https://doi.org/10.20965/jaciii.2015.p0005

Paper:

Run-Length Encoding Graphic Rules Applied to DNA-Coded Images and Animation Editable by Polymerase Chain Reactions

Yuki Hara and Tomonori Kawano

The University of Kitakyushu
1-1 Hibikino, Wakamatsu-ku, Kitakyushu 808-0135, Japan
E-mail: kawanotom@kitakyu-u.ac.jp

[Received August 31, 2012; accepted September 2, 2013]

We previously proposed novel designs for artificial genes as media for storing digitally compressed image data, specifically for biocomputing by analogy to natural genes mainly used to encode proteins. A run-length encoding (RLE) rule had been applied in DNA-based image data processing, to form coding regions, and noncoding regions were created as space for designing biochemical editing. In the present study, we apply the RLE-based image-coding rule to creation of DNA-based animation. This article consisted of three parts: (i) a theoretical review of RLE-based image coding by DNA, (ii) a technical proposal for biochemical editing of DNA-coded images using the polymerase chain reaction, and (iii) a minimal demonstration of DNA-based animation using simple model images encoded on short DNA molecules.

Keywords: artificial gene, DNA animation, RLE

1. Introduction

DNA (**Fig. 1(a)**) was isolated 150 years ago [1], and its role of carrying genetic information was first reported in 1944 [2]. In 1953, James D. Watson and Francis Crick suggested the double-helix model for the DNA structure, ushering in the age of molecular biology [3]. Within the cells of living organisms, amazingly long DNA chains are packed into compact structures called chromosomes [4]. Information coded within this single set of molecules inside each micrometer-length human cell exceeds *ca.* 3 billion base pairs of DNA [5], thus naturally manifesting large-scale integration in the course of evolution. DNA contains genetic information on protein structures and the biological work algorithms used in the development and functioning of all living cellular organisms and DNA viruses. It is highly tempting to handle DNA as a means for "unplugged" information storage and processing.

The approaches to handling pieces of DNA as "unplugged" tools for storing and processing digital information vary widely. They include a series of studies applied to security-related areas such as DNA-based digital barcodes, watermarks, and cryptography.

2. DNA-Based Image Coding Model

Natural genes principally encode proteins [1]. We previously proposed a novel artificial gene-based biocomputing model enabling digitally compressed imaging data to be stored in editable DNA molecules [6].

In the case of natural genes, fragmented information coded on DNA is copied, or transcripted, onto a single strand of messenger RNA (mRNA) on which coded information is still in fragmented form. To obtain correct products (proteins) of genetic code from this fragmented information, mRNAs are spliced to remove noncoding sequences prior to the translating events on ribosomes.

The presence of noncoding regions of DNA with specific structures is required for the regulated function of natural genes (**Fig. 1(b)**). While genes in the biological system principally coded for proteins *via* translation from DNA to RNA and from RNA to amino acid sequences, the artificial genes designed here encode for numeric image data. A key difference between natural genes and artificial image-coding genes is the data-encoding mode. While the former uses strings of nucleotide bases, the latter employs RLE rules as discussed later [6].

In our model, image coding regions analogous to genes within synthetic DNA are fragmented and inserted within the noncoding DNA sequences – the sequences that can be dissected into the portions found within and outside the reading frame (**Fig. 1(c)**). Those noncoding sequences within the reading frame resemble the "introns" inserted among coding sequences, called "exons," in natural genes. Those sequences outside the frame resemble regulatory sequence motifs in promoter regions outside of natural genes. Noncoding DNA sequences designed for image coding/editing play at least five important roles – as (i) positional markers, (ii) tags for editing procedures such as cutting and pasting, (iii) start and end points for sequence copying events by polymerase chain reaction (PCR), (iv) labeling required for filing or addressing, and, as discussed later, (v) embedding hidden information, e.g., in steganography – "covered" or "hidden" writing.

