it began to be evaluated at a crude molecular level using gel electrophoresis techniques that allowed the examination of polymorphisms at many enzyme coding loci.

In the mid 1970s the technological advances of genetic engineering ushered in an entirely new era of molecular biology. The molecular biologist became the successor to the pure geneticist, and the focus switched back to the molecular analysis of development. The molecular biology of recombinant DNA revolutionized the previous concepts of genome organization and function and led to a reappraisal of the importance of neo-Darwinism. Present day,* studies of molecular evolution or molecular population genetics are largely the application of recombinant DNA technologies to traditional evolutionary problems, namely the origin, type and extent of genetic variation in populations. or the determination of phylogenies. However, the application of this technology to eukaryotic genomes has spawned so many surprises. that traditional notions of how evolution works have had to be substantially re-evaluated.

The era of molecular evolution, and what will undoubtedly be its *golden era*, began with the cloning technologies of the mid-1970s, when DNA molecules (reviewed by Felsenfeld, 1985), RNA molecules (reviewed by Darnell, 1985) and protein molecules (reviewed by Doolittle, 1985) could be purified and examined by rapid and ultra-sophisticated new techniques. While evolutionary phenomena can be examined at many levels, it was the genome itself which became the most accessible.

In this essay. I shall examine the new findings that have emerged about eukaryotic genomes using the modern technologies, and how these data alter our perceptions of evolution.

THE STRUCTURE OF EUKARYOTIC GENOMES

DNA Amounts

Eukaryotic genomes can have thousands of millions of base pairs of DNA. The variation in genomic DNA between organisms having the same grade of morphological organization can be quite large (Table 1). Thus in annelid worms, there can be a five-fold difference between two species belonging to the same family. In molluscs, there can be at least a ten-fold difference in genome size. In mammals, the variation is less extreme, but is still at least two-fold (Table 2). It is known for example that two closely related, near identical species of barking deer can differ by at least 500 million base pairs of DNA and have their genomes organized radically differently into three and 23 chromosomes respectively.

TABLE 1
VARIATION IN NUCLEAR DNA AMOUNTS IN VARIOUS INVERTEBRATES
(from Conner *et al.*, 1972; Hinegardner, 1974)

ANNELID WORMS			Genome size in million base pairs
Family	*Nephytidae*	*Nephtys incisa*	6600
		Neptys sp.	2200
	Cirratulidae	*C. luscuriosa*	3000
		C. grandis	600
MOLLUSCS			
Limpets		*Lottia gigantia*	400
		Acmaea mitra	900
Venus Clams		*Tivela stultorum*	900
		Mercenaria compechiensis	2100
Nut Clams		*Nucula proxima*	2900
		Acila castrensis	5000

TABLE 2
VARIATION IN NUCLEAR DNA AMOUNTS AND CHROMOSOME NUMBERS IN VARIOUS VERTEBRATES (from John and Miklos, 1987)

	Genome size in million base pairs	Chromosome number
Mutiacus vaginalis (barking deer)	2400	3
M. reevesi (barking deer)	2900	23
Homo sapiens (humans)	3300	23
Oxcyteropus afer (aardvarks)	5500	10

Since eukaryotic genomic DNA exists as long molecules, millions and even hundreds of millions of base pairs in length, it was a primary requirement of any technology to be able to examine small defined sections of a genome in pure form. The discovery of enzymes which specifically cut DNA, the restriction endonucleases, and the availability of bacterial cloning vehicles, meant that specific pieces of DNA from any organism could be purified, fused to cloning vectors and clonally amplified (reviewed by Weinberg, 1985). The end product is millions of copies of a pure

sequence that is now in sufficient mass to be analyzed by sophisticated molecular biological techniques. This tour-de-force of gene cloning is conceptually similar to locating and cloning the proverbial needle in the haystack. So it is with a genome; once a particular sequence of DNA has been cloned, all the tests that are necessary to characterize it can be performed fairly routinely. Thus any part of any genome can be subjected to sequence analysis and the order of the bases determined.

The sequencing technologies have now led us to what many consider as the ultimate goal – the sequencing of the entire human genome (Robertson, 1986). To put this into perspective, it should be remembered that human beings have approximately 3,300,000,000 base-pairs of DNA in their haploid genome, and if sequencing continues world wide at the current rate of 1,000,000 base-pairs per year, this project would take 3,300 years. However, most eukaryotic genomes contain a fair proportion of noncoding DNA, which accounts for more than half of the genome. Thus, if only those portions of the human genome which code for genes are cloned and sequenced, the task is somewhat simplified. Fifty thousand genes, each with a coding length of say 2,000 bases, yield roughly a century of work based on the present technology. Obviously, it will not be possible to compare the total genomic sequences of all mammals in the foreseeable future, nor is this a sensible avenue along which to proceed. In order to obtain some idea of how present day genomes have altered under the rigors of past evolutionary events, it is sufficient to examine small segments of genomes in order to make comparisons between organisms. The critical decision in an evolutionary context is which parts of a genome are to be compared if, for example, the object is to determine a phylogeny. Some parts of eukaryotic genomes consist of junk DNA and these can change extremely rapidly. Other parts, such as conserva~ tive gene families, change relatively slowly. It is prudent therefore to know how eukaryotic genomes are partitioned in terms of sense and nonsense DNA before we rush off and spend time sequencing and characterizing every piece of putatively interesting DNA.

