Perspective

Retroposons—Seeds of Evolution

JÜRGEN BROSIUS

THE QUESTION OF WHY MAMMALIAN GENOMES ARE "BURdened" with a large number of nonfunctional genes has puzzled molecular and evolutionary biologists. These genes can arise in two ways: (i) by recombination, in which the duplicated gene acquires defects and eventually ends up as an inactive pseudogene or (ii) by retroposition, in which an RNA species is reverse transcribed into DNA copies and dispersed in the genome (*1, 2*). The latter mechanism yields retroposons, a term that includes retropseudogenes (one to several hundred copies per genome) and short or long interspersed elements (up to ~500,000 copies per genome). The hallmarks of a retroposon are the lack of introns, a poly(A) tract at the 3′ end, and remnants of flanking direct repeats. All retroposons can act as insertional mutagens with detrimental or beneficial effects, especially when an active gene is targeted (*3*). Retropseudogenes have been viewed as dead ends of evolution and the more highly repetitive members as selfish, junk DNA that litters genomes (*4*). However, recent findings of active retroposons suggest that this view must be reappraised; retroposition may represent a dynamic route towards evolutionary progress.

Apart from the ability of retroposons to keep the genome in flux (*2*), thus favoring genetic diversity, they can be considered a shotgun approach of nature wherein the majority of these genetic elements are inactive and left to rot in the genomic soil. Nevertheless, some seeds will integrate near a fertile genomic environment, giving rise (usually after mutational alterations) to new genes or gene domains and complementing the conventional gene duplication that is essential to evolution (*5*). Retroposition may also match existing genes with new regulatory elements.

In addition to numerous defective retropseudogenes and the rarer situation where such elements are intact but apparently silent, there is evidence for expressed retroposons. An insulin gene (type I) in rats and mice appears to be a retroposon that was derived from an incompletely processed primary transcript (*2*). Other examples of genes derived by retroposition include a phosphoglycerate kinase gene (*6*) and a pyruvate dehydrogenase E1α subunit gene (7) in mice and humans. While the X-linked *Pgk-1* and *Pdha-1* genes are expressed in somatic cells and contain ten introns each, the autosomal *Pgk-2* and *Pdha-2* genes are testis-specific and have retroposon characteristics. Likewise, *Zfa* is an expressed testis-specific retroposon derived from an alternative transcript of *Zfx*, which encodes a zinc finger protein (*8*). Since *Zfa* is only found in certain mouse species and has perfect hallmarks of a retroposon, its origin must be relatively recent. Furthermore, woodchucks contain an additional form of the N-*myc* oncogene, N-*myc2,* that is a retroposon and is transcriptionally active in liver tumors (*9*). Finally, the promoter regions of human salivary amylase genes (*AMY1A, AMY1B*, and *AMY1C*) are contributed—after some modifications—by the 3′ untranslated region of a γ-actin pseudogene (*10*), which indicates that a retroposon can serve as a new regulatory element for an existing gene.

Fishberg Research Center for Neurobiology, Mount Sinai School of Medicine, New York, NY 10029.

Are inactive retroposons doomed to gradual oblivion as genomic noise? One option for reactivation is mutational removal of stop codons accompanied by creation of an active promoter, which is probably a rare event. Another is recombination of pseudogenes with other genes, as in the generation of diversity in immunoglobulin light chain sequences of the chicken (*11*) or the formation of functional mosaic genes from inactive pseudogenes for variable surface antigens in trypanosomes (*12*). Although it is unknown whether the pseudogenes involved are retropseudogenes, retroposons may represent a reservoir for the creation of new gene variants by recombination.

The presence of expressed retroposons suggests that many, if not most, intronless genes could be of retroposon origin. Fink (*13*) has speculated that almost the entire yeast genome, with its paucity of introns, consists of retroposons that have replaced the founder genes by homologous recombination. This may also have happened in prokaryotic genomes (*14*). In higher eukaryotes there are a number of intronless genes, including members of the potassium channel family and the ever-growing list of G protein–linked receptors. It is conceivable that evolutionary diversity in vertebrate nervous systems, for example, was accomplished, in part, by retroposition.

The role of retroposition in creating new protein genes should also apply to genes encoding small RNAs. For example, BC200 RNA (*15*), predominantly expressed in the primate brain, is similar to a segment of the Alu consensus sequence, and is probably one of the "successful" members of the Alu elements. This retroposon may have yielded a novel gene because of its integration near potential cell type–specific regulatory elements and the subsequent recruitment of the RNA into a function by the evolving nervous system.

Further insight into the roles of retroposition should come from investigations of large genomic sequences in various organisms and should encourage colleagues involved in the human genome project who are unenthusiastic about sequencing huge tracts of junk DNA. Parts of these sequences may be informative after all, even shedding light on the future potential of evolution. In fact, since Alu elements have become very useful for the human gene mapping (*16*) that will help to manage or even eradicate genetic diseases, Alu elements may enable us to evolve into a fitter species—an example of exaptation (*17*) at the molecular level.

REFERENCES AND NOTES

1. J. Rogers, *Int. Rev. Cytol.* **93**, 187 (1985).
2. A. M. Weiner, P. L. Deininger, A. Efstratiadis, *Annu. Rev. Biochem.* **55**, 631 (1986).
3. P. Deininger, in *Mobile DNA*, M. Howe and D. Berg, Eds. (American Society Microbiologists, Washington, DC, 1989), p. 619.
4. W. F. Doolittle and C. Sapienza, *Nature* **284**, 601 (1980); L. E. Orgel and F. H. C. Crick, *ibid.*, p. 604.
5. S. Ohno, *Evolution by Gene Duplication* (Springer, New York, 1970).
6. P. H. Boer, C. N. Adra, Y.-F. Lau, M. W. McBurney, *Mol. Cell. Biol.* **7**, 3107 (1987); J. R. McCarrey and K. Thomas, *Nature* **326**, 501 (1987).
7. H.-H. M. Dahl, R. M. Brown, W. M. Hutchison, C. Maragos, G. K. Brown, *Genomics* **8**, 225 (1990).
8. A. Ashworth, B. Skene, S. Swift, R. Lovell-Badge, *EMBO J.* **9**, 1529 (1990).
9. G. Fourel *et al., Nature* **347**, 294 (1990).
10. L. Samuelson, K. Wiebauer, C. M. Snow, M. H. Meisler, *Mol. Cell. Biol.* **10**, 2513 (1990).
11. C.-A. Reynaud, V. Anquez, H. Grimal, J.-C. Weill, *Cell* **48**, 379 (1987).
12. C. Roth, S. Longacre, A. Raibaud, T. Baltz, H. Eisen, *EMBO J.* **5**, 1065 (1986).
13. G. R. Fink, *Cell* **49**, 5 (1988).
14. H. M. Temin, *Nature* **339**, 254 (1989).
15. J. B. Watson and J. G. Sutcliffe, *Mol. Cell. Biol.* **7**, 3324 (1987); J. Martignetti, H. Tiedge, J. Brosius, unpublished data..
16. D. L. Nelson *et al., Proc. Natl. Acad. Sci. U.S.A.* **86**, 6686 (1989).
17. Exaptations are features that now enhance fitness but were not built by natural selection for their current roles, while the term adaptation is restricted to features built by selection for their current roles. [S. J. Gould and E. Vrba, *Paleobiology* **8**, 4 (1982)].