To directly encode images on DNA, images of interest should be decomposed in the form of binary dot pictures. The two distinct modes for handling numerical data coding for each dot are the use of DNA sequences as strings

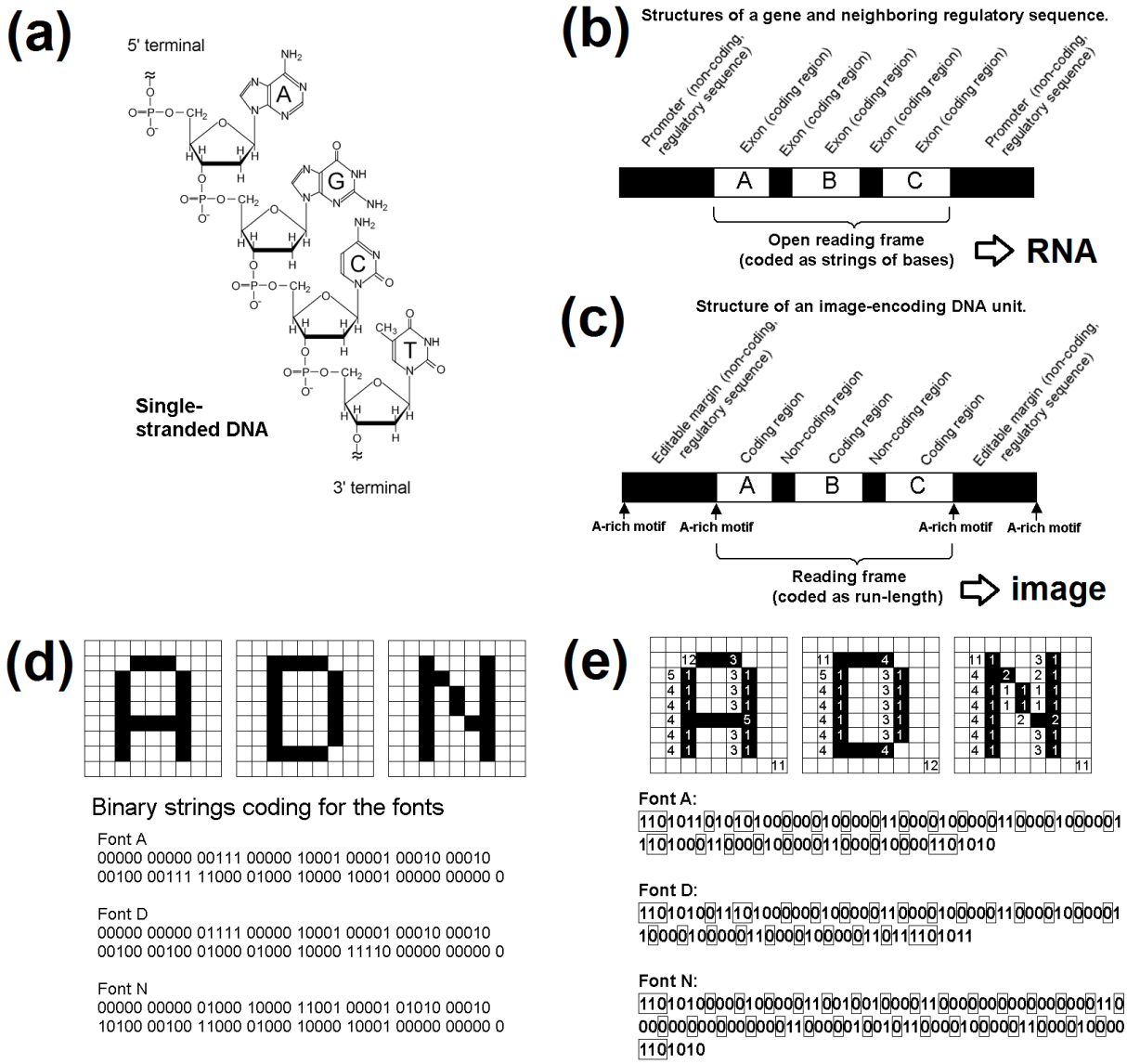


Fig. 1. Proposed DNA structures for coding images. (a) Chemical structure of single-stranded DNA. Coding and reading start from the 5'-terminal toward the 3'-terminal. (b) Schematic model for the structure of a gene and regulatory sequences coded on DNA. (c) Schematic model for an "image-coding DNA unit" structure [6]. Black boxes in the DNA chain represent image coding regions. Within the white box (noncoding regions) are (i) positional markers, (ii) tags for image editing procedures such as cutting and pasting, (iii) start and end points for copying events, (iv) labeling required for filing or addressing, and (v) steganographic information that is embedded. (d) and (e) represent two distinct image-coding protocols. (d) Two-toned model letter font images coded as strings of numbers (1, black; 0, white), to be ciphered within 9×9 blocks. (e) Model letter font images encoded by RLE (Wyle encoding). Runs consisting the letter fonts (A, D, N) defined on 9×9 block-squares are aligned. Runs are expressed in Wyle encoding. Boxed numbers, prefixes; other number, run length. Numbers shown are binary.

of binary data (**Fig. 1(d)**) and as a run-length encoding (RLE) medium (**Fig. 1(e)**).

Rauhe et al. [7] attempted to create digital DNA molecules representing binary data structures based on the programmable self-assembly of DNA oligonucleotides. In their work, plasmid – circular double-stranded DNA – was used as "memory" with programmability that enabled isolation, amplification, and readout based on common genetic techniques.

In their approach, however, oligonucleotide sequences were used as bits rather than single bases such as adenine (A), cytosine (C), guanine (G), and thymine (T). The size of information encoded or handled must therefore be largely limited.

In contrast, we demonstrated coding and deciphering of letter font images coded on DNA molecules [6] based on RLE by modifying Wyle's model [8]. **Figs. 2(a)–(d)** summarizes the flow of thought converting and compressing numeric runs into the DNA sequence. Briefly, unlike Wyle's approach, binary 1 and 0 had to be replaced with

