1973

Reprinted from: MOLECULAR CYTOGENETICS
Edited by Barbara Ann Hamkalo
and John Papaconstantinou
Book available from: Plenum Publishing Corporation
227 West 17th Street, New York, N.Y. 10011

7. A NEW APPROACH TO THE STUDY OF NUCLEOTIDE SEQUENCES IN DNA'S

Giorgio Bernardi, Stanislav D. Ehrlich and Jean-Paul Thiery

Laboratoire de Génétique Moléculaire, Institut de Biologie Moléculaire, Paris 5e, France

We wish to outline here a new procedure for studying nucleotide sequences in DNA's. The procedure is based on our recent demonstration that at least the four deoxyribonucleases (DNases) we have investigated so far (hog spleen acid DNase, snail hepatopancreas acid DNase, bovine pancreas DNase and E. *coli* endonuclease I) hydrolyze specific sets of short nucleotide sequences.

Using methods described elsewhere (Bernardi *et al.*, 1973) for the isolation and analysis of the termini, *i.e.* the nucleotides near the breaks introduced by the enzymes, X↓YZ (the sequence being written in the usual 5'→3' direction and the vertical arrow indicating the position of the break), it is possible to show (Table I) that the base composition of termini a) differs from the values expected for random degradation, in which case the composition of each terminus considered should be equal to the average base composition of the DNA; the 5' penultimate nucleotide is, however, not recognized by the snail enzyme, as shown by the fact that its composition is equal to that expected for the nearest neighbors of the 5' terminal nucleotide; b) differs according to the enzyme used indicating that different sets of sequences are split by different enzymes; c) does not vary, as a rule, according to the level of DNA degradation. The minimum length of the sequences recognized by the nucleases is 4 nucleotides for the spleen enzyme (in which case the 3' penultimate nucleotides were also analyzed), 3 nucleotides for the pancreatic DNase and E. *coli* endonuclease I, and 2 nucleotides for the snail enzyme.

Since these nucleases split specific sets of sequences, the analysis of termini provides information on the frequency of these sequences in a given DNA. In fact, the composition of termini is

Table I. Termini liberated from calf thymus DNA by four DNases

Enzyme		3' term.	5' term.	5' pen.
	T	20	11	14
Spleen	G	43	43	26
DNase*	A	29	18	52
	C	8	28	8
	T	16	14	38
Snail	G	6	45	24
DNase	A	78	10	21
	C	1	31	14
	T	36	38	13
Pan-	G	15	22	36
creatic	A	31	15	30
DNase	C	18	25	21
	T	41	24	28
E. coli	G	8	35	29
endonu-	A	35	17	29
clease I	C	16	23	14

* In this case, the average chain length of oligonucleotides was equal to 15, and 3'-penultimate nucleotides were also analyzed; they were T 22%, G 16%, A 46%, C 16%.

related a) to the average composition of the sequences that can be split by the enzymes; b) to the k_M and V_{max} values associated with each sequence; and c) to their relative amounts in the DNA under consideration.

The latter point is shown by the fact that the composition of termini as obtained from DNA's having different (G+C) contents are different (Fig. 1). If the compositions of termini released from bacterial DNA's are plotted against their (G+C) contents, linear relationships are obtained (Fig. 1). The choice of bacterial DNA's in order to establish such relationships is justified a) by the fact that bacterial DNA's do not contain repetitive sequences and b) by the fact that the doublet frequencies of bacterial DNA's, as determined by the nearest neighbor analysis, show essentially linear relationships with the frequencies predicted for random association, indicating a common type of doublet distribution in these DNA's.

As expected, the composition of termini released from DNA's containing "repetitive" nucleotide sequences deviates, in either direction, from the linear relationship obtained with non-repetitive (bacterial) DNA's. The deviation patterns thus obtained represent

