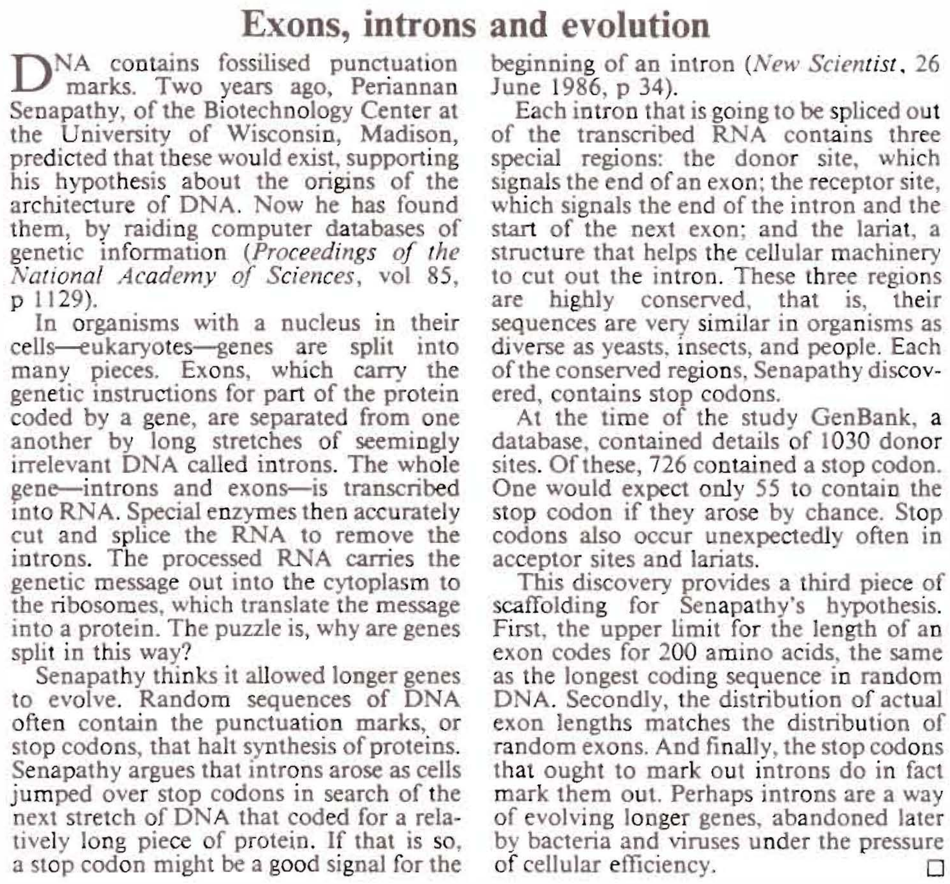
Commentary on the paper:

Possible evolution of splice-junction signals in eukaryotic genes from stop codons

By Periannan Senapathy

Proceedings of the National Academy of Sciences
(1988) Vol. 85: 1129-1133.

In
New Scientist
31 March 1988: page 31



Exons, introns and evolution

DNA contains fossilised punctuation marks. Two years ago, Periannan Senapathy, of the Biotechnology Center at the University of Wisconsin, Madison, predicted that these would exist, supporting his hypothesis about the origins of the architecture of DNA. Now he has found them, by raiding computer databases of genetic information (*Proceedings of the National Academy of Sciences*, vol 85, p 1129).

In organisms with a nucleus in their cells—eukaryotes—genes are split into many pieces. Exons, which carry the genetic instructions for part of the protein coded by a gene, are separated from one another by long stretches of seemingly irrelevant DNA called introns. The whole gene—introns and exons—is transcribed into RNA. Special enzymes then accurately cut and splice the RNA to remove the introns. The processed RNA carries the genetic message out into the cytoplasm to the ribosomes, which translate the message into a protein. The puzzle is, why are genes split in this way?

Senapathy thinks it allowed longer genes to evolve. Random sequences of DNA often contain the punctuation marks, or stop codons, that halt synthesis of proteins. Senapathy argues that introns arose as cells jumped over stop codons in search of the next stretch of DNA that coded for a relatively long piece of protein. If that is so, a stop codon might be a good signal for the beginning of an intron (*New Scientist*, 26 June 1986, p 34).

Each intron that is going to be spliced out of the transcribed RNA contains three special regions: the donor site, which signals the end of an exon; the receptor site, which signals the end of the intron and the start of the next exon; and the lariat, a structure that helps the cellular machinery to cut out the intron. These three regions are highly conserved, that is, their sequences are very similar in organisms as diverse as yeasts, insects, and people. Each of the conserved regions, Senapathy discovered, contains stop codons.

At the time of the study GenBank, a database, contained details of 1030 donor sites. Of these, 726 contained a stop codon. One would expect only 55 to contain the stop codon if they arose by chance. Stop codons also occur unexpectedly often in acceptor sites and lariats.

This discovery provides a third piece of scaffolding for Senapathy's hypothesis. First, the upper limit for the length of an exon codes for 200 amino acids, the same as the longest coding sequence in random DNA. Secondly, the distribution of actual exon lengths matches the distribution of random exons. And finally, the stop codons that ought to mark out introns do in fact mark them out. Perhaps introns are a way of evolving longer genes, abandoned later by bacteria and viruses under the pressure of cellular efficiency. ☐