Cross-species gene transfer: a major factor in evolution?

Michael Syvanen

Recent experimental evidence is discussed in terms of the speculative theory that the transfer of genes from one species to another may be a major factor in macroevolutionary change.

Recently I presented a very speculative theory which suggests that the transfer of genes from one species to another may play a significant role in evolution[1]. Although this theory remains as speculative as the day I wrote it, a number of unexpected experimental results have just been published which were either predicted by the theory or are pertinent to it. Here I briefly summarize the central idea, and review the relevance of three new groups of results: the divergences from the uniform genetic code in ciliated protozoa, new findings on retroviruses, and new findings on recombinogenic aspects of introns. I should add that no author other than myself who is mentioned in this review is in any way associated with, responsible for, or should be caused to suffer by association with these unorthodox ideas – unless, of course, they choose to.

In brief, I argue that the existence of the biological unities, from the uniform genetic code to the cross-species similarity of the stages of embryological development are maintained through the action of cross-species gene transfer. The uniformity of the genetic code would have been maintained as life evolved from its common ancestor because it allows organisms to decipher and use genetic information transposed from chromosomes of foreign species, and the shared sequence of embryological development within each phylum would have been similarly maintained because it would allow an organism to integrate this transposed information, particularly when it affects complex morphological traits. In short, if the rate of gene transfer is high compared with the rate of evolution, then we can view living organisms as being continuously exposed to foreign genes that may enter their germ line and which may, on occasion, become fixed by natural selection should a useful trait be involved. Selective pressure for maintenance of the biological unities, as well as other aspects of the cross-species transfer mechanism, is provided on those occasions, over evolutionary time, when such a foreign gene is selected. If a lineage were to drop out of the universal gene exchange the consequence would be to slow the rate at which adaptive changes are introduced.

Obviously some cross-species gene transfer does occur, especially in the prokaryotic world. Environments with intensified selection – for example, those following antibiotic use by man – yield bacterial populations that show much evidence of cross-species gene transfer. There are also a limited number of examples in the eukaryotic world, but just how extensive they are remains unknown[2~3]. According to

my arguments, cross-species gene exchange must be frequent enough to be a major factor in macroevolutionary change, not just an occasional occurrence. If that is so, according to this theory, then it seems probable, if not necessary, that exchange of genetic information among different species should influence the measured rate of nucleotide sequence divergence at neutral positions; that is, affect the rate of the molecular clock. The transfer of a polynucleotide sequence from the chromosome of one species to that of another would have the effect of producing an apparent convergence of those two species. If species are linked, with varying degrees of intensity by cross-species gene exchange, then the effect would be to decrease their rate of divergence, to render the molecular clock slower than the actual rate of neutral-site evolution. Alternatively, if a lineage were to drop out of the universal gene exchange, then the effect would be to accelerate its rate of sequence divergence relative to other lineages.

The almost universal genetic code

The unity of the genetic code has generally been construed as evidence for the common descent of all life as formulated by the frozen accident theory. I have argued that the code's unity may possibly reflect more than just evidence for common descent but may also reflect the existence of ongoing selective pressure for unity *per se.*

My argument against the frozen accident theory was based on the principle that drift in the code should be possible by at least two means. One begins with the existence of the tRNA nonsense suppressors that can insert amino acids instead of terminating at chain termination codons. tRNA suppressors are not just laboratory creations; it has been known for some time that the tRNA suppressors are found in natural bacterial isolates. Marshall and Levy[4] found, when they examined this point systematically, that 7% of a sample of natural *Escherichia coli* isolates contained suppressors for one specific nonsense mutation. For bacterial populations, this shows that there is a large pool of suppressor tRNAs that could, if established with the right chain terminating or nonsense mutation, become essential. Should such a nonsense suppressor become fixed in a population, then there would be selective pressure to alter translational stop signs that used the codon of that particular nonsense suppressor. Thus, I argued that the conversion of a nonsense codon into a sense codon was possible, at least in principle, and to the extent that it did not occur, there was probably a reason. The second means by which I suggested that the code could change would be by random drift in codon usage frequencies; it seems likely that zero boundaries would occasionally be encountered.