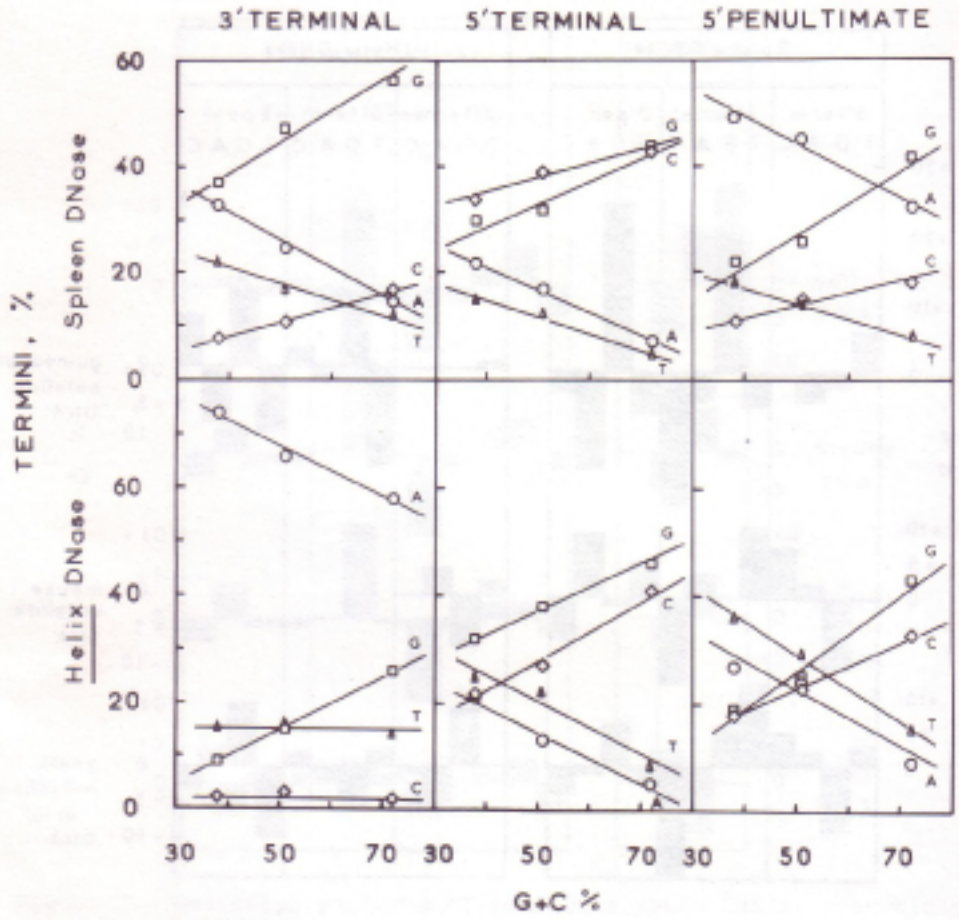

Figure 1. The percentage of A (open circles), G (open squares), C (open diamonds) and T (open triangles) in the 3'-terminal, 5'-terminal and 5'-penultimate nucleotides formed by spleen and the snail DNase from bacterial DNA's, is plotted against the (G+C) contents of DNA's.

a novel way of characterizing "repetitive" DNA's or, more generally, DNA's having sequence distributions different from those of the bacterial DNA's examined here. Fig. 2 shows the deviation patterns of three DNA's containing short repeated sequences: the satellite DNA's from mouse and guinea pig and the mitochondrial DNA from yeast Expectedly, deviation patterns obtained with different enzymes on the same DNA's are different from each other, as are deviation patterns obtained with the same enzyme on different DNA's.

