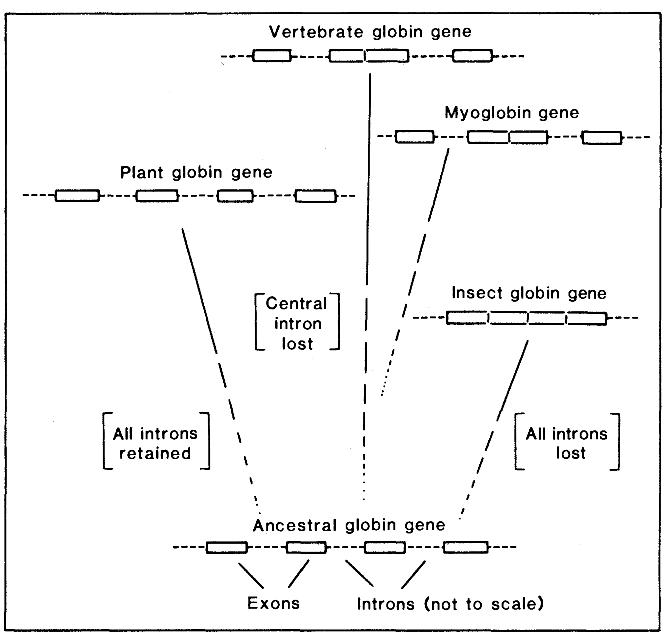
Surprise Finding with Insect Globin Genes

Marianne Antoine and Jürgen Niessing, of the University of Marburg, Germany, have just published the structure of an insect globin gene (*1*), an event that has been long awaited by biologists interested in the slightly baroque relationships among the family of globin genes in the various branches of the living world. Unlike all other globin genes so far characterized, those of the midge *Chironomous thummi thummi*—it has 12 in all—are not interrupted by noncoding sequences, or introns. This result is more than a little unexpected, and it adds another thread of intrigue to this already interesting gene family.

Vertebrate globins, of which there are many variants, all follow the same structural pattern: they comprise three coding regions, exons, interrupted by two introns. In addition, even though the length of the introns may vary between different types, their location within the coding sequences is pretty much constant. In the α- and β-genes of human hemoglobin, for instance, the first introns are much the same size, whereas the second interruption in the β-gene is seven times longer than that in the α-gene. And when the myoglobin gene was added to the family (from seal) it too turned out to be split by two introns at the same positions as in globin, but the introns were immense (*2*).



The conservation of overall configuration of globin gene structure throughout the vertebrate world, and the extension of this to myoglobin, which diverged from globin some 700 million years ago, appeared to bespeak a great stability among the family. When, however, K. A. Marker and his colleagues at the University of Aarhus, Denmark, published the structure of leghemoglobin, which codes for a protein in the root nodules of leguminous plants, family harmony was shaken. Instead of two introns interrupting the coding sequences, the leghemoglobin gene clearly had three, thus giving four exons (*3*).

Close examination of the leghemoglobin gene components revealed that the only real difference between it and the globin gene was that its extra intron simply split the central exon of the globin pattern. Mitiko Gō, of Kyushu University, Japan, showed that this extra intron in the leghemoglobin gene separates two discrete structural modules of the central exon of globin, which made evolutionary sense of the presumably more primitive three intron/four exon pattern (*4*). Vertebrate globin genes have, presumably, been derived by the loss of this "extra" intron.

If the leghemoglobin pattern represents the primitive globin gene pattern, an obvious question is, where did the plant gene come from? Has it been retained in the plant kingdom from the ancestral eukaryote common to plants and animals? Very few plants other than legumes appear to have the gene, so on the face of it this explanation seems unlikely. In which case, could the leghemoglobin gene be a recent acquisition, perhaps introduced by "horizontal transfer" on a viral vector from another branch of the biological world, as Alec Jeffreys, of the University of Leicester, England, has speculated (*5*)? Jeffreys had in mind insects as a possible source of this primitive form of the globin gene, but now says that the demonstration of intronless midge genes "almost certainly rules out the idea." It now seems more likely that leghemoglobin as seen in modern legumes is the expression of a gene that has passed through the plant kingdom since the time of the eukaryotic common ancestor. In which case one would infer that the gene has been lost or is simply not expressed in other nonleguminous plants.

Ever since 1980 genes have been known to lose introns, usually with the production of inactive "pseudogenes," as Philip Leder and his colleagues demonstrated with a mouse α-globin (*6*). Messenger RNA, or perhaps a DNA copy of a messenger, might hybridize with the gene on the chromosome, which would "loop out" the noncoding regions, possibly including controlling elements, and cause them to be lost.

In addition to many intronless pseudogenes so far documented, there are cases where single introns have been excised from otherwise normal genes (for instance, the rat has two insulin genes, one of which has two introns, the other only one). Midge globin represents the first documentation of a gene that has apparently been completely ransacked of all its introns and yet has retained its controlling elements and therefore remained functional. In addition to the coding regions that specify the amino acid sequence of the protein, the midge globin gene has at its 5′ end a short exon that encodes a so-called signal sequence, which is a common feature of proteins that are secreted from cells, as happens with midge globin. Typically, the exon for the signal sequence is separated from the protein encoding sequences by a short exon. In the midge globin gene this intron too has been lost, which might be taken to indicate the total loss of all introns in a signal conversion event rather than separate stochastic events.—**Roger Lewin**

References

1. M. Antoine and J. Niessing, *Nature (London)* **310**, 795 (1984).
2. A. Blanchetot, V. Wilson, D. Wood, A. J. Jeffreys, *ibid.* **301**, 732 (1983).
3. E. O. Jensen *et al.*, *ibid.* **291**, 677 (1981).
4. R. Lewin, *Science* **217**, 921 (1982).
5. A. J. Jeffreys, in *Genetic Engineering*, R. Williamson, Ed. (Academic Press, New York, 1981), vol. 2, pp. 1–48.
6. Y. Nishioka, A. Leder, P. Leder, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2806 (1980).