

Homogenization of Chromosomes Revealed by Oligonucleotide-Stickiness

Ayumu SAITO and Koichi NISHIGAKI*

Department of Functional Materials Science, Saitama University
255 Shimo-okubo, Sakura-ku, Saitama, Saitama 338-8570, Japan

*e-mail: koichi@fms.saitama-u.ac.jp

(Received: May 25, 2004; Accepted for publication: October 5, 2004; Published on Web: November 19, 2004)

Genomic DNA contains sequences that are relics of evolution together with functional protein sequences and regulatory operands. The mutual relationship of chromosomes that constitute the genome is of interest from the viewpoint of evolution as well as from that of systems biology. Chromosome origin, the frequency of chromosomal alteration, and the nature of these alterations can be addressed by analyzing *oligostickiness*, which is highly informative of the hidden properties of genome sequences. *Oligostickiness* analysis is an oligonucleotide-affinity test performed along the genome sequence. This analysis revealed that frequent recombination events occur during the course of evolution, which is a phenomenon inherent to chromosomes. This was further supported by the finding of chromosome homogeneity in the nucleus and by the assimilation of *oligostickiness* for pairs of complementary oligonucleotides. Genomes consisting of multiple chromosomes possess a common property where all chromosomes share the same tendency of *oligostickiness* for any kind of probe (oligonucleotide), indicating that these chromosomes all experienced random-mode frequent inter-/intra-recombinations. The phenomenon observed here, namely the homogeneity of chromosomes, was novel but it could be interpreted in terms of random-mode frequent recombination of chromosomes with a considerable contribution of repeated sequences. Therefore, *oligostickiness* is useful for analyzing the similarity of chromosomes.

Keywords: Genome sequence analysis, Nucleic acid thermodynamics, Chromosome evolution, Negentropy, Recombination

1 Introduction

The genome can be considered to be a chronicle of the evolutionary history of any organism. By scrutinizing the genome, we are able to follow the events that occurred during evolution. Some of these events must have had a positive influence on the evolution of the organism, whereas other events had no effect and so are considered 'neutral' [1]. Extensive gene-level analyses of homolog and analog proteins have elucidated the phylogenetic relationships between genes and organisms [2]. This well-established approach can reveal the frequencies of point mutations, deletions, and insertions that occurred in genes of interest and allow the determination of evolutionary rate, genetic distance, and other variables. On the other hand, as a system of genes, the genome should be

dealt with as a whole. Comparisons of genome sequences have recently disclosed a massive scale of gene arrangements that were caused by recombination events [3–5]. Furthermore, repetitive element analysis [6] and segmental duplication analyses [7, 8] of the human genome unveiled that they probably generated the current genome structures [6–8]. Thus, the genome is less static and less stable than previously believed, making it more important to understand these dynamic genome rearrangements as a whole. Recently, a measure called *oligostickiness* was introduced to characterize genomes [9, 10] and was useful for investigating the degree of recombination between chromosomes. Here, we applied this methodological tool to determine the similarity of chromosomes contained in the nucleus and we found that recombination phenomena occurred with high frequency.

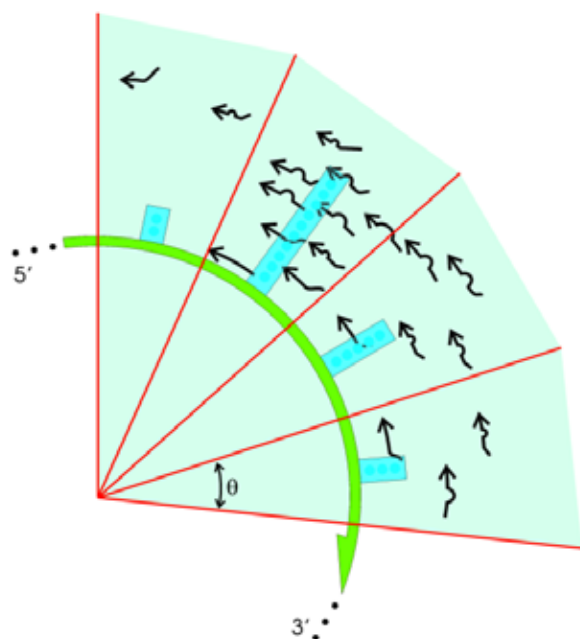


Figure 1. How to calculate *oligostickness*. *Oligostickness* is defined as the normalized frequency of binding to template DNA in various ways as presented here (also see Methods). Each binding structure indicates sufficiently stable binding to template DNA at a particular site. The stability of each structure is calculated thermodynamically [10] and shown with more stable structures placed on a lower layer. Frequency of probe-binding is accumulated within a sector of the angle θ , normalized by actual size of fractional template, and represented as a pillar (or spike), with the height proportional to the normalized frequency. For convenience, *oligostickness* is usually defined regarding registered genome sequence (or database sequence).