Highly Repetitive Sequences (Junk DNA)

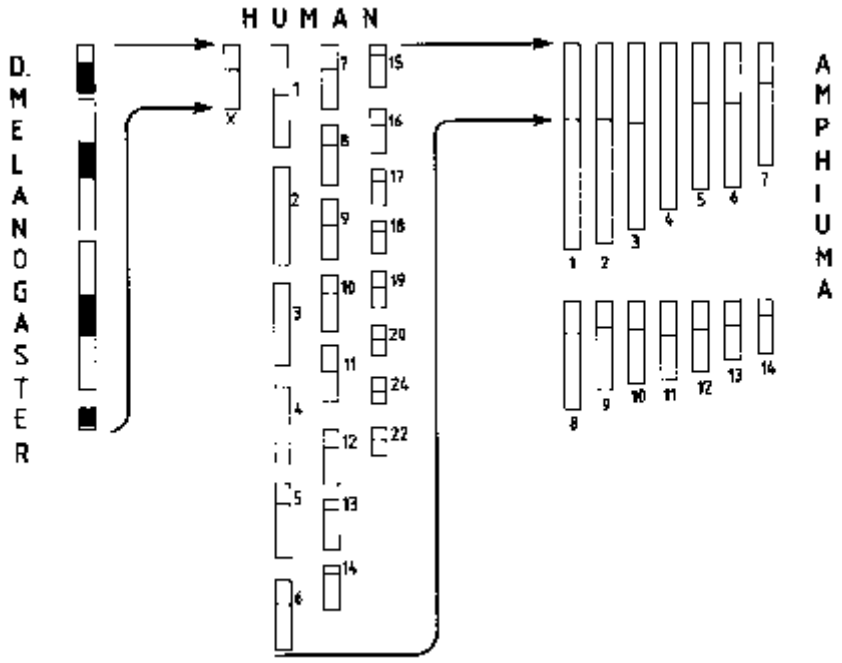

Fig. 1

A major finding concerning most eukaryotic genomes is that a significant proportion of the DNA consists of monotonously repetitive sequences (reviewed in Singer, 1982; Miklos, 1985). In certain crabs, for example, more than 30 percent of the genomic DNA consists of long stretches of DNA made up of alternating AT, yielding a landscape of ATATATATATATATATATAT ... which extends for millions of base pairs. Such findings on localized repetitive sequences turn out to be general, although the basic repeating unit itself is variable in length and sequence. The fly, *Drosophila melanogaster* for example, has a haploid genome of 165 million base pairs or five percent of the human genome. When the fly genome is partitioned, it is found that approximately 50 million base pairs exists as families of repetitive sequences based on variants of repeats such as AACAG, AATAT as well as one long repeat. Since these highly repetitive sequences are not transcribed into RNA, and hence do not code for protein products, nearly a third of the fly genome is silent and is unlikely to have significant evolutionary value (Miklos, 1985). Figure 1 not only illustrates the distribution of localized junk DNA sequences on the chromosomes of the fly (where they are concentrated around the centromeric regions), but puts into perspective the differing genome sizes of three organisms. The entire fly genome can be accommodated in a single human X chromosome, and the *entire* human genome of 3300 million base pairs can fit into a single chromosome arm of the Congo eel *Amphiuma*.

Data such as these are not restricted to invertebrates. In certain American rats such as Dipodomys ordii, more than half of the genome consists of variants of the sequences AAG, TTAGGC and ACACACC= (Fry and Salser, 1977). These three sequence families, known as ms, a and 0,res~ pectively make up 1200 million, 1100 million and 600 million base pairs of DNA, so that there are at least 2900 million base pairs of junk DNA over and above the main 2500 million base pairs that already exist in the basic rat genome (Figure 2).

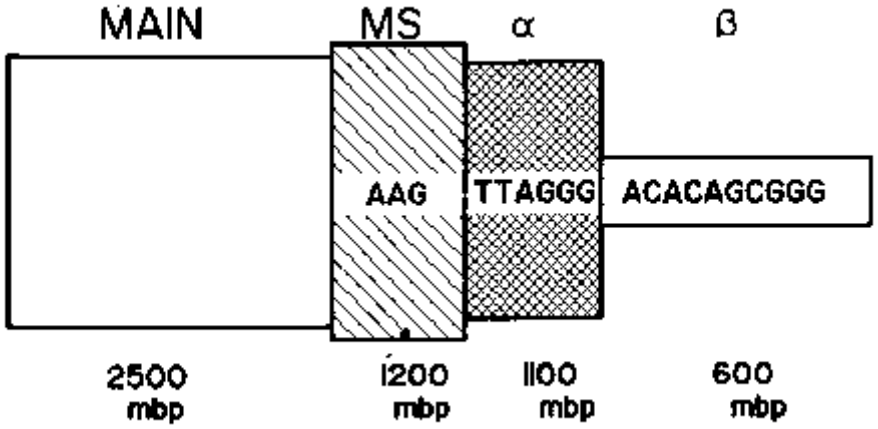

Fig. 2

Nomadic DNA

Genomes of higher organisms are also in a state of flux, containing DNA sequences which are mobile and can move around the genome. These are the so-called 'jumping genes' or nomadic elements which can excise from a particular location in the genome and reinsert elsewhere. The physical consequences of such events may, of course, be profound. If a mobile element inserts into a gene, or into its controlling sequences, a mutation can occur so that the gene is no longer capable of normal expression. Such wanderings of mobile elements are now considered by some scientists to be prime movers in the generation of evolutionary novelty, since not only can

mobile elements turn genes 'on' and 'off' and modulate their activities, but they can act as 'removalists' and relocate other pieces of the genome to new sites, in some cases placing them under the control of different regulatory circuits. When mobile elements insert into other mobile elements, it is relatively easy to see how clustered scrambled arrays of DNA sequences can arise in a genome.

