On Correlation between G+C Contents and Intron Lengths: Longer Introns Tend to be More A+T Rich

Yasuhiro Asakawa [1] s93011ya@sfc.keio.ac.jp

Masaru Tomita [2] mt@sfc.keio.ac.jp

Laboratory for Bioinformatics

[1] Department of Policy Management

[2] Department of Environment Information

Keio University
5322 Endo, Fujisawa, 252 Japan

Abstract

An interesting correlation between G+C contents and the lengths of primate introns have been found by our computer analysis.

All sequences of primate introns were extracted from the Genbank database and classified into subgroups according to their lengths (the number of bases; increment of 100). G+C contents (%) were then calculated for each subgroup.

The results indicate that shorter introns tend to contain more G and C nucleotides, and longer introns contain A and T nucleotides.

Frequencies of each nucleotide for each subgroup are shown in figure 1.

We also computed G+C contents of exons flanking those introns for each subgroup. As we can see in figure 2, the similar but weaker tendencies are observed.

Biological significance of those observations is currently under investigation. We also intend to extend our analysis to other eukaryotes.

Acknowledgment

This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas 'Genome Science' from The Ministry of Education,Science, Sports and Culture in Japan.

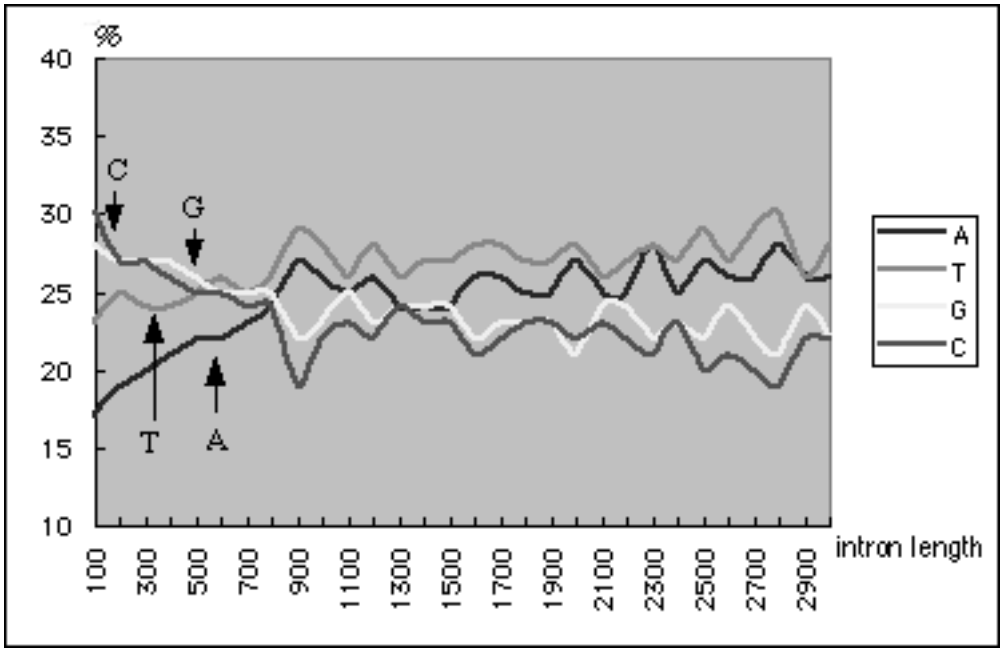

Figure 1: Nucleotide contents and lengths of primate introns

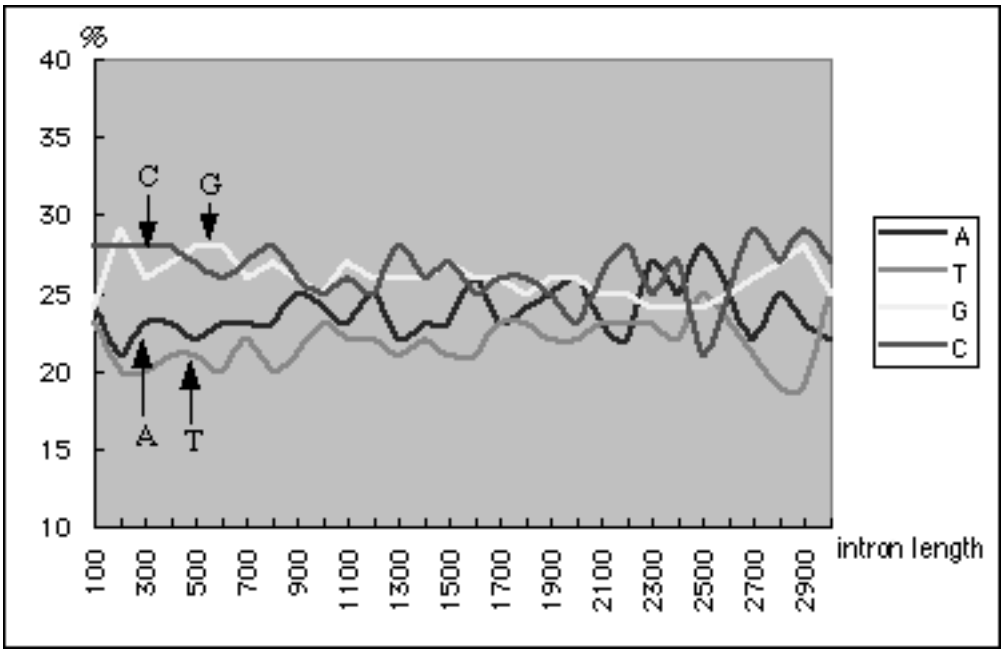

Figure 2: Nucleotide contents of flanking exons(100 bases) and lengths of primate introns