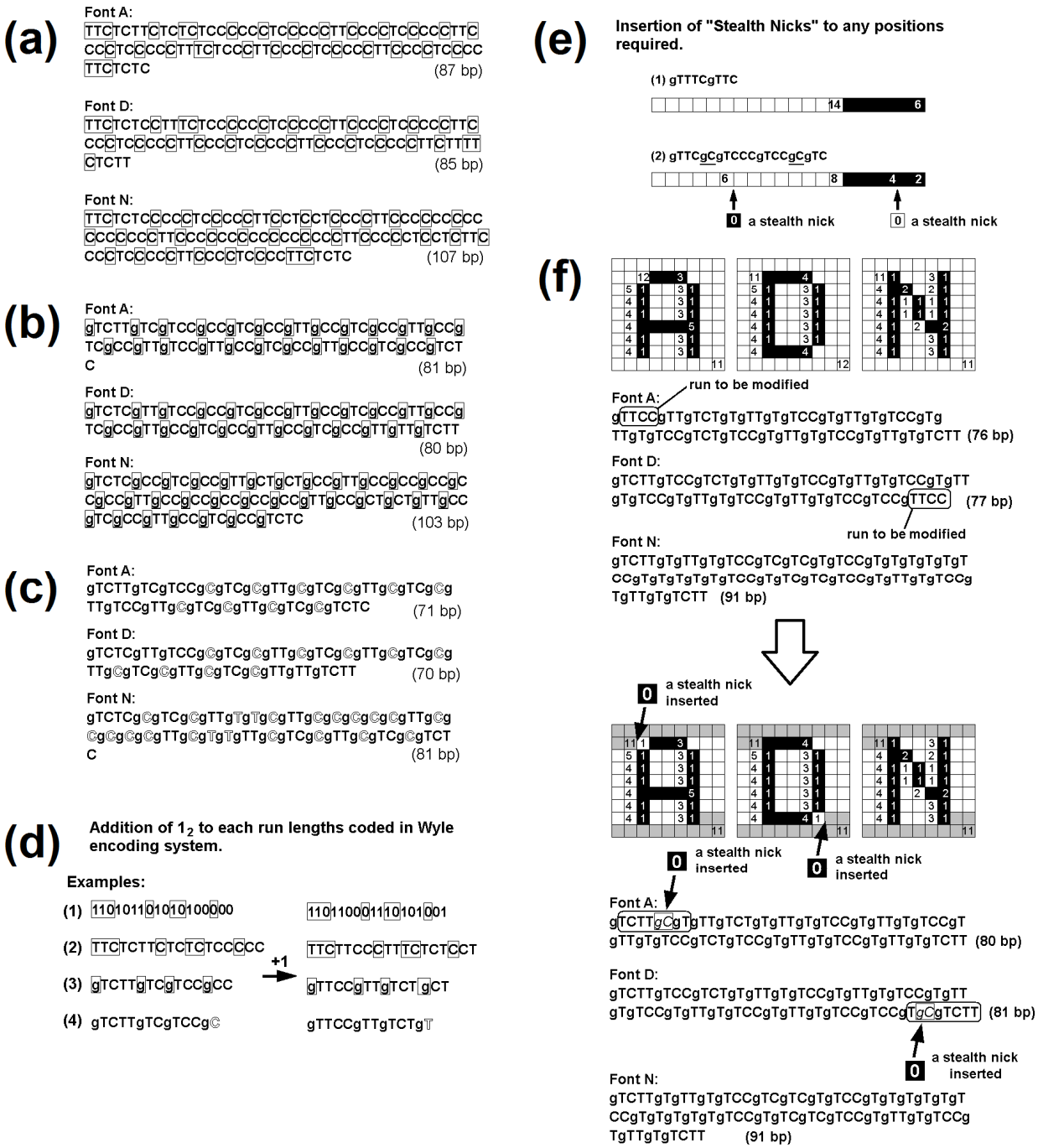


Fig. 2. Wyle's model-inspired RLE approaches using DNA. (a) 1 and 0 in Wyle encoding are replaced with DNA bases T and C. Prefixes are boxed. (b) Prefixes were replaced with DNA base G (boxed g). (c) Outlined letters indicate positions shorten by a simplified coding rule. (d) Use of lengthless (zero-length) runs is allowed after one-binary shift in code. (e) Insertion of lengthless runs (stealth nicks) into existing runs without distorting coded images. (f) Designing common terminal structures for all image-coding DNA chains by inserting two bases, *gC*, as a "stealth nick" within font-coding reading frames. After insertions of *gC* in Fonts A and D at 5'- and at 3'-termini, sequences gTCTTg and gTCTT are common to all DNA chains.

bases T and C (**Fig. 2(a)**). Prefixes to be inserted between runs were replaced by base G (**Fig. 2(b)**), and simplification was further made (**Fig. 2(c)**).

Wyle's RLE displays numbers (n) as ($n-1$), so instead of 1_{10}, 2_{10}, 3_{10}, 4_{10}, etc., 0_2, 1_2, 10_2, 11_2, etc., are used, prohibiting the use of zero as 0_2. We avoided this inconvenience by expressing the run-length as is (**Fig. 2(d)**).

3. Creation of Runs Without Length

The insertion of code (gC) encoding for zero (lengthless runs) is, in practice, allowed in the DNA sequence. By inserting such code for lengthless runs into sites of interest, new positional markers are created without interrupting apparent run-lengths displayed in decoded images

