ibc Interdisciplinary Bio Central

Open Access

IBC 2013;5:5, 1-3 • DOI: 10.4051/ibc.2013.5.2.0005

ESSAY

# Development of Information Biology (III)

**Yoshio Tateno***

School of New Biology, Daegu Gyeongbuk Institute of Science and Technology, Daegu, Korea

**Subject areas;** Bioinformatics/Computational biology/Molecular modeling

**Author contribution;** Y.T. wrote this article.

***Correspondence** and requests for materials should be addressed to Y.T. (yt.tateno@gmail.com)

**Editor;** Hong Gil Nam, Daegu Gyeongbuk Institute of Science & Technology, Korea

**Received** June 13, 2013;
**Accepted** June 27, 2013;
**Published** June 28, 2013

**Citation;** Tateno, Y. Development of Information Biology (III). IBC 2013, 5:5, 1-3. doi: 10.4051/ibc.2013.5.2.0005

**Competing interest;** All authors declare no financial or personal conflict that could inappropriately bias their experiments or writing.



## SYNOPSIS

Introduced were two biological investigations in which information biology played a significant role. In the first case independent findings in cancer research over a long period were united and organized by information biology and led to the outcome that was subject to a Nobel Prize. The outcome has revealed that the cause of human cancer is located in the genome in a dormant condition. The second case shows how to elucidate the function of an unknown DNA sequence or ORF in prokaryotes by a large – scale computer homology search and analyses. For the elucidation the International DNA Databases and a large – scale computer were two key factors.

| *Class* | *Number* |
|---|---|
| *AAAA-A* | 431,672 |
| *BBBB-B* | 10.254 |
| *C* | 7,511 |
| *D* | 107,382 |
| *X* | 697,331 |
| *Total* | 1,254,150 |

**Key Words:** oncogene; proto-oncogene; PDGF; ORF; computer homology analysis

We now want to know how information biology participates in biology. In the following, given are two examples in which information biology played a significant role.

1. Origin of oncogene: In 1911 Payton Rous discovered the virus gene that caused tumor in chickens.[1] The virus was named after Rous as the Rous sarcoma virus. For the discovery Rous was awarded the Nobel Prize for Physiology or Medicine in 1966, more than 50 years after it. The reason for the delay is considered as the fact that the Nobel committee wrongly awarded the Prize for Physiology or Medicine to J. Fibiger for his discovery of the *Spirotera carcinoma* in 1926. The discovery stated that *Sprotera carcinoma*, a roundworm, caused carcinoma, which was totally outrageous. Thereafter, the committee had been too cautious not to make another mistake in awarding particularly in cancer research. One has to live long, if one wants to receive a Nobel Prize.

   The virus gene was named oncogene by R. Huebner and G. Todaro of National Cancer Institute, USA in 1969. Then, in 1976, M. Bishop, H. Varmus and their colleagues at University of California in San Francisco published a paper which stated that they could activate a proto-oncogene to change it into an oncogene that caused cancer.[2] It is now known that a proto-oncogene is activated when it is mutated or expressed more than the normal level.

   Platelet derived growth factor (PDGF) circulates in the blood vessels in humans and other mammals, and participates in hemostasis (to stop bleeding), once they are injured. PDGF is known also as a major growth factor in humans. R. Doolittle and his colleagues at University of California in San Diego searched for a similar protein sequence to PDGF by using the Needleman-Wunsch method at a protein/DNA sequence database, and unexpectedly found that it was very close to p28-sis, which was encoded by v-sis gene in the Rous sarcoma virus. Why was the growth factor so similar to v-sis, an oncogene?