The recent findings of the possible conversion of chain terminating codons into sense codons in the ciliated protozoans is consistent with the hypothesis that drift in the code is possible, with reference to nonsense suppressors. Four different groups have reported that in ciliated protozoans, translational stop codons from the previously universal code are used to code amino acids. In addition, a similar situation has been encountered in mycoplasma[5]. Horowitz and Gorovsky[6] have found UAA used four times in two different histone H3 genes from *Tetrahymena.* The case is convincing that these genes are not pseudogenes. The assignment of glutamine was made by comparing the protein sequence, to the DNA sequence. Caron and Meyer[7] sequenced the gene for one of the G surface antigens in *Paramecium.* No open reading frame was evident unless UAG and UAA were counted as sense codons. By sequencing the mRNA, they showed that this sequence was, in fact, transcribed *in vivo* and, further, this UAG-, UAA-containing open reading frame predicted the correct amino acid composition. When portions of this same reading frame were fused to the N-terminal portion of β-galactosidase (to create a protein fusion) and expressed in *E. coli,* G antigens were synthesized. Based on amino acid composition, Caron and Meyer surmise that the UAA and UAG code for glutamine or glutamic acid. A similar situation was encountered when Preer *et al.*[8] cloned and sequenced parts of one of the immobilization antigen genes from *Paramecium.* They showed that the transcribed sequences had only UAG- and UAA-containing open reading frames and that one of them gave the correct amino acid composition. The fourth example is provided by the complete sequence of the α-tubulin gene[9]. One reading frame that had a UAA yielded an amino acid sequence that showed homology with the rat and chicken α-tubulins. This UAA codon was homologous to a conserved glutamine site. In addition to these observations, all seven genes sequenced to date (three – two *Tetrahymena* histone genes and an *Oxytricha* actin gene – have no UAA or UAG codons within the structural gene) terminate in UGA.

From examining these data, it appears that there might be additional codon differences. The ciliated protozoans appear to have unusually strong biases in codon usage frequencies at other positions. One of the more striking cases is that 91 of 94 arginine codons so far encountered are AGA. Another is that in the first 1623 codons sequenced, the codonsGUG, GGG, CGU, CGA and CGG have not yet been encountered. These examples raise the possibility that another major change may have occurred; namely complete loss of codons.

The ciliated protozoans differ from other organisms not only in their genetic code. There has been an awareness for some time from protein sequence data that yeast, plants and animals show greater similarity among themselves than they do with the ciliates. Glover and Gorovsky'° noticed this pattern with the histone sequences. The same pattern has appeared with actin, tubulin and cytochrome- *c.* The ciliates differ so much that molecular clock calculations place divergence of *Tetrahymena,* for example, back to more than 3 billion years ago; a clearly absurd result. Glover and Gorovsky have therefore suggested that the ciliates did diverge very early but that, in addition, their molecular clock moves at a higher rate than that of other organisms. Now, with the finding that the code for these organisms is different, Horowitz and Gorovsky[6] suggest that the divergence of the ciliates may have in fact occurred before the contemporary codes became 'frozen'.

If we assign a significant role to cross-species gene exchange in evolution, an alternative interpretation for these results can be posed. It is not necessary to place the divergence time for the ciliates back to the time when the genetic code arose if we accept that the contemporary genetic codes are not entirely frozen. It seems more likely that ancestral ciliates once had the universal code but that their usage drifted. Since the dilates read UAA and UAG as amino acid codons, and

never as stop codons, it would presumably make it difficult for ciliates to read large classes of genes introduced from species that use the 'universal' code. This event may have signaled the partial removal of the ciliates from the universal gene exchange, with the consequence that, according to the theory, the rate of divergence of the ciliates from other life would appear fast when compared to the divergence rate, among other organisms. Note, however, that the rate of evolution would not be accelerated by this isolation, rather the rate of divergence from contemporary species would appear accelerated.