Mobile elements or their defective relatives can make up a substantial proportion of a genome. In the case of *Drosophila melanogaster*, 20 million base pairs of its 165 million base pairs of genomic DNA consists of 30 to 50 different families of mobile elements which move around the genome (reviewed by Rubin, 1983). By contrast, its sibling species *D. simulans* has only about three million base pairs of its genome invested in such sequences (Dowsett and Young, 1982). Since these two species are almost identical morphologically, it is unlikely that these different mobile element populations are, or have been, major players in morphological diversification in these two species.

In general, mobile elements consist of a stretch of DNA a few kilobases (kb) in length, flanked by repetitive sequences of a few hundred base pairs in length. In *Drosophila*, the mobile element families have names such as *copia, 297, hobo, gypsy, H.M.S. Beagle*, and *roo*. Their nomadic nature is affectionately recalled in the flamboyance of their designations. There are furthermore many similarities between some of the nomadic element families of *Drosophila*, and the integrated proviruses of avian and murine RNA tumor viruses (Varmus, 1983).

Split Genes and Multigene Families

A third major finding about eukaryotic genomes is that the coding regions of the genes themselves are not single continuous stretches of DNA but are generally fragmented into peptide coding and noncoding regions, or *exons* and *introns*. Thus in the case of the bithorax gene in *Drosophila*, one particular part of the genie landscape that is copied into RNA is over 70,000 nucleotides in length, but after this RNA has been cut and respliced by the cellular machinery, the resultant messenger RNAs are less than 5,000 nueleotides in length. Most of the original RNA transcript has been discarded (Hogness *et al*., 1985).

Multigene Families

Eukaryotic genes, however, are not only internally fragmented, but turn out not to be the solitary and pristine entities that are used in neo-Darwinian theory. "ReaP genes are often members of multigene families which can make the same or very similar products (Hunkapiller *et al*., 1983). Examples of such multigene or multisequence families, with family memberships ranging from two to hundreds of thousands of relatives, are the haemoglobin genes, the antibody genes, the actin genes, the tubulin genes, the collagen genes, the histone genes, the chorion genes, the transfer RNA, the 55, ISS and 28S ribosomal genes, the highly repetitive sequences and the dispersed repetitive sequences. What of course is immediately obvious is that selection acting on a phenotype, which depends on a multisequence family, can only assess the output of that family as a whole, and not that of its individual members. This in fact is very different to selection acting on the end product of a single locus. The situation has been appreciated by Obta (1983); "*in view of the widespread occurrence of multigene families in genomes of higher organisms, the evolutionary theory based mainly on change of gene frequency at each locus would appear to need considerable revision*."

Pseudogenes

However, our thinking needs even further revision in the light of molecular discoveries. Having found that eukaryotic genomes are littered with sundry localized highly repetitive sequences, dispersed repetitive sequences and nomadic elements, it was then revealed that the cellular machinery is capable of taking the RNA transcripts of genes and back copying them to DNA molecules, which are then reiserted into the genome (reviewed by Baltimore, 1985). These are the non-functional pseudogenes which now adorn the landscape. An excellent example is the glyceraldehyde-3-phosphate dehydrogenase (CAPDH) gene. In both human beings and mice there is only one functional CAPDH gene in the genome, but there are 10 to 30 pseudogenes in human beings and over 200 in the mouse genome! (Piechaczyk *et al.*, 1984).

"Conversion" Phenomena

DNA molecules can also "convert" each other, so that given a family of sequences, a particular one can physically replace parts of another. This has been beautifully demonstrated in the case of three serine tRNA genes on different chromosomes of the yeast *Schizosaccharomyces pombe* (Amstutz *et al.*, 1985). The evolutionary implications of such cellular events for multigene and multisequence families are again profound. For instance, the homogeneity seen at the sequence level in a multigene family may have nothing whatsoever to do with its functional attributes, but may be indicative of homogenizing mechanisms (Dover and Tautz, 1986).

The Genomic Overview

Our glimpse into the genome reveals an arena where change is the rule, not the exception. Although many biologists have been slow to accept this view, the molecular data are now overwhelming. Looking at the genomic macromolecules from a physical-chemical viewpoint, genomic flux ought to be an inevitable by-product of enzymic action. DNA is not a sacrosanct molecule, it is a substrate in the cellular jungle. It can be replicated, cleaved, repaired, nibbled, rejoined, supertwisted, converted, moved around, modified and generally tinkered with, all of which may have serious consequences for the cell and the organism. Just as we find fossils on this earth, so in retrospect it is natural that the genome is littered with fossil DNA which represents nature's evolutionary experiments. The existence of enzymes which act on DNA make a whole series of events irrevocable.

Ohno (1970) was the first to appreciate the situation when he pointed out not only that there was so much junk DNA in our genome, but that "*genes in the euchromatic region on mammalian chromosomes can be compared to oases in a barren stretch of desert*" ... and that..."*For every redundant copy of the pre-existent gene that emerged triumphant as a new gene, hundreds of other copies must have degenerated to join the rank of junk DNA.*" (Ohno, 1982).

