Comparison between Linear Chromosomes in Terms of Duplicate Genes: Implication of Whole Genome Duplication in Yeast

Satoshi Fukuchi
sfukuchi@lab.nig.ac.jp
1 Laboratory for Gene-Product Informatics, The Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Yata 1111, Mishima 411-8540, Shizuoka, Japan

Ken Nishikawa
knishika@genes.nig.ac.jp
2 Laboratory of Mathematics, Saga Medical School, Nabeshima 5-1-1, Saga 849-8501, Japan

Katsuhisa Horimoto
horimoto@post.saga-med.ac.jp

Keywords: genome duplication, linear chromosome, quantification, yeast

1 Introduction
Recently, the complete genome structures from eukaryotes have been determined in a steady pace, making the importance of the genome comparisons greater. Here, we present an approach to compare the structures of linear chromosomes in terms of related genes. The performance of the present approach is illustrated by comparisons between the yeast chromosomes in terms of the duplicate genes. And, the result of our approach is discussed in connection with the possibility of the genome duplication during the yeast genome evolution.

2 Method and Results
2.1 Duplicate Genes in Yeast
The amino acid sequences of the 6215 gene products in the yeast genome [2] were compared against all by BLASTP [1]. The amino acid sequence pairs with the BLAST score 200 were listed as the duplicate genes. The blast score for the query sequence itself multiplied by 0.8 was employed as the threshold in the case that the score for query itself is less than 200. In order to define the family of duplicate genes, the single-linkage clustering was performed on the listed genes above, resulting in 576 gene families consisting of 1,571 genes.

2.2 Quantification Method II
To compare the chromosomes in terms of the locations and the families of the duplicated genes, we adopted the Quantification Method II (QM2) [3], which is a multivariate analysis for the discrimination of categorical data. The chromosome, the gene family, and the gene location are categorized in the application of QM2 to the linear chromosome. The QM2 discriminate between the duplicate genes according to the chromosomes in which they reside. In the quantification of the duplicated genes, therefore, the genes that reside in the chromosomes with a similar structure share similar discriminant scores.

2.3 Chromosome Distance
A chromosome distance is defined as follows. First, we calculate the average discriminate score of i th chromosome, \( A\nu_i \), i.e.,

\[
A\nu_i = \frac{1}{N_i} \sum_{j=1}^{N_i} Y_{ij},
\]
where $Y_{ij}$ is the discriminant scores for $j$ th gene on $i$ th chromosome, and $N_i$ is the number of the duplicate genes on $i$ th chromosome. Then, the Euclidean distance between the average scores for each chromosome is calculated as follows,

$$D_{ij} = \sqrt{(A_{ui} - A_{uj})^2}$$

(2)

Consequently, the distance quantitatively expresses the difference between the chromosomes in macroscopic view from the locations and the family of duplicated genes.

2.4 Dendrogram between Chromosomes

The distance matrix for the 16 yeast chromosomes was created by using equation (2). The neighbor-joining tree [4] was shown in Fig. 1 based on this matrix. Taking 50% bootstrap values (denoted by each node in Fig. 1) as a cutoff for the clustering, the 16 yeast chromosomes were clustered into 7 clusters: {I, VII and VIII}, {II and XVI}, {III and XIV}, {IV and XV}, {V and IX}, {VI, XII, and XIII}, and {X and XI}.

Figure 1: Dendrogram of the 16 yeast chromosomes.

3 Discussions

Given that the clusters reproduced an ancient form of the yeast chromosomes, there would be 7 chromosomes in the ancestral yeast. Thus, the present comparison supports the occurrence of single genome duplication in the yeast genome evolution [5]. Further evidence for the 7 clusters of chromosomes was obtained by the sequence comparisons between the duplicate genes. Notably, the present approach is easily applicable to compare the whole chromosome structures in different organisms. The application of the present approach to the global comparisons of chromosome structures will be present in near feature.

References