If my explanation concerning the origin of the different code in the ciliates is correct, it does not, of course, explain how they could have escaped from universal gene exchange. Perhaps certain organisms are so well adapted, so isolated, or otherwise unchallenged, over sufficiently lengthy periods, that improvement is not necessary for survival. In other words, perhaps the genetic code can diverge in certain slowly evolving species. Perhaps, in other cases, elements are so perfect that improvement in them would not help an organism to compete. This has been my suspicion in the case of the changed genetic code in mitochondria; here the laws of thermodynamics preclude increased efficiency in mitochondrial function, thus freeing them from the constraints that cross-species gene exchange imposes. I want to emphasize that I am not suggesting that the findings concerning divergence from the code offer direct evidence for the evolutionary role of crossspecies gene exchange. My interpretation does, however, offer a parsimonious explanation for both the divergence in genetic code as well as the apparent increased rate of the ciliates' molecular clock. In addition, it raises the possibility that the genetic code is not entirely frozen, thus making the re-examination of the explanation for the unity of the genetic code seem a little less gratuitous (as one critic euphemistically put it) than it did until this data became available.

Retroviruses and transposable elements

If cross-species genomic sequence exchange does occur among higher eukaryotic species, the mechanism of transmission will most likely turn out to be viruses. A variety of observations during the past ten years indicates that a particularly attractive viral vector may be the retroviruses. These are the single-strand, RNA-transforming viruses with reverse transcriptase activity. The viruses of this class share a similar chromosome – three genes designated *gag, pol* and *env* that are flanked by a directly repeated sequence, called *ltr*. These viruses may or may not contain an oncogene. This is the organization of both the RNA sequences in the virion and the DNA sequences when integrated in some host.

It has been known for some time that nucleotide sequences found in retroviruses are present in eukaryotic hosts. In fact, lateral gene transfer was first indicated by Benveniste and Todaro" in their study of the degree of homology to a specific retroviral sequence within the chromosomes of a series of mammals. They found that sequences homologous to a cat leukemia virus were found in the chromosomes of domestic cats and some of their close relatives and also in old world monkeys. However, these sequences were absent from other members of the primate or cat families. The only reasonable explanation for this result was that a retrovirus carried through the chromosome of one of the lineages was transmitted to the chromosome of another. With the more recent finding that retroviral oncogenes have cellular counterparts, it may be that such cellular genes were being compared. Whatever, the evidence is reasonable for a lateral nucleotide sequence transfer that was most likely mediated by a retrovirus.

Retroviruses have additional properties that mark them as likely vectors for the transmission of nonviral genes. Ikawa *et al.12* showed that when Friend leukemia virus was grown on reticulocytes that were actively transcribing the globin genes, some of the resulting virus particles had packaged ß-globin mRNA. If we consider that most viruses have host-ranges that extend beyond a single species, it seems fairly likely that such globin mRNAs could eventually be deposited into cells of a foreign species. I know of no example where the genetic information from such a foreign mRNA became integrated into a cell's chromosome, but we can surmise that the mechanism exists for such an integration event; the structure of many pseudogenes implies as much. Two early examples are the structure of the mammalian α-globin [13,14] and a small nuclear RNA pseudogenes [15], both of which contain the linear sequences of an mRNA. including the absence of a complete promoter, deleted introns and the 3' poly(A) region. The only reasonable explanation for these structures is that the respective mRNAs were reversely transcribed by reverse transcriptase and that the resulting DNA molecules were inserted into a germ-line chromosome. It may be that the majority of pseudogenes arose by such a process, as evidenced by the structures of five different mammalian cytochrome- *c* pseudogenes'[6].

