2 Results and Discussion

Orthologous genes were selected from HOVERGEN (Homologous Vertebrate Genes Database, release 13) [1]. We obtained 456 introns from 106 human/rodentia genes, 93 introns from 35 mouse/rat genes, 76 introns from 30 human/artiodactyla genes, and 26 introns from 7 human/old world monkey genes. The last data set (human/monkey) was used for the estimate of sequence diversity.

Natural mutageneses, mainly insertions and deletions, change the size of the genes. The accumulation of insertions and deletions expected to be faster in introns than in coding sequences. For human/rodentia genes, the difference of intron length varies from 0 up to 2 kbp. Compared to the conservative length of coding sequences, length constraints on intron sequences seem to be very low.

The distribution profile of intron size also exhibits a tendency of longer human introns from human/rodentia comparison: in all 106 comparisons 68 % of human introns are longer than rodentia introns. Such a significant shift of intron size is not observed from other interspecies comparisons (human/artiodactyla, mouse/rat). Therefore, the shift may be a reflection of events that occurred differently between the evolution of rodentia and other lineages (primate/artiodactyla). It has been proposed that deletions occur more frequently than do insertions [2], and they are related to meiosis. The short life cycle and the accumulation of deletions may be a major source of the smaller size of rodentia introns.

Base compositions of different regions of genes are compared between species. When homologous genes are compared, compositional differences are larger in the 3rd codon positions and introns than in the 1st and the 2nd codon positions, although the difference in introns are smaller than that of the 3rd codon position. This is consistent with the substitution rate of each region given from the human/monkey comparison.

Futher study on homologous intron will lead us to a better knowledge of the evolutionary history and the role of intron.

Acknowledgement

This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas, "Genome Informatics", from The Ministry of Education, Science, Sports and Culture in Japan. The computation time was provided by the Supercomputer Laboratory, Institute for Chemical Research, Kyoto University.

References

[1] L. Duret, D. Mouchiroud, M. Gouy "HOVERGEN: a database of homologous vertebrate genes," *Nucleic Acids Res.*, Vol. 22, pp. 2360-2365, 1994.

[2] X. Gu, W.-H. Li "The size distribution of insertions and deletions in human and rodent pseudogenes suggests the logarithmic gap penalty for sequence alignment," *J. Mol. Evol.*, Vol. 40, pp. 464-473, 1995.

Statistical features identified from comparison of homologous introns

H. Ogata W. Fujibuchi M. Kanehisa
{ogata, wataru, kanehisa}@kuicr.kyoto-u.ac.jp

Institute for Chemical Research, Kyoto University,
Gokasho, Uji, Kyoto 611 Japan

Abstract

In order to develop an understanding of the evolutionary history of introns, we compared intron sequences from homologous genes of different vertebrate species. We observed the following statistical features of introns. From interspecies comparisons, the length difference of an intron from each gene exhibits considerable variations, which is in consistent with the previous observation on the wide distribution of insertion and deletion sizes. The distribution profiles also indicate that 68 % of introns are longer in human than in rodentia, although such a significant shift in the intron length is not observed in mouse/rat and human/artiodactyla comparisons. From the analysis of diversities in base compositions and sequence data between intron and coding regions, a similar behavior was observed between intron and the third codon position, although intron sequences appeared to be more constrained. Those descriptive analysis will help us to understand the functional constraints and the recent history of introns.

1 Introduction

One of the major features in gene structures that differentiate eukaryotes from prokaryotes is intron/exon organizetion in eukaryotes. From the discovery of intron, there has been an extensive body of research to characterize and understand the function and the evolution of introns. The research focused on size distributions, base compositions, evolutionary rates, and so on. Although a lot of interesting features of introns were clarified, the analyses of interspecies comparisons have been limited; some of them were concerned with one or several genes and some of them were restricted to one species. Among those which compared introns between species, there is little work to describe detailed differences between species. Recent accumulation of sequence data has enabled us to inspect diversities of many introns by orthologous comparison. We report here length distributions, base compositions, and sequence diversity of introns. Our analysis focuses on the differences of homologous introns by pairwise comparisons.

緒方博之, 藤渕航, 金久實：京都大学化学研究所生体分子情報部門，〒 611 京都府宇治市五ヶ庄