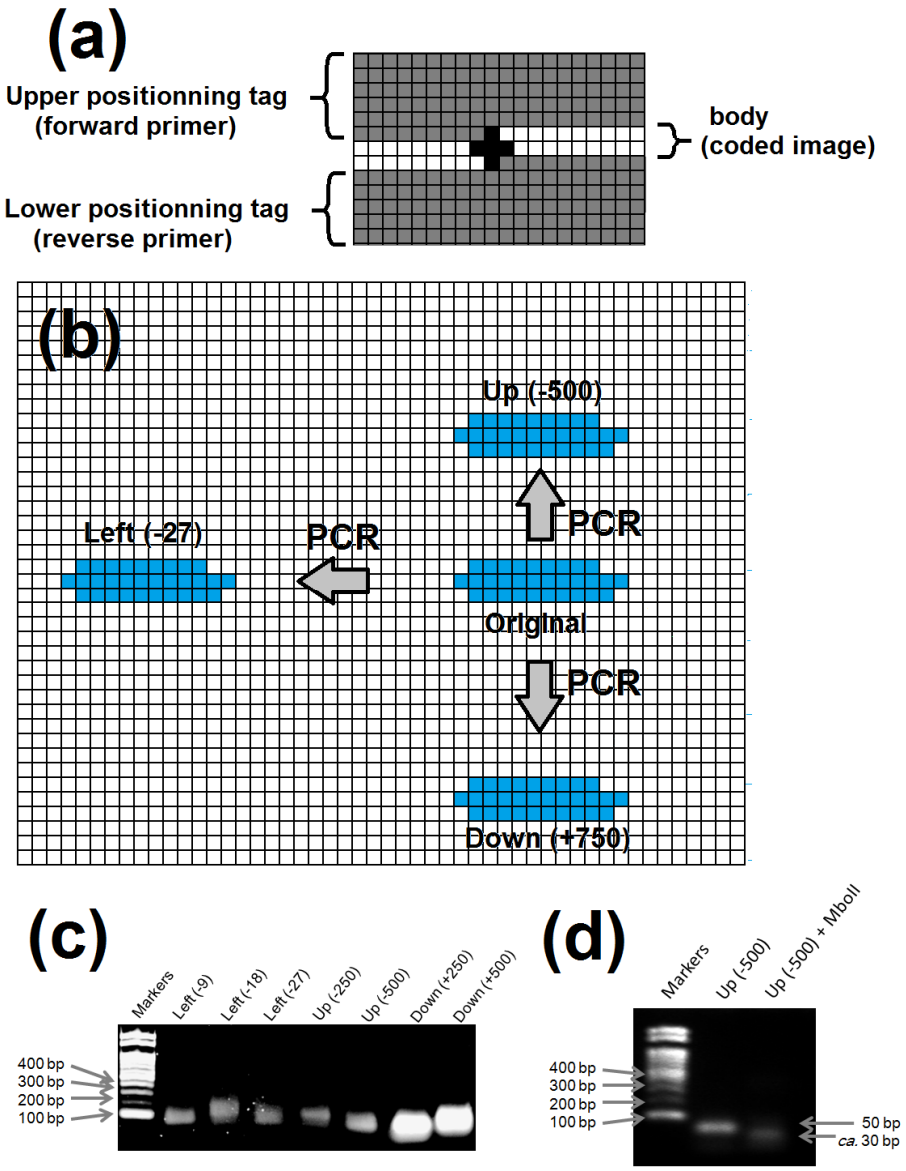


Fig. 3. PCR-based editing of RLE DNA images for simple animation. (a) Structure of a model image coded on DNA to be used for PCR-based amplification and editing. (b) Repositioning of DNA-coded images of bars on a 40 × 50 RLE imaginary square after designed PCR. (c) Image-coding DNA fragments edited after PCR. (d) Molecular evidence that Up (−500) molecules derived from original DNA molecule coding for the image of a bar on a 40 × 50 RLE imaginary square was processed successfully by PCR designed using a specific pair of primers.

(**Fig. 2(e)**). Note that the interruption of coded run-lengths by the insertion of single *gC* or odd-numbered *gC*s could not be detected graphically after decoding of images, so such interruptions are now referred to as "stealth nicks." We have shown that one of the attractive uses of these stealth nicks is in designing common terminal structures for all individual image-coding DNA chains by insertions of pairs of G (guanine used as a gap) and C (cytosine) as "stealth nicks" within image-coding reading frames [6]. Examples demonstrate – after the insertion of *gC* – that common sites of oligosequences designed to interact with common primers for PCR could be created in individual DNA chains (**Fig. 2(a)**).

4. DNA Animation Edited by PCR

As indicated above, images may be encoded in the DNA sequence as RLE numerical data. In addition to the expected decoding procedure, i.e., reading of images from binary code written on DNA, the use of DNA further enables us to perform biochemical data processing through the PCR-engineered termini of chains [6].

We hereby propose a procedure for PCR-mediated edition of DNA-coded images for the purpose of having the automated editing of animated DNA-coded images (**Fig. 3**). We divided a DNA-coded image into three parts, i.e., an upper positioning tag spanning the first blank run positioned between the first pixel and the image-coding

Table 1. PCR-dependently animated DNA molecules derived from original image-coding DNA.