The involvement of retroviruses in the evolution of their hosts is becoming apparent in another area. It has often been suggested that transposable elements play a major role in the evolution of their hosts [17,18]. Among eukaryotes, two of the better understood transposable elements, are *Ty* in yeast and *copia* in *Drosophila.* These were both seen initially as classes of middle-repetitive chromosomal DNA whose members are responsible for causing insertion mutations. The complete nucleotide sequence of *Ty* (Refs 19 and 20) and *copia* (Ref. 21) show that they are related to the retroviruses; they are homologous to vertebrate retroviruses and show the same *ltr, gag, pol, env, ltr* organization. In addition**,** Boeke *et al.22* have provided direct evidence that *Ty* transposition proceeds through an RNA intermediate. Emori *et al.*[23] show that *copia* and an RNA isolated from a *Drosophila* virus-like particle are very closely related. Thus, it seems fairly certain that these transposable elements were once, and may still be, retroviruses. It is very difficult to imagine how this retroviral gene organization was conserved within the lineages leading to yeast and *Drosophila;* the most likely explanation is that these elements transferred into the lineages from a common retroviral pool after the lineages diverged.

Why introns?

The final area of new data that might be relevant to a theory of cross-species gene exchange concerns the function of introns. Recently I proposed[24] that one of the functions of introns is to promote gene conversion between homologous genes from different species.

Gilbert [25, 26] has suggested that the functional role of introns is evolutionary; they create new arrangements of protein functional domains that result in new proteins. Thus, introns evolved in the course of evolving new proteins; otherwise, they have no necessary role to play in gene expression or chromosome structure. This idea has found support[26], though in its simplest form some questions remain unanswered. For example, once a functional protein evolved, what pressure maintains the intron? Efficient mechanisms apparently exist to delete introns from genes[27]. I extended the exon-shuffling model to suggest that introns can also promote recombination events between homologous introns and their adjacent exon sequences. This proposal arose from the observation that the exon sequences flanking the two introns in the mammalian ß-globin gene are highly conserved at silent positions in four mammals[24]. The conserved exon regions extend over 63 and 51 nucleotides in the region of the two introns. This exceeds the number of exon nucleotides thought to be needed for the processing of the introns from pre-mRNA. An explanation I offered for this result was that these highly conserved exon regions were not conserved by functional constraint but were rather homogenized by cross-species gene conversion events. This led to the suggestion that these recombination events were initiated by the introns and, further, that introns have evolved, in part, because they promote the exchange of genetic information between different species.

Though this model seems *ad hoc*, if not far-fetched, it does at least offer a straightforward prediction – that introns promote gene conversion. Evidence in support of this is accumulating. Yeast mitochondria encode a 21S rRNA gene that contains an intron and variants of it exist that lack the intron. Two groups [27, 28] have shown that crosses between intron-plus and intron-minus strains result in an efficient, non-reciprocal, gene conversion event that renders the intron-minus copies intron-plus[27, 28]. It therefore appears that the intron sequence promotes a crossing-over event that results, in effect, in integration of the intron from one mitochondria genome to another. The crossing-over event necessarily occurs through exon sequences. Other evidence supporting gene conversion events mediated by introns is seen from an analysis of the ß-globin gene families. Slightom *et al.*[29] reported evidence for gene conversion between the A and G fetal y-hemoglobin loci in humans and they proposed that a T-G rich region in the large intron initiated the conversion. Scott *et al.30* report a similar situation for the γ loci in gorilla. Erhart *et al.31* have presented a similar story for the adult ß-globin gene family in mice. Mice have four adult ß-globin genes (two genes actually, each with two 'haplotypes'), three of whose evolution is concerted; i.e. appear to convert among themselves. The fourth does not. And what distinguishes the fourth is that its large intron is not homologous to those of the other three. In addition, this eccentric large intron lacks a T-G-rich region that these authors propose promotes the gene conversion events that occur among the other three.