2 Methods

2.1 Calculation of *Oligostickness*

Oligostickness, σ , was calculated as explained in the legend to Figure 1. It can be expressed as follows:

$$\sigma = \frac{1}{n} \sum_{i=l_0+1}^{l_0+n} \delta(p, T(i)) \quad , \quad (0 \leq \sigma \leq 1) \quad (1)$$

where $l_0 + 1$ and n are the genome sequence positions at which the sampling region begins and the sampling size for *oligostickness*, respectively, and δ is a determinant that takes the value of 1 when the primer (p) binds stably to the i -th local sequence of the genome ($T(i)$), in other words, a fragmental sequence that has a fixed 5'-end at the sequence position i , or the value of 0 when not bound. In this formula, the $p - T$ binding is determined based on the thermodynamic stability of the $p - T$

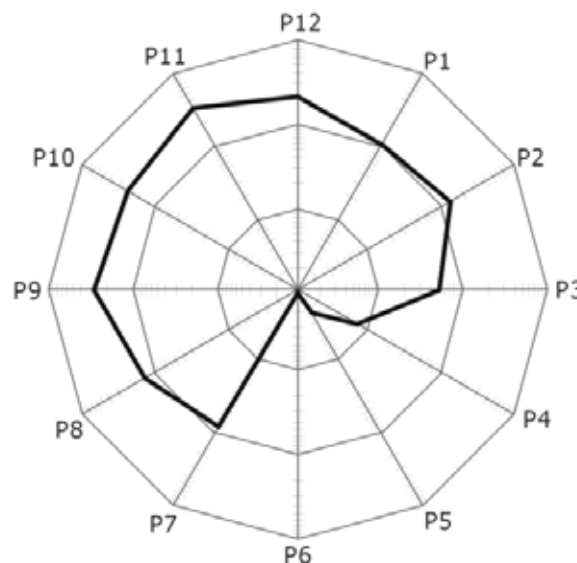


Figure 2. Spider-web chart of *oligostickness* of *Escherichia coli* genome. Probes were P1~ P12 (oligodeoxyribonucleotides written from 5' to 3'): P1, dGGGGTCGAGGGG; P2, dTGGGTGGGTGGG; P3, dGAGAGAGAGAGA; P4, dGCTAAAAAAAAA; P5, dAAAAAAAAAAAA; P6, dATATATATATAT; P7, dGTGCTGGGATTA; P8, dCCAGGCTGGTCT; P9, dCCG-GCCGGCCGG; P10, dGGGGTCGAGGCG; P11, dA-GACCGCGCCTG; P12, dACGACGACGACG. *Oligostickness* values are plotted on the radial axes.

complex under given conditions (25°C, 1M NaCl) [11]. Figure 2 shows the stickiness of 12 oligonucleotides to *Escherichia coli* genome DNA in the mode of a spider-web chart [10]. In this mode, the *oligostickness* of each oligonucleotide is expressed as a value between 0 (the center) and 1 (the outermost circle) and represents the frequency of binding of the oligonucleotide to the genome DNA (where 1 means that the oligonucleotide can bind at any site of the genome DNA with required or higher stability and zero means no binding throughout). The oligonucleotides were arbitrarily selected except with respect to the following considerations: i) oligonucleotides with different properties (G+C contents, sequence complexities, and thermodynamic stabilities) were selected as much as possible and ii) representative oligonucleotides from the viewpoint of *oligostickness* based on data collected from over 20 genomes were selected [10].

In addition, the obtained results are available at the following URL. <http://gp.fms.saitama-u.ac.jp/stickiness/>

2.2 Spider-web representation

The representation of chromosome/genome properties used here is called the ‘*spider-web*’ presentation of global *oligostickiness*’ [9]. Each global *oligostickiness* value with respect to a chromosome probed by a particular oligonucleotide was plotted on an axis radially extended from a common center, following a logarithmic scale (Figure 2). In this paper 12 axes per round were adopted with 12 different oligonucleotide probes. The nearby plots were connected with a line to define a characteristic pattern for each chromosome. This type of representation appears to be more effective in presenting a feature of a chromosome in depth than reducing it to a single figure (or value).

2.3 Genome sequences

The genome sequences were retrieved from databases published in relation to genome sequencing projects about each genome. The source of *A. thaliana* chromosome sequences was the Kazusa DNA Research Institute (CONTIGs: Chr1, pseudo v211200b; Chr3, MEC18 T15D2). The *E. coli* genome sequence was from the National Institute of Genetics and *H. sapiens* chromosome sequences were from the National Center for Biotechnology Information (CONTIGs were as follows: chromosomes Chr1, NT_004359.5/NT_004873.5; Chr2, NT_005194.5/NT_005204.5; Chr3, NT_005718.5/NT_005787.5/NT_005997.5/NT_006022.5/NT_022517.5; Chr4, NT_006051.5/NT_006052.5/NT_006098.5/NT_006316.5; Chr5, NT_006455.5/NT_006547.5/NT_006576.5/NT_023115.5; Chr6, NT_027049.2/NT_029309.1; Chr7, NT_007867.5/NT_007918.5; Chr8, NT_007988.5/NT_008157.5/NT_008227.5; Chr9, NT_008484.5/NT_023967.5; Chr10, NT_008682.5/NT_008895.5; Chr11, NT_009107.5/NT_009243.5/NT_009325.5/NT_009368.5; Chr12, NT_009711.5/NT_009775.5; Chr13, NT_009799.5; Chr14, NT_010101.5; Chr15, NT_010178.5/NT_010310.5/NT_010351.5/NT_024680.5; Chr16, NT_010530.5/NT_010552.5/NT_024827.5/NT_027182.2/NT_029459.1; Chr17, NT_010641.5/NT_010783.5; Chr18, NT_010874.5/NT_010990.5/NT_011054.5/NT_024981.5; Chr19, NT_011098.5/NT_011141.5/NT_011157.5/NT_011196.5/NT_011225.5; Chr20, NT_028391.3; Chr21, NT_011515.6;+A24 Chr22, NT_011523.7; ChrX, NT_011584.5/NT_011618.5/NT_011657.5/NT_011793.5; ChrY, NT_011896.6). For other information, see Table 1.

