

## The HIV "9749" and HBV "3215" issues and beyond over the past three years

Chun Yang, MD., [Energinity.com](http://Energinity.com)

20 E. Brookline St. Suite 86, Boston, MA 02118

September 12, 2012

This paper reports how the fixed genome size of each species as a doctrine may have been on the edge of revision. Before replication, each individual of the same species or individual may reorganize its genome according to its environment, and therefore, the replicated genome may vary in both size and nucleotide sequences. See Yang, C. DNA transtruction: A major mechanism of genome reorganization before replication. <[www.energinity.com/2010proceedings1.pdf](http://www.energinity.com/2010proceedings1.pdf)>. Here I use two viruses to further demonstrate the principle of DNA transtruction - an RNA virus, Human Immunodeficiency Virus type 1, HIV-1, which has been well documented as a fixed 9749 nucleotides of its RNA genome, but now has been found not the case at all but varied with its genome sizes, and a DNA virus, Hepatitis B Virus, HBV, which has been well documented as a fixed 3215 nucleotides of its DNA genome, but now has been found may not the case, and more and more reports show that HBV genome size has been varied according to each resources.

2010

The HIV "9749" and the HBV "3215" stories and the size variation of viral genomes have been an interesting one, but I did not have time to write them down until now.

My study of consciousness using virus as the simplest model has led me to conclude that the majority of genomes may be changed with their both sizes and base sequences, in particular, when the internal and external environments are changed – a theory is much different from the standard textbooks that DNA replications or RNA reverse transcriptions are always identical size with their parent origin.

In the early 2010, I figured out that the genome can be defined as two parts, one is stable region (S), and another is adjustable region (A), or variable part. Taking virus for example, HIV most likely will show different sizes from different patient or resources in addition to sequence mutation variations. I went to check the gene bank. There were a lot of "9749" nucleotides sequences submitted (actually the majority), and I thought most of them probably due to a contamination (or a type of questionable practice).

I began to talk about the size variations of genomes after replications to a few scholars including students at an annual mini symposium 2010, of Biophysics, Harvard main campus. One student from MGH told me clearly that the HIV genome was a fixed size of base pairs around 9k, but exactly he could not remember. I told him it might actually be varied through each life cycle. He did not agree with me. I also talked to another student who was working on a commonly used plasmid. I told her the same idea of mine. She was also dubious.

I checked the web, books, introduction courses, or websites etc. all of them I checked have stated the same - HIV-1 RNA genome is 9749 nucleotides. I was surprised – how differently my research conclusions have been from the recorded gene bank database.

It comes out that it has been a doctrine that genome sizes are fixed at each type of species, including virus. When the genome is replicated it always replicated the same original size. Hence my recent research results have challenged the long time doctrine.

I only generated two sets of data on March 14, 2010 in my notebook. Using HBV, a DNA virus, as the first example, I found most of the submitted in Gene bank are 3215 nucleotides. From about 600 entries I found 6 reports that are not 3215, ranges from 2043 to 3248. Then I used HIV, an RNA virus, as the second example, and I found most of the submitted in Gene bank are 9749 nucleotides. From about 600 entries, I found 6 reports that are not 9749, ranging from 6859 to 10280. I collected these data for them

be principle proven at that time – I have developed DNA transtruction etc. theories, saying genomes of organisms can well be changed to adapt the environment during replications,  
< [www.energinity.com/2010proceedings1.pdf](http://www.energinity.com/2010proceedings1.pdf)>

and the proposed transtruction principle has been part of the complementary theory of the central dogma, which in turn has been part of the whole theory of faith science.

2011

Early last year, I talked to Dr. Bruce Walker before a pathology seminar he was giving at Harvard Medical School (a meeting room at Harvard Institute of Medicine, January 26, 2011). He immediately expressed in agreement with me on my idea of "possible contamination" about the fixed sequence size of HIV.

About one month later in an afternoon symposium held on HMS, "Next Generation Sequencing of HIV" (February 25, 2011), I talked to Dr. Todd Allen, who told me he was working on the variations of HIV for many years and he was sure the sequences were varied. But when I told him the sizes should be varied as well - he, like all others I have talked with, seemed not agreeing and showed a bit surprising. But Dr. Allen was nicely asking me to work together on some data. Though I have thought I should follow up with him, I have not done so due to my other issues. I should do it after I finish this essay. (I shared with Dr. Walker and Dr. Allen very recently, Sep 11, 2012).

Sometime later, I dropped by the annual mini symposium of biophysics at Harvard again for a very short time. I happened to see a presentation about virus mutations and different environmental conditions. I did not follow up in details however.

Sometime even later last year, I checked the NIH gene bank again and found that most of the "9749" submissions were gone. Listed sequences are more believable to me, namely, HIV-1 isolates with various sizes. By this time, after my sharing my conclusion with the community, it had been for more than a year.

2012

Today, September 11, 2012, I checked the NIH gene bank again, and I found no 9749 nucleotides entries for HIV-1. The textbook doctrine of HIV-1 virus sequence of 9749 nucleotides is gone in the NIH Gene bank.

The submissions of HIV virus sequences, as compared my searched 2010 entries with that of today, have been dramatically increased, by nearly doubling the total entries in less than two and a half years (2612 of March 14, 2010 vs. 4942 of September 11, 2012). These are not adjusted by the removing the entries of 9749s.

I also checked HBV entries. In less than two and a half years' time, the entries have also increased by nearly 43 percent, (3200 of March 14, 2010 vs. 4554 of September 11, 2012). The majority HBV entries are no longer 3215 nucleotides. Among the 200 recent entries I found 37 are 3215, 117 are not 3215, though they are all complete genomes, and the remains of the 200 entries are fragments.

What had happened behind the scene was unclear to me so far, though the part of my self is clear to me.

Apparently the past few years' revolutionary understanding of the size variations of viral genomes has greatly promoted the more accurately sequencing of the viral genomes as well as the submissions to NIH gene bank. The significant increasing of the recent submissions should have been studied and further characterized by researchers.

Disclosed September 11, and modified September 22, 2012