In terms of evolutionary perspectives, this cursory survey of eukaryotic genomes has left at least one clear message. We would be foolish to go looking for the action in every piece of DNA. It is firstly necessary to sort the non-functional from the functional material. Many generations of biologists were brought up on the notion that everything ought to be functional since it was assumed that natural selection should have pruned out all aspects of an organism that did not serve a specific function. What had not been readily perceived is that natural selection cannot act on changes which it does not see. It is an editor, not a composer as has been pointed out by King and Jukes (1969). Consequently we have suffered from the malaise of seeking functional explanations first. The recombinant DNA revolution has helped to partially reverse this thinking as has the concept of selfish DNA propounded by Orgel and Crick (1980) and Doolittle and Sapienza (1980). DNA is an information transmission system and it suffers from biochemical noise at many levels (Tautz et al., 1986). Some of the noise causes meaningful changes, some just

causes chaos. How then do we sift the wheat from the chaff and ask about the mechanistic changes which brought about significant evolutionary changes?

THE EUKARYOTIC GENOME IN DEVELOPMENT

Mammals such as the whale, the mouse and the bat are morphologically very dissimilar and are adapted to radically different ways of life. However all three have about the same genome size, and many of their genes and gene families will undoubtedly be very similar. What factors then are responsible for the enormous morphological differences between these organisms? As yet we do not know, but as a result of the molecular data base, we have a very good idea where not to search. The highly repetitive sequences, the mobile elements, the dispersed simple repeats and the pseudogenes can undoubtedly be ignored from a developmental viewpoint. This still leaves nearly 50,000 genes in a vertebrate only some of which will be the important decision-making ones. As Jackson (1986) has so clearly warned us, "*The hip young gunslingers of modern developmental biology shoot hard, fast and often inaccurately. Any likely gene which raises its head is quickly cloned and analyzed.*" However, what we have learnt from our initial forays into the genome is that it is a cruelly unequal maelstrom. Initially it would be foolish to clone genes at random. What is required is to sort the most significant and interesting ones from their less important colleagues.

Executive Genes

How do we distinguish the executive genes from those which dutifully and unerringly carry out the more mundane cellular tasks? This is not at all an easy assignment, but it has been approached most readily in lower organisms such as the worm, *Caenorhabditis* (Horvitz *et al.*,1983) and *Drosophila*, where genes which affect early embryogenesis for example are very rapidly being cloned and sequenced (reviewed by Gehring, 1985). Why do the lower organisms show such promise for unravelling gene circuits involved in development and differentiation and hence in understanding how such genes and circuits produce morphological novelties? The answers are relatively straightforward when compared to mammals. The crucial gene circuits of mammalian embryology largely carry out their functions when the embryo is experimentally inaccessible; in human beings, most major formative events are well and truly over in the first 12 weeks post fertilization and these embryos are thus unsuitable for direct study. Secondly, the developmental genetics of human beings is in a rudimentary state and cannot easily be interfaced with molecular biology as yet. On the other hand, organisms such as *Drosophila* have a short life cycle, manipulative genetics and a burgeoning recombinant DNA data base. The ability to reintroduce genes into the *Drosophila* genome rapidly and precisely mean that significant inroads are now being made into the molecular biology of early development.

The apprehension of many biologists is that what is true for the fly will be irrelevant to human beings. As far as the basic developmental principles are concerned, however, these fears are largely being allayed. The finding that nearly half of the gene products which occur in the brain of the fly cross-hybridize to the human brain was significant in this regard (Miller and Benzer, 1983). Clearly many components of the neuronal circuitry are likely to be very similar in vertebrates and invertebrates.

Owing to the enormous input into the genetics of this fly over the last half century, its genome has been saturation mutagenized so that most genes which affect a particular phenotype can be identified and then subsequently mapped and cloned. The 'importance' of the gene can also be evaluated, since in many cases it is possible to construct a homozygous deficiency just for the locus in question and determine the null phenotype, a procedure not routinely available in mammals.

Many genes which effect the anterior-posterior and dorsal-ventral gradients in the *Drosophila* egg have been discovered (Nusslein-Volhard, 1979; Anderson and Nusslein-Volhard, 1984). The *principles* which underlie these critical developmental gradients are likely to be variations on a theme, so it is sensible to start wherever a foothold presents itself. In *Drosophila*, the first steps in this area have already been taken. Genes affecting the segmentation of the body have been tracked down (reviewed in Nusslein-Volhard and Wieschaus, 1980), and many have been cloned. In the case of the gene, *Kruppel*, its predicted protein product turns out to be homologous to an important transcription factor in the 5S gene system of *Xenopus*. The significance of this finding is as yet unclear.

Similarly, in the *Drosophila* nervous system, the power of the genetic analysis has revealed genes which affect differentiation of parts of the ectoderm into neuroblasts. These seven genes, *Notch, almondex, big brain, mastermind, neuralized, Delta* and *Enhancer of split*, all cause specific hypertrophy of the nervous system at the expense of epidermal structures (Lehmann *et al*., 1981). They all have affects on a switching process in which cell fates are altered from one state to another. The important point to note here is that by specifically targeting the genome for certain phenotypes, many of the genes affecting particular processes can be uncovered, a situation not yet feasible in mammals.

The message from *Drosophila* is that executive developmental decision-making genes are relatively few in number (Raff and Kaufman, 1983), maybe of the order of but a few hundred. Since this organism has about 10,000 transcription units, it is apparent that the bulk of the genes provide the metabolic backup for the decisions of the executive genes. Thus it is crucial, in the case of a mammal, where there are maybe 50,000 genes or so, to try and sift those few of developmental importance from the remainder. Deciding which of these may have been instrumental in the generation of evolutionary novelty is then the crux of the problem.