Table 1. Size and source of genome DNA sequences.

Sample	Size (base)	Reference and note
<i>A. pernix</i>	1,669,695	[12]
<i>A. thaliana</i> Chr 1	22,743,551	[13]
<i>A. thaliana</i> Chr 3	9,821,447	- Same as above -
<i>B. subtilis</i>	4,214,814	[14]
<i>C. elegans</i> Chr I	13,467,562	[15]
<i>C. elegans</i> Chr II	15,116,321	- Same as above -
<i>C. elegans</i> Chr III	12,476,799	- Same as above -
<i>C. elegans</i> Chr IV	15,919,001	- Same as above -
<i>C. elegans</i> Chr V	20,668,416	- Same as above -
<i>C. elegans</i> Chr X	17,432,311	- Same as above -
<i>D. melanogaster</i>	429,858 ^a	[16]
<i>E. coli</i>	4,636,552	NIG ^e
<i>H. influenzae</i>	1,830,138	[17]
<i>H. pylori</i>	1,667,867	[18]
<i>H. sapiens</i> Chr 1	6,906,110	NCBI ^f
<i>H. sapiens</i> Chr 2	6,163,434	NCBI
<i>H. sapiens</i> Chr 3	7,256,523	NCBI
<i>H. sapiens</i> Chr 4	7,611,217	NCBI
<i>H. sapiens</i> Chr 5	6,506,832	NCBI
<i>H. sapiens</i> Chr 6	6,705,352	NCBI
<i>H. sapiens</i> Chr 7	6,270,794	NCBI
<i>H. sapiens</i> Chr 8	7,083,280	NCBI
<i>H. sapiens</i> Chr 9	6,270,222	NCBI
<i>H. sapiens</i> Chr 10	6,709,963	NCBI
<i>H. sapiens</i> Chr 11	6,717,522	NCBI
<i>H. sapiens</i> Chr 12	6,308,324	NCBI
<i>H. sapiens</i> Chr 13	6,050,474	NCBI
<i>H. sapiens</i> Chr 14	6,050,829	NCBI
<i>H. sapiens</i> Chr 15	6,319,005	NCBI
<i>H. sapiens</i> Chr 16	6,389,377	NCBI
<i>H. sapiens</i> Chr 17	6,147,036	NCBI
<i>H. sapiens</i> Chr 18	6,408,213	NCBI
<i>H. sapiens</i> Chr 19	6,229,141	NCBI
<i>H. sapiens</i> Chr 20	2,426,223	NCBI
<i>H. sapiens</i> Chr 21	3,427,677	NCBI
<i>H. sapiens</i> Chr 22	2,728,705	NCBI
<i>H. sapiens</i> Chr X	6,810,333	NCBI
<i>H. sapiens</i> Chr Y	6,386,620	NCBI
Lambda phage	48,502 ^b	NCBI
<i>M. thermoautotrophicum</i>	1,751,377	[19]
<i>M. pneumoniae</i>	816,394	[20]
<i>P. falciparum</i> Chr 3	1,060,106 ^c	SC ^g
<i>P. horikoshii</i>	1,738,505	[21]
<i>S. cerevisiae</i>	12,069,299	[22]
<i>Synechocystis</i> sp.	3,573,470	[23]
Random 4636552-1	4,636,552 ^d	-
Random 4636552-2	4,636,552 ^d	-

^a Source: ‘CONTIG. p1 contig 1’

^b Source: ‘NUCLEOTIDE. NC.001416’

^c Source: ‘GENOME. NC.000521’

^d Artificially generated at random (equivalent to the size of *E. coli* genome)

