2 Construction

We have constructed codon usage database (CUTG: Codon Usage Tabulated from GenBank), which can be accessed on the WWW *http://tisun4a.lab.nig.ac.jp/codon/CUTG.html* [6]. For all full–length ORFs of *E. coli* and *S. cerevisiae*, Fop values are compiled using CUTG. Distributions of Fop values along the chromosome are presented as the plotted data shown by image files. A program allows to calculate a Fop value for a sequence of *E. coli* and *S. cerevisiae*.

3 Access

The data can be accessed on the following WWW server:

http://tisun4a.lab.nig.ac.jp/fop/fop.html

Comments on the database can be sent to *yanakamu@ddbj.nig.ac.jp* by E-mail.

Acknowledgement

The authors wish to thank Dr. Y. Ugawa at the DNA Information and Stock Center, National Institute of Agrobiological Resources, for help in compiling the data.

References

[1] T. Ikemura, "Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes" *J. Mol. Biol.*, Vol. 146, pp. 1–21, 1981.

[2] T. Ikemura, "Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes." *J. Mol. Biol.*, Vol. 158, pp. 573–597, 1982.

[3] Pedersen S., Bloch, P.L., Reeh, S. and Neidhardt, F.C., "Patterns of protein synthesis in *E. coli*: A catalog of the amount of 140 individual proteins at different growth rates." *Cell.*, Vol. 14, pp. 391–402, 1978.

[4] T. Ikemura, "Correlation between codon usage and tRNA content in microorganisms." In Transfer RNA in Protein Synthesis (CRC Press), pp. 87–111, 1992.

[5] Y. Nakamura, T. Fukagawa, K. Sugaya and T. Ikemura, "Estimation of protein–production levels for ORFs found by *E. coli* and yeast genome projects, basing on levels of "optimal codon" usage, in connection with feasibility of their protein coding ability and with assignment to foreign–type genes." *Proc. Genome Informatics Workshop 1994*, pp. 188–189, 1994.

[6] Y. Nakamura, K. Wada, Y. Wada, H. Doi, S. Kanaya, T. Gojobori and T. Ikemura, "Codon usage tabulated from the international DNA sequence databases" *Nucl. Acids Res.* in press.

Fop (frequency of optimal codon usage): WWW service with its distribution analysis

Y. Nakamura
yanakamu@ddbj.nig.ac.jp

T. Ikemura
tikemura@ddbj.nig.ac.jp

National Institute of Genetics and The Graduate University for Advanced Studies, Yata 1111, Mishima, Shizuoka 411 Japan

Abstract

We have been constructing a WWW server which shows: 1) Fop value for all of the full–length ORFs of E. coli and S. cerevisiae in GenBank DNA sequence database and 2) distribution of Fop value along the chromosomes. A program available on the server allows to calculate a Fop value for a sequence of these species.

1 Background

As an index to show the optimization level of synonymous codon choice in each gene to translation process, we introduced frequency of usage of "optimal codon" (Fop) mainly on the basis of experimental data quantifying tRNA contents of *E. coli* and *S. cerevisiae* [1][2].

$$\mathrm{Fop} = \frac{\text{the numbers of "optimal" codons}}{\text{sum of the numbers of "optimal" and "nonoptimal" codons}}$$

Optimal codons have been defined as the codons which are recognized by the most abundant isoaccepting tRNA. The cellular content of a wide variety of *E. coli* proteins has been measured [3]. The correlation between Fop and protein production level was revealed, showing Fop in each gene to be related to its protein content [4]. We have calculated Fop values of ORFs proposed for the DNA sequences registered by genome project of *E. coli* and *S. cerevisiae*, and have estimated protein–production levels and feasibility of their protein coding ability [5].

中村保一, 池村淑道: 国立遺伝学研究所進化遺伝研究部門, 総合研究大学院大学生命科学研究科, 〒 411 三島市谷田 1111