Genomic Raiding

Certain key decision-making genes in Drosophila contain a small sequence termed the homeo-box whose exact function is yet to be elucidated. When cloned probes of this sequence are used to challenge other organisms to see if they have sequences homologous to this probe, it is found that they do. Equivalent sequences are found in man, mouse, chicken and the amphibian *Xenopus* (Shepherd *et al*., 1984). In *Xenopus* the sequence is expressed during early development beginning at gastrulation, just as in *Drosophil*a (Carrasco *et al*., 1984). Whilst the significance of such homologies is not yet totally clear, it is obvious that much of relevance to higher eukaryotes can be gleaned from the cross-raiding of vertebrate and invertebrate genomes. For example, the calmodulin and pyrimidine biosynthesis genes of the rat have been used to isolate the corresponding genes from *Drosophila*. The oncogenes of certain vertebrate retroviruses have also been found in the fly. A heat shock gene of *Drosophila*, hsp70, has nearly 80 percent homology with its counterpart in the chicken. Thus there is clearly a component of eukaryotic genomes which through sheer conservation over evolutionary time, will be sufficiently similar between diverse organisms to allow genes to be isolated through cross homologies. What roles such highly conserved genes will play in the generation of evolutionary novelty is unclear, since the more conserved they are, the lower is their probability of contributing to innovation. However, one of the significant questions as far as evolutionary emphases are concerned is whether genes or pieces of genes which have such obvious homologies are performing the same suite of functions in the different organisms, or whether in fact their functional attributes have largely diverged. This is as yet a largely unknown area.

The invertebrate molecular data base has provided the first glimpses into how early developmental events are controlled at the level of the DNA. However, in spite of the breathtaking insights that possession of some of the cloned genes has given us, the real hurdles are yet to come. For example, when we consider a series of events as apparently simple as the dichotomy between feather or skin formation in vertebrates, we need to approach the problem at the level of *cell* interactions in the developing epidermis (Oster and Alberch, 1982). In this system, if the epidermal layer *evaginates*, scales or feathers are formed, whereas if the layer *invaginates*, skin and hair are the consequence. The molecular events at the *cellular* level which control the folding of the epidermal layer and the epidermal-dermal interactions are clearly the places on which to focus experimentally. This, however, involves an understanding of the underlying basis of cell *shape* changes and reminds us how little progress we have made in bridging the gap between DNA molecules and phenotype. In lower organisms, we are just at the beginning of determining which developmental gene circuits affect morphology. In mammals, our molecular data base is in a rudimentary state.

Nervous Systems

There is one particular facet of higher organisms, namely the nervous system, whose molecular development has yet to be explored in an evolutionary context. Prime attention has always focussed on external morphologies and with good reason. Moreover, since the limelight has fallen on the contributions of various morphological novelties such as the development of jaws, or the appearance of feathers, to evolutionary 'progress', the contribution of the nervous system has been less appreciated.

However, when the vertebrate brain is examined molecularly, a few shocks are already apparent. The total informational content of different messenger RNAs which are transcribed in tissues such as the liver and kidney, averages 30 million nueleotides. The informational content of the brain, however, is in excess of 110 million nucleotides (reviewed in Davidson and Britten. 1979). The cloning data of Milner and Suteliffe (1983) are impressive (Table 3). Of 191 clones selected at random from a eDNA library of the rat brain, more than half are absolutely brain specific with the remainder being differentially expressed in the brain, but also occurring in liver and kidney. In mammals, most of the genes may well be concerned with neural functioning! If this turns out to be so, then some of the cornerstones of evolution may not have been in those morphological areas that have received the most attention, but may well have been in producing the sophisticated computers that power our behavioural repetoires. This is indeed food for thought.

TABLE3
PARTITTIONING AND EXPRESSION OF POLYADENYLATED MESSENGER RNAs IN THE RAT BRAIN (from Milner and Sutcliffe, 1983)

	Per Cent	
CLASS 1	18	Present equally in Brain, Liver and Kidney
CLASS 2	26	Differentially expressed in Brain, Liver, Kidney
CLASS 3	30	In brain only
CLASS 4	26	In brain only; very rare brain mRNAs