   The answer Doolittle and his colleagues reached was that v-sis was newly included in the virus genome by recombination between the virus genome and PDGF gene or a similar gene of the virus's host (wooly monkey). The answer means that the oncogene originated from wooly monkey. They published their results and discussion in 1983.[3]

   Bishop, Vermus and their colleagues demonstrated that an oncogene stimulated cell division when an embryo developed, and ceased its function and became dormant as a proto-oncogene when the embryo grew up to an adult, and that when a proto-oncogene was activated again by something in an adult, it caused cancer that endlessly stimulated cell division from one tissue to another. For that achievement Bishop and Vermus won the Nobel Prize for Physiology or Medicine in 1989. The paper by Doolittle and his colleagues must have supported their achievement.

2. Finding the function of a DNA sequence: When I was in charge of DDBJ, we organized a research team there in 2005 for the purpose of elucidating the functions of submitted DNA sequences with unknown functions. Many researchers submitted their sequence data without describing their functions to DDBJ and the other two databases in Europe and USA (see Development of Information Biology I). Actually, many submitters did not know the functions of their sequences.

   Our first task was then to select complete bacterial genomes from the data at DDBJ, because most bacterial genomes were completely sequenced, and their genes were annotated more thoroughly than those of eukaryotes. The total number of the sequences thus selected was 183. By applying Glimmer2 to the selected sequences, we picked up 1,254,150 open reading frames (ORFs) or protein coding genes from the 183 sequences. Then, the question was whether those ORFs were genuine or not. To answer the question, we first translated the ORFs into the proteins, and examined if the proteins had authentic motifs by using InterProScan of the UniProt database in Switzerland. After the examination we classified the ORFs into the following classes: AAAA) the ORF in question was a known protein having a known motif and at least one similar sequence in DDBJ with the similarity of 70% or more, AAA) it was the same as AAAA except that the motif was unknown, AA) it was the same as AAA except that it had no motif, A) it was the same as AA except that it was a hypothetical protein. BBBB) it was the same as AAAA except for no similar sequences, BBB) it was the same as AAA except for no similar sequences, BB) it was the same as AA except for no similar sequences, B) it was the same as A except that it had no similar sequences. C) it was the same as A except that it had no motifs, D) it was the same as C except that it was an unknown protein, X) it was a totally unknown sequence with no motifs. The classification took a very long CPU time even by using a large-scale computer. The Table 1 shows the 1,254,150 ORFs classified into the each category mentioned above (see Kosuge *et al.*[4] for more details).

It is interesting to ask about how many genes exist in the all prokaryotes, and then in the all organisms on the Earth, because all organisms ever existed in the past and exist now on the Earth originated from a common ancestor with some genes. How the

**Table 1.** Classification of the ORFs in the order of clarity of functions

| Class | Number |
|---|---|
| AAAA-A | 431,672 |
| BBBB-B | 10.254 |
| C | 7,511 |
| D | 107,382 |
| X | 697,331 |
| Total | 1,254,150 |

genes in the common ancestor have evolved into so many genes in the present organisms?

## REFERENCES

1. Rous, P. (1911). A Sarcoma of the Fowl Transmissible by an Agent Separable from the Tumor Cells. *J Exp Med* 13, 397-411.
2. Stehelin, D., Guntaka, R. V., Varmus, H. E., and Bishop, J. M. (1976). Purification of DNA complementary to nucleotide sequences required for neoplastic transformation of fibroblasts by avian sarcoma viruses. *J Mol Biol* 101, 349-365.
3. Doolittle, R. F., Hunkapiller, M. W., Hood, L. E., Devare, S. G., Robbins, K. C., Aaronson, S. A., and Antoniades, H. N. (1983). Simian sarcoma virus onc gene, v-sis, is derived from the gene (or genes) encoding a platelet-derived growth factor. *Science* 221, 275-277.
4. Kosuge, T., Abe, T., Okido, T., Tanaka, N., Hirahata, M., Maruyama, Y., Mashima, J., Tomiki, A., Kurokawa, M., Himeno, R., et al. (2006). Exploration and grading of possible genes from 183 bacterial strains by a common protocol to identification of new genes: Gene Trek in Prokaryote Space (GTPS). *DNA Res* 13, 245-254.