^e National Institute of Genetics

^f National Center for Biotechnology Information

^g Sanger Centre

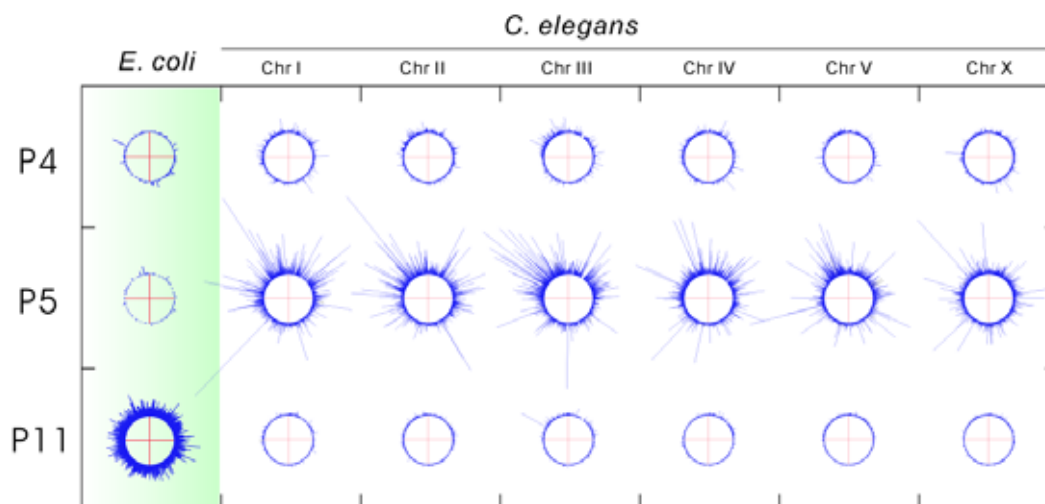


Figure 3. Map presentation of *oligostickness* against the six chromosomes of *C. elegans* and *E. coli* genome. Chromosomes are shown as a circle beginning at the top and running clockwise. A complete circle is equivalent to the entire chromosome and the diameter of each circle is drawn equal to a *oligostickness* value 0.36. Each *oligostickness* value is expressed by the length of a spike at each position where the running average of *oligostickness* along a 1/3600 part of the entire chromosome is shown. The chromosome nomenclature and the probes used are shown on the top and to the side, respectively. The sizes of the chromosomes are 4.6 (*E. coli*), 13.5 (Chr I), 15.1 (Chr II), 12.5 (Chr III), 15.9 (Chr IV), 20.7 (Chr V) and 17.4 (Chr X) in Mb. The sequences of the probes are: P4, dGCTAAAAAAAAA; P5, dAAAAAAAAAAAA; and P11, dAGACCGCGCCTG.

3 Results and Discussion

3.1 Chromosome homogenization

Genomes of various species can be characterized by *oligostickness* maps using a number of oligonucleotide probes [10]. Typical examples of *oligostickness* maps are shown for the *E. coli* and *C. elegans* genomes using three probes (P4, P5, and P11) in Figure 3. As the length of the spike is proportional to *oligostickness* σ (for definition see Methods), the probe, P11, can be considered to be of higher *oligostickness* to the *E. coli* genome than the other probes, P4 and P5. Aside from the *E. coli* genome, we have examined the genomes of other bacteria, such as *B. subtilis*, *H. influenzae*, *H. pylori*, and *S. aureus* (see Table 1. Data partly shown in Ref. 10). A clear advantage of *oligostickness* maps is that they can represent local *oligostickness*, γ , with ease, which can not be achieved by numerical and statistical approaches. Another benefit of this analysis is the fact that any size of genome (or chromosome) can be depicted by a circle of the same diameter, preserving Equation 1 when the whole genome (or chromosome) is divided into the same number of portions (e.g., 3600). Local *oligostickness* can be expressed as follows:

$$\gamma = \alpha \frac{(\text{spike length})}{(\text{circle diameter})} \quad (2)$$

where α is a constant arbitrarily set for the sake of clear representation (we adopted 0.36 for α throughout our representations). The size of the genome can be shown in proportion to its real genomic size (if so, Eq. 2 does not yet hold). The same presentation is used to represent six chromosomes of *Caenorhabditis elegans* (Figure 3). It is clear that the tendency of *oligostickness* is quite different between *E. coli* and *C. elegans*. However, it is remarkable that all six chromosomes of *C. elegans* have the same tendency (or texture [10]) against the three probes used here. We examined this further using 42 probes and with reference to the genomes of *Saccharomyces cerevisiae* and *Homo sapiens*. These results are shown as *spider-web* mode representations in Figure 4, where 12 probes were used in common. Figure 4B to D shows superimposed *spider-web charts* of all 16 chromosomes of yeast, 6 of nematode and 24 of human, respectively. All the *spider-web charts* are genome-specific except for the common dent around P4 to P6 corresponding to the *oligosticknesses* probed by AT rich sequences. Apart from the fact that each eukaryotic organism has a species-specific shape of *spider-web chart*, all of the chromosomes of each organism have essentially the same shape of *spider-web chart*. Especially, the 24 human chromosomes (22 autosomes and 2 sexual chromosomes) have very similar shapes (Figure 4D), considering that some vertices such as P1, P2, and P9-12, which are rather widely distributed, are connected in parallel to each chromosome