This kind of recombination activity, if associated with introns in general, could account for both the conserved regions in exons that I documented previously, and conservation of the introns themselves. Furthermore, if this phenomenon is general, a means for carrying out locus directed gene conversion events using recombinant DNA should now be accessible for experimental manipulations.

References

1 **Syvanen, M. (1985)** *j. Theor.* ***Biol.*** **122,** 333—343
2 **Lewin, R. (1982)** *Science* **217,** 42-43
3 **Lewin, R. (1985)** *Science* **227,** 1020
4 **Marshall, B. and Levy, S. B. (1980)** *Nature* **286.** 524—525
5 **Yamao, F.** *et al.* **(1985)** *Proc. Natl Acad. Sci. USA* **82,** 2306-2309
6 **Horowitz, S. and Gorovsky, M. (1985)** *Proc. Natl Acad. Sci.* ***USA*** **82,** 2452—2455
7 **Caron, F. and Meyer, E. (1985)** ***Nature*** **314,** 185—188
8 **Preer, J. R.** ***et al.*** **(1985)** *Nature* **314,** 188—190
9 **Helftenbein, E. (1985)** *Nucleic Acids Res.* **13,** 415
10 **Glover, C. and Gorovsky, M. (1979)** *Proc. Natl Acad. Sci.* ***USA*** **76,** 585—589
11 **Benveruste, R. E. and Todaro, G. J. (1974)** Nature **252,** 456—459
12 **Ikawa, Y., Ross, J. and Leder, P. (1974)** *ProcNatlAcad. Sci USA* **71**, 1154—1158
13 **Leder, P.** ***et al.*** **(1980)** *Science* **209**, 1336—1342
14 **Vanin, E. F.** ***et al.*** **(1980)** ***Nature*** **286**, 222—226
15 **Van-Arsel, S. W.** ***et al.*** **(1981)** ***Cell*** **26,** 11—17
16 **Scarpulla, R. C. (1984)** ***Mol. Cell. Blot.*** **4,** 2279—2288
17 **Campbell, A. (1983)** In: *Evolution* ***of*** *Genes and Proteins* **(Nei, M. and Koehn, R., eds)**, pp. 258-279, Sinauer
18 **Syvanen, M. (1984)** *Annu. Rev. Genet.* **18**, 271—293
19 **Clare, J. and Farabaugh, P. (1985)** *Proc. NatlAcad. Sci. USA* **82,** 2829—2833
20 **Hauber,J., Nelböck-Hochstetter and Feldmann, H. (1985)** *Nuc!eic Acids Res.* **13,** 2745—2758
21 **Saigo, K.** ***et al.*** **(1984)** *Nature* **312,** 659—661
22 **Boeke, J. D.** ***et al.*** **(1985)** ***Cell*** **40,** 491—500
23 **Emori, Y.** ***et al.*** **(1985)** *Nature* **315,** 773—776
24 **Syvanen, M. (1984)** *J. Theor. Biol.* **107,** 685-696
25 **Gilbert, W. (1978)** ***Nature*** **271,** 501
26 **Gilbert, W. (1985)** *Science* **228,** 823
27 **Jacquier, A. and Dujon, B. (1985)** ***Cell*** **41,** 383—394
28 **Macreadie, I. G.** ***et al.*** **(1985)** ***Cell*** **41,** 395-402
29 **Slightorm, J. L. (1980)** ***Cell*** **21,** 627—638
30 **Scott, A. E. (1984)** ***Mol. Biol. Evol.*** **1,** 371—389
31 **Erhart, M., Kimberlee, S. and Weaver, S. (1985)** ***Mol.*** *Biol.* ***Evol.*** **2,** 304—320

Michael Syvanen is ***at the*** *Department* ***of*** *Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA 02115, USA.*

Tarr, G.E. & Fitch (1976) Biochem J. **159**, 193-199.
cytochrom *e* - C from tetrahymen