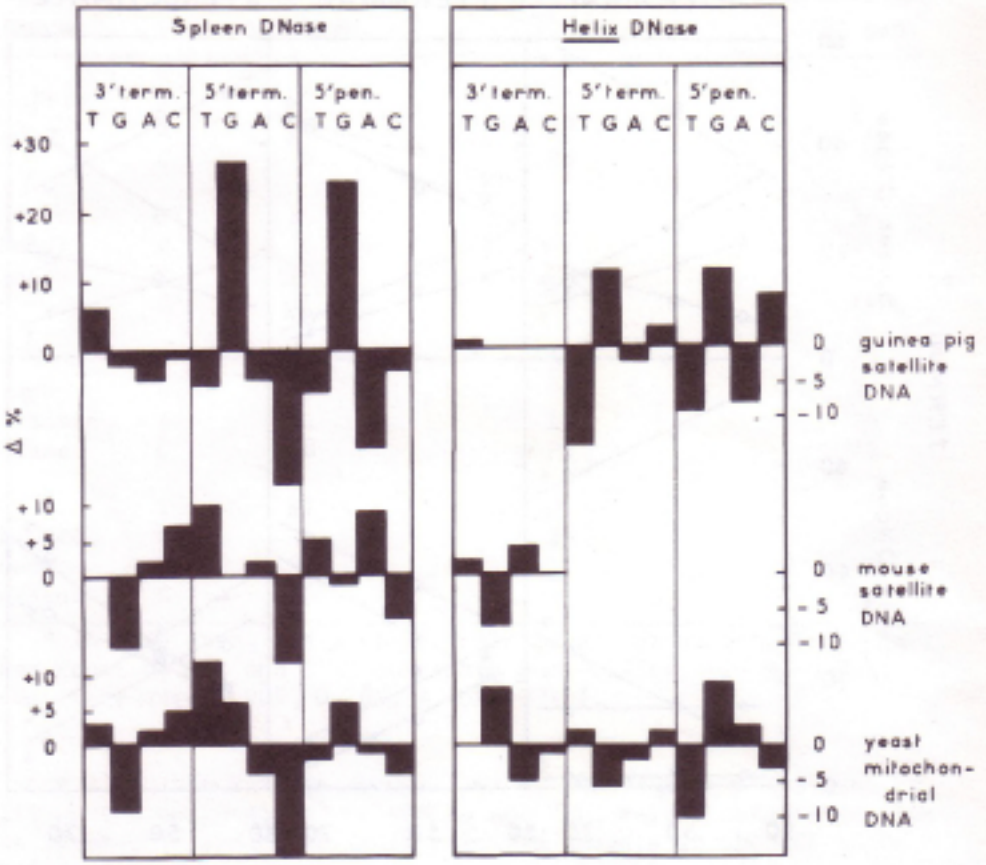

Figure 2. Deviation patterns of three repetitive DNA's. The histograms show the differences between the composition of termini formed from guinea pig satellite, mouse satellite and yeast mitochondrial DNA's by spleen and snail DNases and the compositions expected for bacterial DNA's having the same (G+C) contents. The delta values represent differences in the percentages of each terminus.

As another example, Fig. 3 shows deviation patterns obtained with calf, mouse and guinea pig DNA's and with yeast nuclear DNA. It is evident that in this case, too, a number of sequences are present at greater or lower frequencies in the eukaryotic DNA's compared to bacterial DNA's of identical (G+C) composition. Interestingly, the mammalian DNA's have similar deviation patterns whereas the yeast nuclear DNA pattern is quite different; mammalian DNA's appear to share the sequence features that are responsible for the similarity of their deviation patterns and that do not exist in

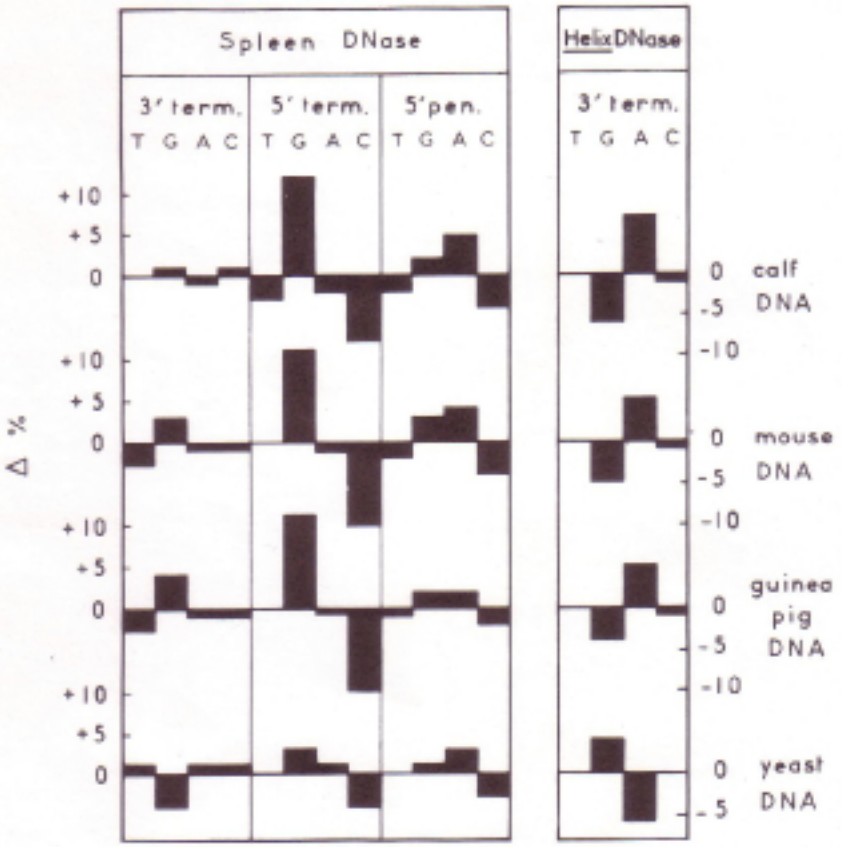

Figure 3. Deviation patterns of four eukaryotic DNA's. The histograms show the differences between the composition of termini formed from eukaryotic DNA's by spleen and snail DNases and the compositions expected for bacterial DNA's having the same (G+C) contents. Delta values represent differences in the percentages of each terminus.

yeast nuclear DNA. The possibility that the deviations observed in the mammalian DNA's arise from their satellite DNA's is ruled out by the completely different deviation patterns exhibited by the latter (Fig. 2).

REFERENCE

Bernardi, G., S. D. Ehrlich and J. P. Thiery. 1973. A new approach to the study of nucleotide sequences in DNA: the analysis of termini formed by DNases. Methods in Enzymol. in press.