(thus, providing a proportionally expanded shape of the *spider-web chart*). In addition, we confirmed the same tendency for all the other probe oligonucleotides tested (additional 20 sequences). This is surprising since each chromosome must contain distinct genes consisting of different sequences. This indicates that chromosomes in a cell (*i.e.* in the nucleus) have a similar texture of sequence, or in other words possess a monochronous sequence. The rational explanation for this can come from only two sources: (i) all chromosomes originate from a common ancestor and (ii) frequent recombinations occur between the chromosomes. The former asserts a rather unnatural consequence, which is that all chromosomes evolved from the same sequence to different sequences and acquired novel functions. Segmental duplication [7] and chromosomal duplication must have occurred during the course of evolution, as originally suggested by Ohno [24]. However, it does not mean that the 24 human chromosomes evolved from a single ancestral chromosome confined in a cell, since this would be highly unrealistic. On the other hand, chromosome recombinations have been observed on various occasions, such as meiotic segregation [25] recombinational repair [26], SOS responses [26], and movable elements [6, 27]. Furthermore, recent observations of segmental duplications [7, 8] must have also contributed to generating the current genome state (it is necessary to note that chromosomes might have gathered and evolved prior to the establishment of the current species). Actually, the result shown here clearly demonstrates the possibility of frequent non-homologous recombinations between chromosomes. The ‘*Frequent recombination*’ hypothesis can be further supported by the fact that eukaryotic genomes contain a high ratio of non-coding regions such as introns and intergenic sequences that are not only permissible to non-homologous recombinations (which do not always lead to gene disruption) but also to a high probability of recombination. Here, we propose a natural interpretation of the phenomenon of chromosome homogenization: the ancestral chromosomes that had been recruited by successive duplication of chromosomes and/or by cell-fusion of different organisms evolved to their present state through a highly sophisticated mechanism of evolution, which must have involved recombinations such as exon [28] and module [15] shuffling. Interestingly, since the overwhelmingly larger portion of eukaryotic chromosomes are composed of non-coding sequences such as introns and intergenic sequences, the monochronous property observed here must be attributed mostly to such sequences, thus indicating that such non-coding sequences retain the genome-wide common yet species-specific property detected by *oligostickiness*. This is not yet interpretable but what is clear is that it can not be attributed only to highly repetitive sequences such as the *Alu* and L1 families (over 40% of the whole genome sequence in the human genome

[6]) since the nematode genome contains only 7% repetitive sequence [29, 30]. Repetitive sequences such as the *Alu* and L1 families do not show high *oligostickiness* to probes P2 and P5 and they can not substantially contribute to the high *oligostickinesses* of P2 and P5 in human chromosomes (Figure 4D). This is contrary to the hypothesis of repetitive sequence-caused homogeneity of chromosomes, although they may be responsible for a considerable contribution of homogeneity. Nevertheless, recombination seems to be the main cause of this phenomenon. From Figure 4, the featuring *oligostickiness* for each organism can easily be read. (i) P12 (dACGACGACGACG) is commonly sticky to yeast and nematode genomes but of very low *oligostickiness* to the human genome. (ii) P3 (dGAGAGAGAGAGA) is most sticky to the nematode genome. (iii) P5 (dA₁₂) is the best discriminator between bacteria and eukaryotes, since *oligostickiness* to the bacterial genome is much suppressed, whereas that to the eukaryote genomes is high. The advantage of this *oligostickiness*-based representation is that it allows such features to be easily recognized.

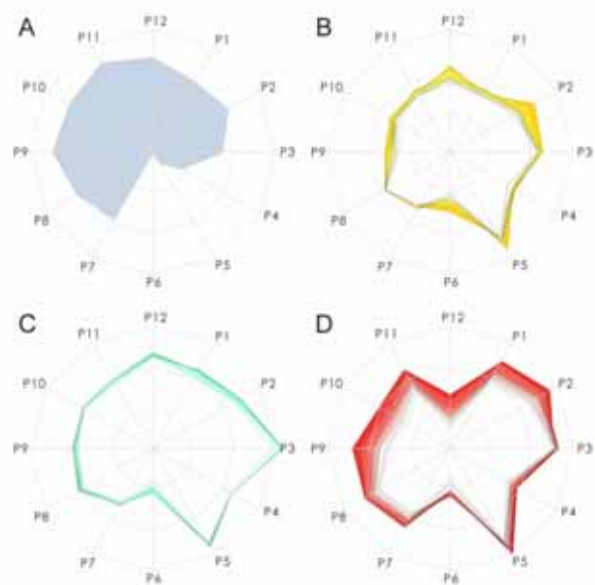


Figure 4. Chromosome structures of four organisms represented by a *spider-web chart* of *oligostickiness*. The same representation and symbols are used here as in Figure 2. In addition, *oligostickiness* values plotted on the radial axes are connected with a line to form a circle for each chromosome for multi-chromosomal genomes (B~D). These circles are superimposed. *Oligostickiness* is plotted on a logarithmic scale with three polygons crossing at 10^{-3} , 10^{-2} , and 10^{-1} (inner to outer). A, *Escherichia coli* genome; B, *Saccharomyces cerevisiae* (16 chromosomes); C, *Caenorhabditis elegans* (6 chromosomes); D, *Homo sapiens* (22 autosomes and X and Y chromosomes).

