# Chaos Game Representation (CGR)

## a HyperCard application for presenting nucleotide sequence data

Recently, M. Joel Jeffrey (Jeffrey 1990) published a way of presenting nucleotide sequences in graphical form, that he called Chaos Game Representation (CGR). It is, to my knowladge, the first published attempt to apply the nonlinear approach to nucleotide sequences data (excluding Ohno 1988). While the complete mathematical description of this approach waits its completion, I felt its present state to be important enough to make it widely available and let as many molecular biologists as possible to try it out.

Shortly, CGR plots nucleotides in a square where the corners are labelled A, C, G and T. Beginning at the centre, successive nucleotides in a sequence are plotted halfway between the corner carrying the corresponding label and the previous point. Estimates of relative oligonucleotide (<10 b) frequencies can be seen in a glance from the fractal plot. The possible implications of this approach are discussed in detail in the original article.

CGR HyperCard stack reads sequence data in EMBL or GenBank form or in plain sequences from text only files. Introns or other sequence regions can be excluded from the plot by entering their base range. Contrary to what is suggested in the original article, both plotting and calculation are interrupted these regions. For easy modification, the file and excluded region information are transferred to the new card when “New card” button is pressed.

Also included in the stack is my own modification of the CGR approach for comparing nucleotide frequencies in codons. Three plots, one for each base in codons are plotted separately. For most sequences, the resulting plots show clearly the increasing randomness from first to third codon base and can be used, for example, to determine the actual reading frame from overlapping ORFS. This approach loses much of the accuracy of a mathematical approach (e.g. Tavaré and Song 1989), but gains in presenting the data in easily understood graphical form. For molecular biologists, this might be more useful.

The users are strongly encouraged to try their own ideas and modify the scripts of CGR. I am interested in any developments in CGR approach to gene structure.

CGR 1.0 is available from the EMBL Network File Server (Internet address: NETSERV@EMBL.BITNET).

**References:**

Jeffrey JM 1990: Chaos game representation of gene structure. Nucl Acid Res 18:2163-21270.

Ohno S 1988. Codon preference is but an illusion created by the construction principle of coding sequences. Proc Natl Acad Sci USA 85: 4378-4382.

Tavaré S and Song B 1989: Codon preference and primary sequence structure in protein-coding regions. Bulletin of Mathematical Biology 51:95-115

8 August 1990

Heikki Lehväslaiho
Cancer Biology Laboratory, Departments of Pathology and Virology
University of Helsinki, Haartmaninkatu 3, SF-00290 Helsinki, FINLAND
E-mail: LEHVASLAIHO@CC.HELSINKI.FI