Name	Run lengths (binary), <DNA sequence>		
	Upper positioning tag (blank run)	Body (coded image, bar)	Lower positioning tag (blank run)
Original	981 (1111010101)	9, 40, 12, 38, 10 (1001, 101000, 1100, 100111, 1010)	909 (1110001101)
	<5'-TTTTCTCTCTgTCCTgTCTCCCgTTCCgTCCTTTgTCTCgTTTCCCTTCT-3'>		
Left (-9)	972 (1111001100)	9, 40, 12, 38, 10 (1001, 101000, 1100, 100111, 1010)	918 (1110010110)
	<5'-TTTTCCTTCCgTCCTgTCTCCCgTTCCgTCCTTTgTCTCgTTTCCTCTTC-3'>		
Left (-18)	963 (1111000011)	9, 40, 12, 38, 10 (1001, 101000, 1100, 100111, 1010)	927 (1110011111)
	<5'-TTTTCCCCTTgTCCTgTCTCCCgTTCCgTCCTTTgTCTCgTTTCCTTTTT-3'>		
Left (-27)	954 (1110111010)	9, 40, 12, 38, 10 (1001, 101000, 1100, 100111, 1010)	936 (1110101000)
	<5'-TTTCTTTCTCgTCCTgTCTCCCgTTCCgTCCTTTgTCTCgTTTCTCTCCC-3'>		
Up (-250)	731 (1011011011)	9, 40, 12, 38, 10 (1001, 101000, 1100, 100111, 1010)	1159 (10010000111)
	<5'-TCTTCTTCTTgTCCTgTCTCCCgTTCCgTCCTTTgTCTCgTCCTCCCCTTT-3'>		
Up (-500)	481 (111100001)	9, 40, 12, 38, 10 (1001, 101000, 1100, 100111, 1010)	1409 (10110000001)
	<5'-TTTTCCCCTgTCCTgTCTCCCgTTCCgTCCTTTgTCTCgTCTTCCCCCCT-3'>		
Down (+250)	1231 (10011001111)	9, 40, 12, 38, 10 (1001, 101000, 1100, 100111, 1010)	659 (1010010011)
	<5'-TCCTTCCTTTTgTCCTgTCTCTCCCgTTCCgTCCTTTgTCTCgTCTCCTCCTT-3'>		
Down (+500)	1481 (10111001001)	9, 40, 12, 38, 10 (1001, 101000, 1100, 100111, 1010)	409 (110011001)
	<5'-TCTTTCCTCCTgTCCTgTCTCTCCCgTTCCgTCCTTTgTCTCgTTCCTTCCT-3'>		

runs, a body (the image-coding runs), and a lower positioning tag spanning the last blank run positioned between the image-coding runs and the last pixel in the coded image (**Fig. 3(a)**). Note that each upper and lower positioning tag must consist of a single blank run. By using pairs of primers corresponding to run-length-modified upper and lower positioning tags, the position of the body (coded images) on an imaginary plane could be determined, suggesting that objects to be printed may be moved based on the choice of primers, thus enabling an image to move, or migrate, vertically, horizontally, or diagonally.

Here, we report on our minimal demonstration of DNA-based animation using a simple model image (bar) encoded on short DNA molecules designed to derive minimal animated images on a 40×50 RLE imaginary square (**Fig. 3(b)**). The original image to be printed on the 40×50 RLE imaginary square was a bar (position of the bar labeled "Original," **Fig. 3(b)**), coded by a 50 bp double-stranded DNA sequence (5'-TTTTCTCTCTgTCCTgTCTCCCgTTCCgTCCTTTgTCTCgTTTCCCTTCT-3') with 7 runs – i.e., 981, 9, 40, 12, 38, 10, and 909 in decimal and 1111010101, 1001, 101000, 1100, 100111, 1010, and 1110001101 in binary. Seven different molecules coding for repositioned images designated Left (-9), Left (-18), Left (-27), Up (-250), Up (-500), Down ($+250$), and Down ($+500$) were created by PCR (**Fig. 3(c)**; **Table 1**).

To ensure that PCR products represented, without doubt, the molecules coding modified images, we examined PCR-mediated changes in the DNA sequence through digestion of the newly obtained band of DNA with restriction enzymes. The digestion of the Up (-500) molecule with an MboII restriction enzyme is one such example (**Fig. 3(d)**). Since no MboII restriction site exists in the original sequence, the PCR-dependent creation of an MboII site on the Up (-500) molecule suggests that the coded image was successfully repositioned after PCR.