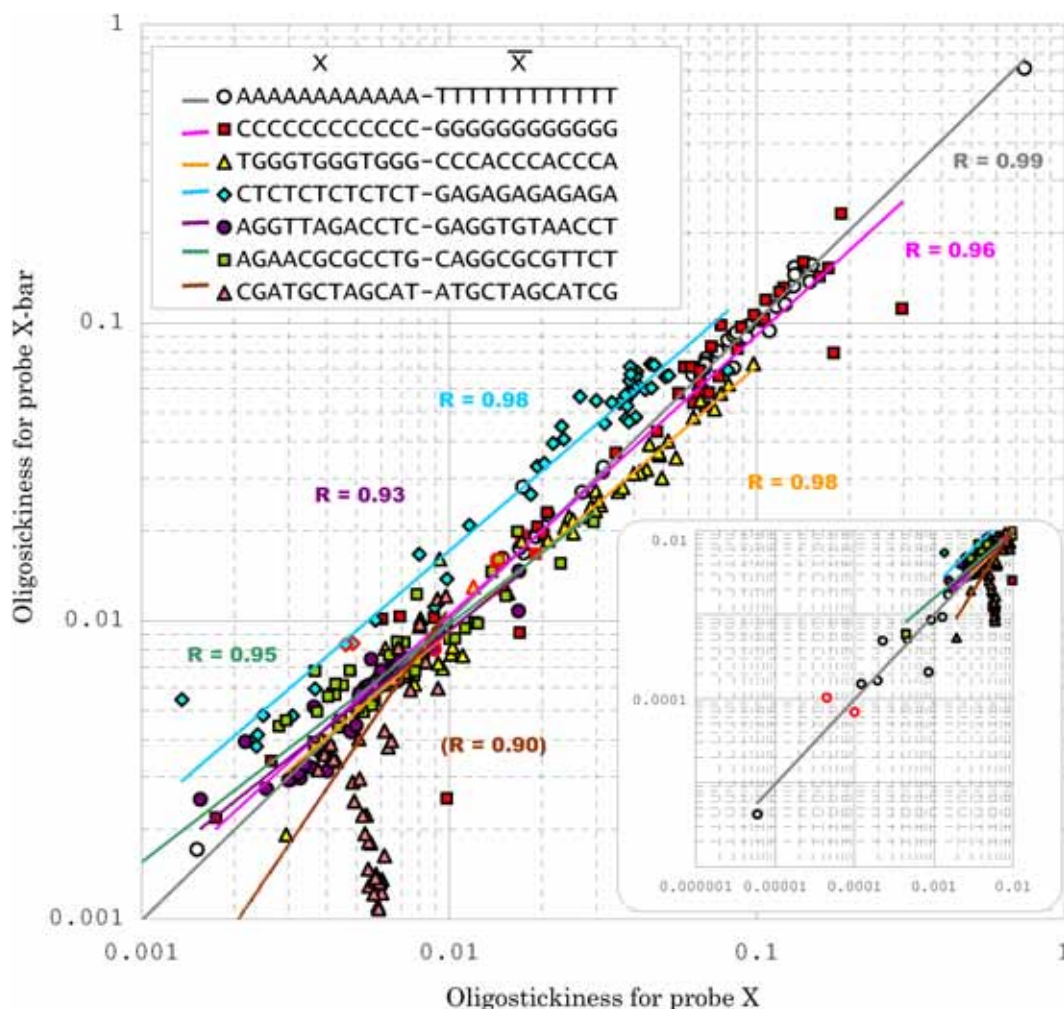


Figure 5. Correlations of *oligostickness* values for pairs of complementary probes obtained with 45 chromosomes from 16 species and 2 random sequences. Seven pairs of complementary oligonucleotides (probes) are adopted: colored symbols for these pairs (pairs of X and X-bar) are shown as an Insert in this figure (upper-left). Correlation coefficients (R) calculated for each set, which is composed of values obtained with 47 chromosomes or similar, are shown beside the least-square line. The width of the lines corresponds to the range of distribution of values (thus, values for probes of A₁₂/T₁₂ give the highest correlation coefficient (0.99) and widest range (0.000006~ 0.73)). Insert shows range of lower values of *oligostickness* (right-bottom). Global *oligostickness* is referred to here (in other words, averaged *oligostickness* over entire range).

