

Seeking Genome Rearrangements in Prokaryotic Genomes via Tandem Repeats

Satoshi Mizuta^{1,2}, Hikaru Munakata¹, Abulimiti Aimaiti¹,
Ichitaro Oya¹, Kenji Oosawa³, Toshio Shimizu^{1,4}

Keywords: tandem repeats, color-coding, genome rearrangement, whole-genome duplication

1 Introduction

Although tandem repeats (TRs) were caused by small scale local duplications on genomes, they can be used as probes to seek genome rearrangements such as a large scale segmental duplication or whole-genome duplication (WGD). WGD is thought to have played an important role in genome evolution[4]. Actually, the indications of WGD have been reported in several eukaryotic genomes ranging from yeast[1, 5] to human[2]. In prokaryotic genomes, however, no explicit evidence of WGD has been observed. Because genes were strongly affected by natural selection, duplicated genes would have vanished even though they were slightly changed if the changes were disadvantageous for surviving. On the contrary, TRs located on intergenic regions are rather freer from the natural selection. Even considerably changed TRs, therefore, would be conserved to some extent, which is thought to be one of the advantages of studying genomes by analyzing TRs.

In this work, we analyzed 13,542 TRs detected by the color-coding (CC) method[3, 6] in 44 prokaryotic genomes in detail and investigated genome rearrangements by examining locations of equivalent TR-pairs.

2 Materials and Methods

Genome sequences of the 44 prokaryotic species analyzed in this work were downloaded from GenBank, and TRs were searched by the CC method following to the procedure given originally by Yoshida et al.[6] with the aid of a Java applet CC viewer, 2D-PaCC (2-Dimensional Pattern Formation with Color Coding), developed by some of the authors [unpublished].

All the detected TRs were compared pairwise by SSEARCH (the University of Virginia [<http://fasta.bioch.virginia.edu/>]) with the default parameter values, and TR-pairs were extracted with a condition of E -value < 0.001 . When the overlap of the query TR with the target TR was larger than 50% and the TR-pair was extracted in two-way directions, it was identified as an equivalent TR-pair. The equivalent TR-pairs were located on a circle which symbolically expressed each genome.

If a genome had experienced a one-round WGD, we will find equivalent TR-pairs having central angles of around 180° on the circle; if a two-round WGD, we will find those of around 90° and 270° as well as 180° , and so on. We classified the genomes into four categories with respect to the distribution of the equivalent TR-pairs: $2R$ -like(C_1) in which all the central

¹Department of Electronic and Information System Engineering, Faculty of Science and Technology, Hirosaki University, Hirosaki 036-8561, Japan.

²E-mail: slmizu@cc.hirosaki-u.ac.jp

³Department of Nano-Material Systems, Graduate School of Engineering, Gunma University, Kiryu 376-8515, Japan. E-mail: kenji@nms.gunma-u.ac.jp

⁴E-mail: slsimi@si.hirosaki-u.ac.jp

Table 1: Genomes in which the indication of WGD were found

Species	Species
Proteobacteria	Other Bacteria
<i>Escherichia coli</i> K-12 MG1655	<i>Synechocystis</i> sp. PCC6803
<i>Neisseria meningitidis</i> MC58	<i>Thermosynechococcus elongatus</i> BP-1
<i>Yersinia pestis</i> CO92	Archaea
Firmicutes	<i>Methanococcus jannaschii</i> DSM2661
<i>Bacillus halodurans</i> C-125	<i>Methanopyrus kandleri</i> AV19
<i>Mycoplasma pneumoniae</i> M129	<i>Pyrobaculum aerophilum</i> IM2
<i>Staphylococcus aureus</i> N315	<i>Pyrococcus horikoshii</i> OT3
<i>Streptococcus pneumoniae</i> R6	

angles of around 90° , 180° , and 270° were explicitly observed to some extent, *Ambiguous*(C_2) in which the feature of *2R-like* was ambiguously observed, *Almost nothing*(C_3) in which only the central angles of less than 60° were observed, and *Nothing*(C_4) in which no equivalent TR-pair was found.

3 Results

The numbers of genomes classified into C_1 , C_2 , C_3 , and C_4 are 13, 16, 8, and 7, respectively. Table 1 lists the genomes classified into C_1 (the full list of the 44 genomes analyzed in this work is available elsewhere). Figure 1 shows the central angle distribution for *Escherichia coli*, which was classified into C_1 . TR-pairs of around 90° , 180° , and 270° were all observed to a large extent there. And besides, it has a characteristic pattern of distribution which is explicitly distinguished from the distribution shown by randomly located TR-pairs.

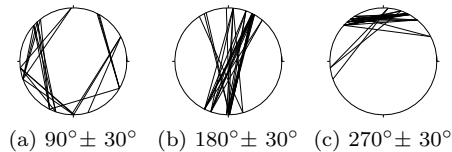


Figure 1: Central angle distribution of equivalent TR-pairs for *Escherichia coli*

The distributions for other genomes classified into C_1 are similar to that for *Escherichia coli*, although they are different in degree. These results are expected to show the indication of the occurrence of WGD in these genomes.

References

- [1] Kellis, M., Birren, B. W., and Lander, E. S. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**:617–624.
- [2] McLysaght, A., Hokamp, K., and Wolfe, K. H. 2002. Extensive genomic duplication during early chordate evolution. *Nature Genetics* **31**:200–204.
- [3] Mizuta, S., Munakata, H., Aimaiti, A., Oosawa, K., and Shimizu, T. 2004. Evaluation of the color-coding method for searching tandem repeats in prokaryotic genomes. *Chem-Bio. Inf. Journ.* **4**:133–141.
- [4] Ohno, S. 1970. *Evolution by Gene Duplication*. Springer-Verlag, New York.
- [5] Wolfe, K. H., and Shields, D. C. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**:708–713.
- [6] Yoshida, T., Obata, N., and Oosawa, K. 2000. Color-coding reveals tandem repeats in the *Escherichia coli* genome. *J. Mol. Biol.* **298**:343–349.