

# Genome Visualization by Isometric Projection Focused on Replication Origins and Termini, and Genome Instability

**Hirokazu Nishio**<sup>1,2</sup>      **Ken-nosuke Wada**<sup>2</sup>      **Yoshiko Wada**<sup>2</sup>  
 Nishio\_hirokaz@plamid.apgeo.com      kwada@m3.kcn.ne.jp      ywada@m3.kcn.ne.jp  
**Atsushi Fukushima**<sup>3</sup>      **Toshimichi Ikemura**<sup>4</sup>      **Hirotsada Mori**<sup>5</sup>  
 a-fukusi@bs.aist-nara.ac.jp      tikemura@lab.nig.ac.jp      hmori@gtc.aist-nara.ac.jp  
**Taku Oshima**<sup>5</sup>      **Shigehiko Kanaya**<sup>6</sup>  
 taku@bs.aist-nara.ac.jp      skanaya@gtc.aist-nara.ac.jp

<sup>1</sup> Applied Mathematics and Physics Course, Faculty of Engineering, Kyoto Univ.

Sakyo-ku, Yoshidahonmachi 606-8501, Japan

<sup>2</sup> UNTROD, Inc., 4-4-123 Aoyama, Nara 630-8101, Japan

<sup>3</sup> Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0101, Japan

<sup>4</sup> Division of Evolutionary Genetics, Department of Population Genetics, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan

<sup>5</sup> Research and Education Center for Genetic Information, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0101, Japan

<sup>6</sup> Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0101, Japan

**Keywords:** Fourth-isometric projection, replication origin and terminus, visualization of genome

## 1 Introduction

Intragenomic heterogeneity of GC content is generally more evident in higher eukaryotes than in prokaryotes and unicellular eukaryotes. Genomes of warm-vertebrates have long-range GC% mosaic structures that are related to chromosome bands [1, 2]. GC content is considered a fundamental characteristic of each genome and of genomic regions. However, we believe that GC content is too simple a parameter to characterize the genome and specific genome regions. Each genomic sequence registered by the International DNA Databank corresponds to one of the two DNA strands in a genome. Therefore, the G and A contents are not necessarily equal to the C and T contents, respectively. This tendency is more significant with respect to local sequence characteristics, and therefore, a method for effectively visualizing distribution of the four nucleotides in a three-dimensional space would be useful. We developed new techniques to visualize DNA sequence based on the fourth dimensional isometric projection and used this method to predict replication origins and termini in bacterial genomes.

## 2 Method

To represent the nucleotides in DNA sequence numerically, a vector for the  $i$ th nucleotide was denoted by  $\mathbf{b}_i = (b_{iA}, b_{iT}, b_{iG}, b_{iC})$ . Here,  $b_{iA}=1$ ,  $b_{iT}=0$ ,  $b_{iG}=0$ , and  $b_{iC}=0$  if the  $i$ th nucleotide is A and so on. With vector  $\mathbf{b}_i$ , the cumulative coordinate is defined by

$$\mathbf{x}_i = \sum_{k=1}^i \mathbf{b}_k.$$

when we focused on the  $i$ th position  $\mathbf{x}_i$  relative to  $\mathbf{x}_{i-w}$  and  $\mathbf{x}_{i+w}$ , the angle becomes an important parameter for understanding the relative relations of  $\mathbf{x}_i$  to  $\mathbf{x}_{i-w}$  and  $\mathbf{x}_{i+w}$ . This is done by calculating

$$\cos\theta = (\mathbf{x}_{i-w} - \mathbf{x}_i) \cdot (\mathbf{x}_{i+w} - \mathbf{x}_i) / \{|\mathbf{x}_{i-w} - \mathbf{x}_i| \cdot |\mathbf{x}_{i+w} - \mathbf{x}_i|\}$$

here,  $w$  is the window size. This is referred to as the  $\cos\theta$ -representation. An isometric projection of a DNA sequence composed of four nucleotides into a three-dimensional space is possible for cumulative coordinate  $y_i$  where

$$\mathbf{y}_i = (y_{i1}, y_{i2}, y_{i3}) = \left( \sum_{k=0}^i t_{k1}, \sum_{k=0}^i t_{k2}, \sum_{k=0}^i t_{k3} \right)$$

by assigning vector  $\mathbf{t}_i = (\alpha, \alpha, \alpha)$  for A,  $\mathbf{t}_i = (-\alpha, -\alpha, \alpha)$  for T,  $\mathbf{t}_i = (-\alpha, \alpha, -\alpha)$  for G, and  $\mathbf{t}_i = (\alpha, -\alpha, -\alpha)$  for C. This projection is the fourth-dimensional isometric projection. The software that generates the  $\cos\theta$ -map and fourth-dimensional isometric projection which is called “4D Isometric Genome Analyzer”, was purchased from UNTROD, Inc. (Nara-shi, Nara-ken, Japan).

### 3 Results and Discussion

The  $\cos\theta$ -value from 1 to -1 is shown by the graduation from white to black. Positions in the white zone extending from small to large base sizes correspond to switch points in long-range base composition across the genome. Six switch points, N1 to N6 were obtained from the  $\cos\theta$ -map. This finding was confirmed by the configuration of the fourth-dimensional isometric projection. The switch points, N1, N2, and N5, correspond to islands of horizontally transferred DNA (IHTs). Bacterial genomes are known to

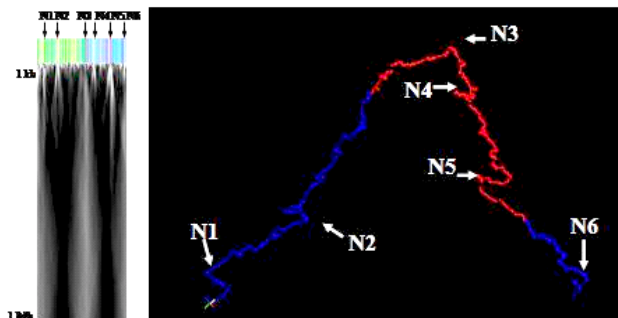


Figure 1:  $\cos\theta$  (left) and fourth isometric projections (right) for *N. meningitides*.

undergo frequent rearrangements, and regions of DNA derived from horizontal gene transfer often have atypical DNA composition relative to the entire genome. The isometric projection and  $\cos\theta$ -map can be used to visualize genome heterogeneity in base composition and to distinguish zones that differ significantly from the background base composition. One switch point (N3) that is continuously invariant from small to large window sizes in the  $\cos\theta$ -map is thought to correspond to a replication origin.

### References

- [1] Ikemura, T., Codon usage and tRNA content in unicellular and multicellular organisms, *J. Mol. Biol.*, 2:13–34, 1985.
- [2] Wada, K., Watanabe, I., Tsuchiya, R., and Ikemura, T., *G+C% mosaic structures of the higher vertebrate genome and distribution of dinucleotide frequencies. New Aspects of the Genetics of Molecular Evolution*, Kimura, M. and Takahata, N. (eds.), Japan Sci. Soc. Press, Tokyo/Springer-Verlag, Berlin, 199–210, 1991.