3.2 Similar *oligostickness* between complementary probes

Another piece of evidence supports the frequent recombination hypothesis. Figure 5 shows that each pair of oligonucleotides of complementary sequences has very close *oligostickness* values (statistically shown as correlation coefficients (R): most of these correlation coefficients are very close to one). This relationship is valid for all recombinations of chromosomes and complementary probes with the sole exception of human chromosomes examined with the probe dCGATGCTAGCAT(=X)/ dATGCTAGCATCG(=X-bar) (this exception

is very prominent as can be seen by the vertically-stacked triangles around 0.006 of *oligostickness* for probe X in Figure 5). This indicates the extremely abnormal phenomenon that probe X binds equally to all human chromosomes while X-bar differs from chromosome to chromosome. So far we are unable to find any artifact to explain this phenomenon. Therefore, this might indicate a specific unknown function of this sequence for human chromosomes. On the other hand, the R values regarding non-complementary pairs of probes, such as dCGATGCTAGCAT/ dTACGATCGTAGC or dTGTCTGTCTGCTGC/ dCGTCGTCGTCGT, which have a relationship in terms

of mutually-inverted sequences, were not so high (0.51 and -0.1, respectively) (not included in Figure 5 for clarity). Another important point that should be noted is the wide range of *oligostickiness* (e.g. 0.000006~ 0.73 for probe dA₁₂ and 0.002~ 0.3 for probe dC₁₂). Considering the occurrence of frequent recombinations established above, this phenomenon can only be explained by introducing two modes (the parallel and antiparallel modes; if the orientation of either of two adjacent chromosomes is reversed (face about) just before recombination, then 'parallel' is converted to 'antiparallel') during recombination. These results must represent the well-shuffled state of those sequences, which is attained after vigorous recombinations between chromosomes irrespectively of the direction of DNA. This should not be taken simply as a matter of course because, in completely random artificial genome sequences (Table 1) as indicated in Figure 5 (the insert plotted with the circles outlined in red), the *oligostickiness* can be the same for a pair of complementary oligonucleotides that is, riding on the line of slope 1. Evidently, this is not valid for those sequences that have highly deviated from the random state like poly (dA), to which the probe dT₁₂ bound well but to which its complementary probe dA₁₂ did not. There is no need for probe X to have a similar *oligostickiness* as the probe X-bar with regard to the same chromosome. Therefore, this represents evidence that supports the frequent recombination hypothesis. The other wide distribution of *oligostickiness* among genomes is a type of representation of species diversity acquired during the course of evolution (leading to the concept of *negentropy* [31]) while equal *oligostickiness* for complementary primers (e.g. dCGATGCTAG-CAT/ dATGCATGCATGC) represents the opposite *entropic* effect during evolution.

Therefore, chromosomes have maintained their identity by rendering themselves isolated in the nucleus while they have been mangled to homogeneity with respect to their hyper-structure, as observed in this study, by entropic force due to confinement in the nucleus. The resultant chromosomes still retain the information essential for the organism, and the improvement of the quality of information occurs by rare chance events (*evolution*).

Finally, it is quite natural to suppose that any pair of chromosomes taken from two species in close relationship have mutually similar *oligostickiness* values since their genomes should have originated from a near, common ancestor. *Oligostickiness* analysis may be useful for estimating such an evolutionary distance between close species.

4 Conclusions

Software for the analysis of genome *oligostickiness* was applied to detecting the homogeneity of chromosomes that could be interpreted in terms of random-mode fre-

quent recombination of chromosomes with a considerable contribution of repeated sequences. The visual representation, known as the *spider-web chart* of *oligostickiness*, is useful in simplifying the huge amount of information from genome sequences to a manageable database and yet is sufficiently powerful to present the facts clearly. Several features of the human genome as well as those of the yeast and nematode genomes were depicted by *oligostickiness*, providing novel implications.

The authors are grateful for financial assistance from the Kouhi fund from the Japanese government.

References

- [1] M. Kimura, *The Neutral Theory of Molecular Evolution*, Cambridge University Press, New York (1983).
- [2] C. H. House and S. T. Fitz-Gibbon, *J. Mol. Evol.*, **54**, 539-547 (2002).
- [3] K. Hiramatsu, L. Cui, M. Kuroda and T. Ito, *Trends Microbiol.*, **9**, 486-493 (2001).
- [4] T. Baba, F. Takeuchi, M. Kuroda, H. Yuzawa, K. Aoki, A. Oguchi, Y. Nagai, N. Iwama, K. Asano, T. Naimi *et al.*, *Lancet*, **359**, 1819-1827 (2002).
- [5] E. Dawson, G. R. Abecasis, S. Bumpstead, Y. Chen, S. Hunt, D. M. Beare, J. Pabial, T. Dibling, E. Tinsley, S. Kirby *et al.*, *Nature*, **418**, 544-548 (2002).
- [6] W. H. Li, Z. Gu, H. Wang and A. Nekrutenko, *Nature*, **409**, 847-849 (2001).
- [7] E. E. Eichler, *Trends Genet.*, **17**, 661-669 (2001).
- [8] J. Cheung, X. Estivill, R. Khaja, J. R. MacDonald, K. Lau, L. C. Tsui and S. W. Scherer, *Genome Biol.*, **4**, R25 (2003).
- [9] K. Nishigaki and Y. Sakuma, *J. Chem. Software*, **2**, 96-107 (1994).
- [10] K. Nishigaki and A. Saito, *Bioinformatics*, **18**, 1153-1161 (2002).
- [11] W. Salser, *Cold Spring Harb. Symp. Quant. Biol.*, **42**, 985-1002 (1978).
- [12] Y. Kawarabayasi, M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekin, S. Baba, H. Kosugi, A. Hosoyama *et al.*, *DNA Res.*, **6**, 83-101, 145-152 (1999).
- [13] E. Asamizu, Y. Nakamura, S. Sato and S. Tabata, *DNA Res.*, **7**, 175-180 (2000).
- [14] F. Kunst, N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessieres, A. Bolotin, S. Borchert *et al.*, *Nature*, **390**, 249-256 (1997).