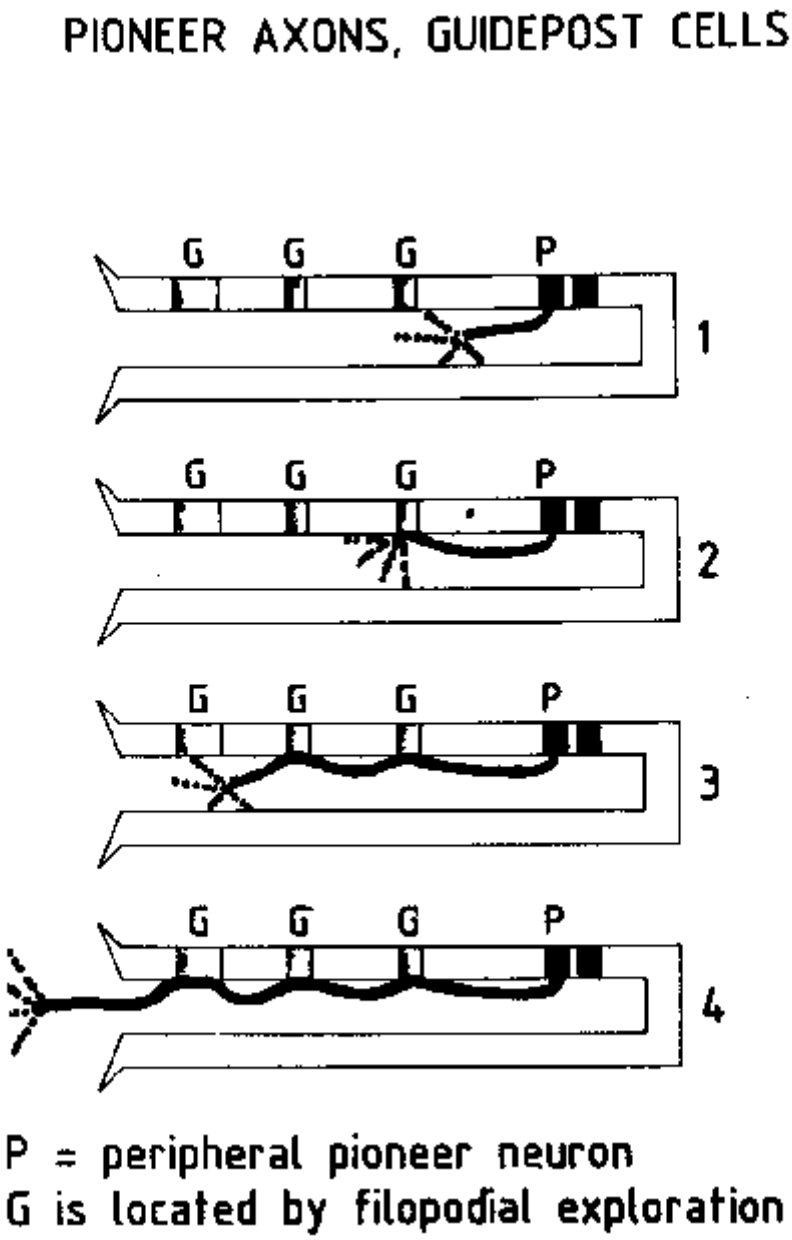

Fig. 3

It is easy to be overawed by the molecular complexity of the functioning mature brain. However, this pales into insignificance when one considers how the brain is set up in the first place. The problems of how neurons make correct synaptic connections is an old one, but only now are some molecular approaches being seriously explored. Once again it is the invertebrates, with their less complex systems, that are providing the first inroads. A simple example reveals the complexity of the problem. In the developing grasshopper limb, certain cells, destined to be pioneer neurons, trace a pathway from the periphery of the developing limb to the developing central neurons system (Figure 3). They accomplish this by a process of filopodial exploration. When they come in contact with a guidepost cell, which presumably has characteristic cell surface properties relative to its neighbours, the correct direction of further exploratory movements is assured (Goodman and Bastiani, 1984; Goodman *et al.*,1984; Thomas *et al.*, 1984). Following such directional pathfinding, the pioneer neuron must still have the information to

generate a correct synaptic connection, at its final destination. Imagine the molecular information just to wire up the 'simple' 100,000 neuron computer of *Drosophila*, and compare it to the wiring task of the human brain, which has approximately 100,000,000,000,000 synaptic connections. When one considers the evolution of such biological computers, the problems of the generation of novel evolutionary wiring pathways seem somehow just as important as the evolution of morphological novelties, maybe in some cases even more so.

THE GENOME IN EVOLUTION

Facts and Theory

We return yet again to asking how morphological or neuronal novelty is generated and spread throughout a population. Has there been a significant evolution of new genes, or have the old genes, or duplicate variants of them been assembled into new circuits, which have themselves now led to new morphologies? We still do not know. The problem is even harder than this since, as we saw earlier, eukaryotic genomes are not simple collections of single genes, but consist also of multigene families. Neo-Darwinian theory has held that the spread and fixation of variants through a population is largely the consequence of natural selection of single gene systems with most evolutionary changes being thought to arise as adaptive responses to the environment.

The theory is experiencing great difficulty with multigene families even in their simplest form. Clearly when a variant family member arises, selection has to survey the family as a whole. Furthermore the theory misses the essential point that is so aptly summarized in the following, 'Selection may account for the survival of the fittest, but fails to account for the arrival of the fittest'. It is in the latter category that modern molecular biology is having its greatest impact.

However, a far more interesting and novel mode of change, the process known as molecular drive has been proposed and analyzed by Dover (1982, 1986) and it is especially applicable to multigene families. This mode of change is predicated on the observed molecular data which stem from the flux that goes on within a genome. DNA sequences promote their own amplification, dispersion, conversion and so forth and can obviously cause recurrent changes within a genome. This genomic turnover in multigene and multisequence families may well be the prime mover in evolution, in that DNA turnover mechanisms cohesively altering the genomes of a population could result in evolutionary novelties (Dover and Flavell, 1984; Dover and Tautz, 1986). Given all of this, however, we still have been unable to determine the underlying causes of morphological differentiation between the species we began with, namely whales, mice and bats. We will be unlikely to do this until we move our efforts away from the abstractions of neoDarwinian theory, and return to molecular embryology. Once again, as pointed out long ago by Bateson in 1922, it is "*in embryology (that the) quintessence of morphological truth* (is) *most palpably presented*."

The evolutionary problem comes home with even more force when we consider fossil sequences and try to speculate on the underlying bases of change. While some groups such as the Proboscidea, show a progressive and gradual series of morphological changes, others appear abruptly in the fossil record. Thus turtles appear fully formed as do pterodactyls and early amphibians (Lull, 1940). We will ultimately need to decide whether the underlying genomic changes were of a major kind or not. The root of our current problems is out inability to determine whether any morphological change is initially the result of a change in a single gene, a multigene or multisequence family, or whether it is a novel gene.