Recently, we have been engaged in the studies of biologicals components such as living cells [9], proteins [10], and DNA [6] to be used as tools or media in bioinformatics. In the present study, we demonstrate, for the first time we know of, that DNA can be used as the platform for coding images and biochemically editable animation. In future work, we plan to testify that DNA can be used as the carrier of both the gene and of memes as first coined by Richard Dawkins [11].

Acknowledgements

This work was supported in part by a grant from the Regional Innovation Strategy Support Program implemented by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

References:

[1] R. Dahm, "Discovering DNA: Friedrich Miescher and the early years of nucleic acid research," Human Genetics, Vol.122, No.6, pp. 565-581, 2008.

[2] O. Avery, C. MacLeod, and M. McCarty, "Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Inductions of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III," J. Exper. Medicine, Vol.79, No.2, pp. 137-158, 1944.

[3] J. D. Watson and F. H. Crick, "A structure for deoxyribose nucleic acid," Nature, Vol.171, pp. 737-738, 1953.

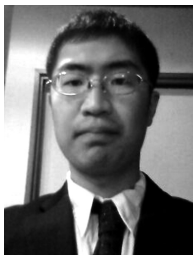
[4] E. A. Carlson, "Defining the gene: an evolving concept," Amer. J. Human Genetics, Vol.49, No.2, pp. 475-487, 1991.

[5] J. C. Venter et al., "The sequence of the human genome," Science, Vol.291, pp. 1304-1351, 2001.

[6] T. Kawano, "Run-length encoding graphic rules, molecular biologically editable designs, and steganographical on-image numeric data embedment for DNA-based cryptographical coding system," Commun. Integr. Biol., Vol.6, No.2, e23478, 2013.

[7] H. Rauhe, G. Vopper, U. Feldkamp, W. Banzhaf, and J. C. Howard, "Digital DNA molecules," Proc. 6th DIMACS Workshop on DNA Based Computers, Leiden, Netherlands, pp. 13-17, 2000.

[8] H. Wyle, T. Erb, and R. Banow, "Reduced-time facsimile transmission by digital coding," IRE Trans. on Communication Systems, Vol.9, No.3, pp. 215-222, 1961.
[9] T. Kawano, F. Bouteau, and S. Mancuso, "Finding and defining the natural automata acting in living plants: Towards the synthetic biology for robotics and informatics in vivo," Commun. Integr. Biol., Vol.5, No.6, pp. 519-526, 2012.
[10] T. Kawano, "Biomolecule-assisted natural computing approaches for simple polynomial algebra over fields," ICIC Express Lett., Vol.7, No.7, pp. 2023-2028, 2013.
[11] S. Blackmore, "The meme machine," Oxford University Press, Oxford, UK, 1999.



Name:
Yuki Hara

Affiliation:
Faculty of Environmental Engineering, The University of Kitakyushu

Address:
1-1 Hibikino, Wakamatsu-ku, Kitakyushu 808-0135, Japan

Brief Biographical History:
2013- Graduate Student, Graduate School of Environmental Engineering, The University of Kitakyushu

Name:
Tomonori Kawano

Affiliation:
Faculty of Environmental Engineering, The University of Kitakyushu

Address:
1-1 Hibikino, Wakamatsu-ku, Kitakyushu 808-0135, Japan

Brief Biographical History:
2000- Post-doctoral Scientist, Institut National de la Recherche Agronomique (INRA), France
2001- Research Associate, Graduate School of Science, Hiroshima University
2003- Associate Professor, Graduate School of Environmental Engineering, The University of Kitakyushu

Main Works:
• "Run-length encoding graphic rules, molecular biologically editable designs, and steganographical on-image numeric data embedment for DNA-based cryptographical coding system," Communicative and Integrative Biology, Vol.6, e23478, 2013.

Membership in Academic Societies:
• Information Processing Society of Japan (IPSJ)
• Society of Plant Signaling and Behavior
• Japanese Society of Agricultural, Biological and Environmental Engineers and Scientists