- [15] S. Kaneko, S. Iwamatsu, A. Kuno, Z. Fujimoto, Y. Sato, K. Yura, M. Go, H. Mizuno, K. Taira, T. Hasegawa *et al.*, *Protein Eng.*, **13**, 873-879 (2000).
- [16] M. D. Adams, S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, S. E. Scherer, P. W. Li, R. A. Hoskins, R. F. Galle *et al.*, *Science*, **287**, 2185-2195 (2000).
- [17] R. D. Fleischmann, M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick *et al.*, *Science*, **269**, 496-512 (1995).
- [18] J. F. Tomb, O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty *et al.*, *Nature*, **388**, 539-547 (1997).
- [19] D. R. Smith, L. A. Doucette-Stamm, C. Deloughery, H. Lee, J. Dubois, T. Aldredge, R. Bashirzadeh, D. Blakely, R. Cook, K. Gilbert *et al.*, *J. Bacteriol.*, **179**, 7135-7155 (1997).
- [20] R. Himmelreich, H. Hilbert, H. Plagens, E. Pirkl, B. C. Li and R. Herrmann, *Nucleic Acids Res.*, **24**, 4420-4449 (1996).
- [21] Y. Kawarabayasi, M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekine, S. Baba, H. Kosugi, A. Hosoyama *et al.*, *DNA Res.*, **5**, 55-76 (1998).
- [22] Nature The Yeast Genome Directory
http://www.nature.com/genomics/papers/s_cerevisiae.html
- [23] T. Kaneko, S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirosawa, M. Sugiura, S. Sasamoto *et al.*, *DNA Res.*, **3**, 109-136 (1996).
- [24] S. Ohno, *Evolution by Gene Duplication*, Springer-Verlag, New York (1970).
- [25] F. W. Stahl, *Genetic Recombination; Thinking About It in Phage and Fungi*, W. H. Freeman and Company, San Francisco (1979).
- [26] J. A. Nickoloff and M. F. Hoekstra, *DNA Repair in Prokaryotes and Lower Eukaryotes, In DNA Damage and Repair*, Humana Press, New Jersey (1998).
- [27] A. I. Bukhari, J. A. Shapiro and S. L. Adhya, *DNA Elements, Plasmids, and Episomes*, Cold Spring Harbor Laboratory Press, New York (1977).
- [28] J. A. Kolkman and W. P. Stemmer, *Nature Biotechnol.*, **19**, 423-428 (2001).
- [29] T. C. e. S. Consortium, *Science*, **282**, 2012-2018 (1998).
- [30] M. V. Katti, P. K. Ranjekar and V. S. Gupta, *Mol. Biol. Evol.*, **18**, 1161-1167 (2001).
- [31] N. Wiener, *Cybernetics or Control and Communication in the Animal and the Machine*, M.I.T. Press, Cambridge (1961).

オリゴスティッキネスによる染色体均質化現象の発見

斉藤 あゆむ, 西垣 功一*

埼玉大学工学部機能材料工学科, 〒 338-8570 埼玉県さいたま市桜区下大久保 255

*e-mail: koichi@fms.saitama-u.ac.jp

ゲノム DNA の塩基配列情報が明らかになるにつれて、ゲノムが進化的にどのように形成されてきたのかという興味深い問題への挑戦が始まっている。本研究はこのテーマにオリゴスティッキネス解析を用いて取組んだものであり、これまで他の方法で示唆されてきた染色体の組換え現象を全く新しい観点からはじめて定量性をもって明らかにしたものである。オリゴスティッキネス分析はゲノム配列に対するオリゴヌクレオチド配列の類似性検索の一種でコンピュータによる膨大な計算を前提にしている。今回、ゲノムが複数染色体で構成されている生物を代表し酵母、線虫およびヒトの染色体について調べたところ同一生物内の染色体間ではこれらすべての生物種においてオリゴスティッキネスのパターンが酷似することを発見した (Figure 4)。さらに興味深いことにはオリゴヌクレオチド配列が相補の関係にある場合のオリゴスティッキネスも高い相関を示すことを見出した (Figure 5)。これらの現象を合理的に解釈するものとして、核内染色体間に頻繁な組換えがあったことと、その結果として染色体均質化現象が起きたと考えられた。このことは、現存するゲノム DNA が一般に高頻度な組換え現象を経て形成されたものであることを示し、ゲノム進化のプロセスを考える上で重要な事実となる。

キーワード: ゲノム配列解析, 核酸熱力学, 染色体進化, ネグントロピー, 組換え