CONCLUSIONS

Let me summarize then some of what has been learnt about the structure and function of eukaryotic genomes and how this information has influenced our perspectives on the underlying bases of phenotypic change:

1. The study of genomic DNAs has revealed the existence of junk DNA, mobile elements, pseudogenes, split genes, multigene families, and sundry DNA turnover processes. We are still evaluating which of these are more or less important for phenotype.
2. The ubiquity of multigene families, when considered together with the cellular enzymic machinery, means that there is enormous potential for the generation of evolutionary novelty from within. There is now a distinct possibility that evolutionary changes stemmed more from within the genome than from external forces.
3. The demonstration of key executive genes in *Drosophila*, together with their homologous sequences in higher eukaryotes, has yielded a clearer view of the circuits involved in particular morphological and neuronal programmes. It is now necessary to determine how easy or how difficult it is to perturb such circuits. This will yield some idea of what it costs to evolve in a genetic sense. It is further necessary to determine in which directions a circuit can be modified in order to be able to guess how a given structure can alter.

The mechanistic side of evolution is now much more exciting than it has ever been, with the very real prospect that we shall soon be able to understand what makes the genome tick in a structural sense. Then we can go on and think about the harder problems, protein-protein interactions, and ultimately cell-cell interactions through time. Then and only then will we be able to guage how difficult a problem we have set ourselves, in enquiring about the molecular bases of morphological novelty and the molecular costs of neuronal novelty.

It ought to be obvious that the molecular analysis of development offers some hope of unravelling the origins of phenotypic structural change. The conventional approaches of 'explaining' the origins of morphological novelties, via the neo-Darwinian selectionist-neutralist debate is a sterile one, because it is not addressed to the genesis of the changes themselves. Levin (1984) has pointed out that "*the neutralist-selectionist controversy is more a product of the sociology of science (the two camp advocacy approach) than its substance*." Whilst this controversy has dominated molecular population genetics since the early 1970s, its contribution, when measured against the burgeoning molecularly precise data bases, has not only been oblique to central evolutionary issues, but in the light of multigene families and molecular drive, may well turn out to have been an irrelevant side issue.

REFERENCES

Amstutz, H_ Munz, P_ Heyer, W.D., Lenpold, U. and Kohli, J., 1985. Concerted evolution of tRNA genes: intergenic conversion among three unlinked serine tRNA genes in *S. pombe. cell, 40*, 879-886.

Anderson, K.V. and Nusslein-Volhard, C., 1984. Information for the dorsal-ventral pattern of the *Drosophila* embryo is stored as maternal mRNA. *Nature, 311*, 223-227.

Baltimore, D., 1985. Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome. *Cell, 40*, 481-482.

Bateson, W., 1922. Evolutionary faith and modern doubts. *Science, 55*, 55-61.

Carrasco, A.E., McGinnis, W., Gebring, W.J. and De Robertis, E.M., 1984. Cloning of an *Xenopus laevis* gene expressed during early embryogenesis coding for a peptide region homologous to *Drosophila* homeotic genes. *Cell, 37*, 409-414.

Conner, W.G., Hinegardner, R. and Bachmann, K_ 1972. *Experientia, 28*, 1502-1504.

Darnell, J.E., 1985. RNA. *Sci. Amer. 253*, 54-64.

Davidson, E.H. and Britten, R.J., 1979. Regulation of gene expression: possible role of repetitive sequences. *Science, 204*, 1052-1059.

Doolittle, R.F., 1985. Proteins. *Sci. Amer., 253*, 74-83.

Doolittle, W.F. and Sapienza, C., 1980. Selfish genes: the phenotypic paradigm and genome evolution. *Nature, 284*, 601~603.

Dover, G.A., 1982. Molecular drive: a cohesive mode of species evolution. *Nature, 299*, 111-117.

Dover, G.A., 1986. Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends in Genetics, 2*, 159-165.

Dover, C. A. and Flavell, R.B., 1984. Molecular coevolution: DNA divergence and the maintenance of function. *Cell, 38*, 622-623.

Dover, G.A. and Tautz, D., 1986. Conservation and divergence in multigene families: alternatives to selection and drift. *Phil. Trans. R. Soc. Lond. B, 312*, 275-289.

Dowsett, A.P. and Young, M.W., 1982. Differing levels of dispersed repetitive DNA among closely related species of *Drosophila. Proc. Natl. Acad. Sci., 79*, 4570-4574.

Felsenfeld, C., 1985. DNA. *Sci. Amer. 253*, 44-53.

Fry, K. and Salser, W_ 1977. Nueleotide sequences of HS-a satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. *Cell, 12*, 1069-1084.

Gehring, W., 1985. The molecular basis of development. *Sci. Amer., 253*, 136-146.

Goodman, C.S. and Bastiani, M.J., 1984. How embryonic nerve cells recognize one another. *Sci. Amer., 251*, 50-58.

Goodman, C.S., Bastiani, M.J., Doe, C.Q., Dulac, S., Helfand, S.L., Kuwada, J. and Thomas, J.B., 1984. Cell recognition during development. *Science, 225*, 1271-1279.

Hinegardner, R., 1974. Cellular DNA content of the Mollusca. *Comp. Biochem. Physiol., 47A*, 447-460.

