

# 1/f Noises in DNA Base Sequence

Miyako Kawachi

Supervisor E.Takushi

## 1 Introduction

We wish to report a result of seeking for an existence of new functions hidden in the DNA base sequences coding the proteins for virus such as parainfluenza type-1 virus parasitizing in bacteria and human mitochondrial genomes [1]. The DNA base sequence is transformed into numerical sequence and is calculated for the power-spectrum by Fourier transformation through correlation functions [2]. At large  $f$  power spectrum is dominated by the white noise of uncorrelated random process. Within the white noise there is a sharp peak at  $f = 1/3$  that is probably related to the nucleotide triplet that specifies one of 20 amino acids. At low  $f$ , power-spectrum shows power-law scaling  $1/f^\alpha$  noise-like. The DNA base sequences and amino acid sequences have characteristics of the repeated structure. In the previous study, it have been calculated the power-spectra for Coliphage  $\phi k$  Bacteriophage  $\phi x174$ , human  $p53$  and  $p51$  tumor suppressor gene (TSG) and human mitochondrial DNA sequences [3, 4]. These results may give a possibility of new way of finding an appearance of function in coding and non-coding regions. In the present study, the DNA base sequence of parainfluenza type-1 virus has been analyzed (NCBI Accession Number NC\_03465). The parainfluenza type-1 virus genome has 15600 bases and 10 kinds of genes which have been defined by virtue of their protein products; 10 of these ( N, C', P, C, Y1, Y2, M, F, HN, and L ) are an essential for virus reproduction.

## 2 Results and Discussion

The power-spectrum is calculated by Fourier transformation software "GeneFFT". For the calculation of exponent  $\alpha$  of  $1/f^\alpha$  noise, making graph software "gnuplot" is used.

In summary, there exists a long-range correlation in parainfluenza type-1 virus RNA base sequence. It is observed that the value of exponent  $\alpha$  of long-range correlation  $1/f^\alpha$  noise is related to the types of gene and the location of DNA base sequence. Finally, it is discussed

that four modes of cooperative group genes such as groups  $[\Sigma, L, P], [P, HN, F], [P, N, M, C', C, Y1],$  and  $[P, Y2]$  are seemed to be correlated with each other through the gene P (Fig.2).

[1]S. Anderson et al and F. Sanger, Sequence and Organization of the Human Mitochondrial Genomes, Nature 290,457 - 465,1981. [2]Li,W.and Kaneko, K.,DNA correlations,Nature,360 : 635-636,1992; Voss R. F, Evolution of long-range correlations and  $1/f$  noise in DNA base sequences, Phys. Rev. Lett., 68 : 3805 - 3808, 1992; Stanley , H.E. et al, Long-range correlations in nucleotide sequences, Nature, 356 : 168 - 170, 1992. [3]Takushi, E., Power spectra of DNA sequences in phage and tumor suppressor genes (TSG), Genome Informatics, 13, 412-413, 2002. [4]Takushi, E., and Chinen, Y., Power Spectrum of DNA Base Sequence in Human Mitochondrial Genome, Proc. Int. Conf. Biological Physics, 780, Goeteborg, Sweden, 2004.

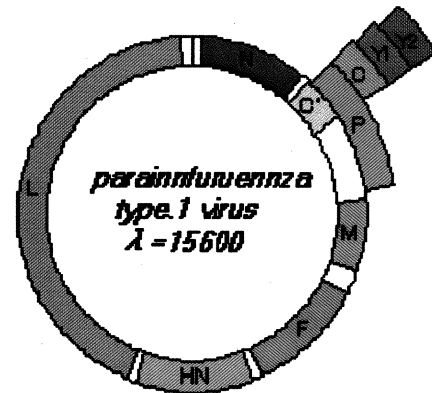


Fig1: The gene map of parainfluenza.

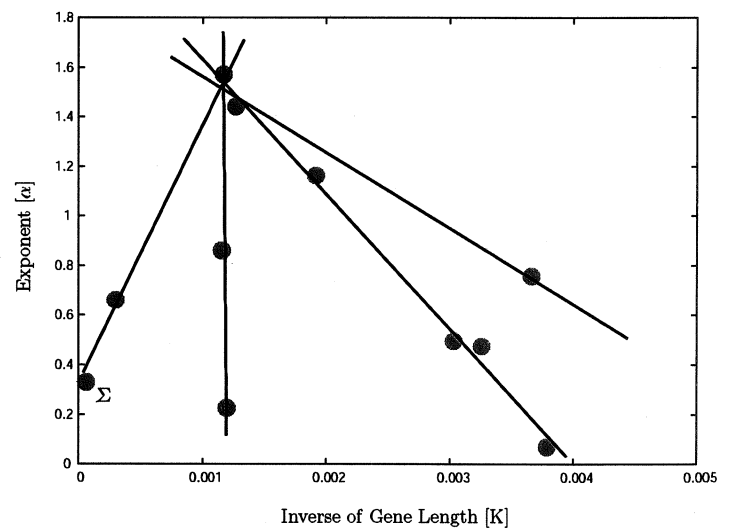


Fig2: Size effect and cooperative functions for parainfluenzal genes.