Hogness, D.S., Lipshitz, B.O., Beachy, P.A., Peattie, D.A., Saint, R.B., Coldschmidt-Clermont, M., Harte, P.J., Cavis, E.R. and Helfand, S.L., 1985. Regulation and Products of the ubx domain of the Bithorax Complex. *Cold Spring Harbor Symposia on Quantitative Biology, 50*, 181-194.

Horvitz, H.R., Sternberg, P.W_ Creenwalk, I.S., Fixen, W. and Ellis, H.M., 1983. Mutations that affect neural cell lineages and cell fates during the development of the Nematode *Caenorhabditis elegans*. *Cold Spring Harbor Symposia on Quantitative Biology, 48*, 453-463.

Hunkapiller, T., Huang, H., Hood, L. and Campbell, J.H., 1983. The impact of modern genetics on evolutionary theory, *in* PERSPECTIVES ON EVOLUTION pp 164-189. R. Milkman (Ed.). Sinauer Associates, Sunderland, Mass.

Jackson, I., 1986. Solid Foundation for Developmental Biology. *Trends in Genetics, 2*, 193.

John, B. and Miklos, G.L.G., 1987. THE EUKARYOTE GENOME IN DEVELOPMENT AND EVOLUTION. George Allen and Unwin (in press). London.

King, J.L. and Jukes, T.H.J969. Non-Darwinian evolution. *Science, 164*, 788-798.

Lehmann, R., Dietrich, V., Jimenez, F. and Campos-Ortega, J.A., 1981. Mutations of early neurogenesis in *Drosophila*. *Wilhelm Roux's Archives, 190*, 226-229.

Levin, B.R., 1984. Science as a way of knowing Molecular Evolution. *Amer. Zool., 24*, 451-464.

Lull, R.S., 1940. ORGANIC EVOLUTION. Maemillan Publishing Co., New York.

Miklos, G.L.G., 1985. Localized highly repetitive DNA sequences in vertebrate and invertebrate genomes, *in* MOLECULAR EVOLUTIONARY GENETICS, 241-321, R.J. MacIntyre (Ed.). Plenum Press, New York.

Miller, C.A. and Benzer, S., 1983. Monoclonal antibody cross-reactions between *Drosophila* and human brain. *Proc. Natl. Acad. Scl U.S.A., 80*, 7641-7645.

Milner, R.J. and Sutcliffe, J.G., 1983. Gene expression in rat brain. *Nucleic Acids Res.,11*, 5497-5520.

Nusslein-Volhard, C., 1979. Maternal effect mutations that alter the spatial co-ordinates of the embryo of *Drosophila melanogaster*, *in* DETERMINANTS OF SPATIAL ORGANIZATION, pp 185-211. S. Subtelny and I.R. Korisberg, (Eds.). Academic Press, New York.

Nusslein-Volhard, C. and Wieschaus, E., 1980. Mutations affecting segment number and polarity in *Drosophila*. *Nature 287*, 795-801.

Ohno, S., 1970. So much "junk" DNA in our genome *in* EVOLUTION OF GENETIC SYSTEMS, pp 366-370, H.H. Smith (Ed.). Gordon and Breach, New York.

Ohno, S., 1982. The common ancestry of genes and spacers in the euchromatic region: *omnis ordinis hereditarium a ordinis priscum minute Cytogenet. Cell. Genet. 34*, 102-111.

Ohta, T., 1983. On the evolution of multigene families. *Theoretical Population Biology 23*, 216-240.

Orgel, L.E. and Crick, F.H.C., 1980. Selfish DNA: the ultimate parasite. *Nature, 284*, 604-607.

Oster. C. and Alberch, P., 1982. Evolution and Bifurcation of Developmental Programmes. *Evolution 36*, 444-459.

Plechaczyk, M., Blanchard, J.M., Riaad-El Sabouty, S., Dani, C., Marty, L. and Jeanteur, P. 1984. Unusual abundance of vertebrate 3-phosphate dehydrogenase pseudogenes. *Nature, 312*, 469-471.

Raff, R.A. and Kaufman, T.C., 1983. EMBRYOS, GENES AND EVOLUTION. Macmillan Publishing Co., New York.

Robertson, M., 1986. The proper study of mankind. *Nature, 322*, 11.

Rubin, C.M., 1983. Dispersed repetitive DNAs in *Drosophila*, *in*: MOBILE GENETIC ELEMENTS. J.A. Shapiro, (Ed.). Academic Press, New York.

Shepherd, J.C.W., MeGinnis, W., Carrasco, A.E., De Robertis, E.M. and Gchring, W.J., 1984. Fly and frog homeo domains show homologies with yeast mating type regulatory proteins. *Nature, 310*, 70-71.

Singer, M.F., 1982. Highly repeated sequences in mammalian genomes. *Int. Rev. Cytol. 76*, 67-112.

Tautz, D., Trick, M. and Dover, G.A., 1986. Widespread cryptic simplicity in DNA: a major source of genetic variation. *Nature* (in press).

Thomas, J.B., Bastiani, M.3., Bate, M. and Goodman, C.S., 1984. From grasshopper to *Drosophila*: a common plan for neuronal development. *Nature, 310*, 203-207.

Varmus, H.E., 1983. Retroviruses, *in*: MOBILE GENETIC ELEMENTS, pp 411-503. J.A. Shapiro, (Ed.). Academic Press. New York.

Weinberg, R.A., 1985. The molecules of life. *Sci. Amer. 253*, 34-43.

Return to Top