Understanding SARS with Wolfram Approach

Da-Wei LI1,2, Yu-Xi PAN1,2, Yun DUAN2,1, Zhen-De HUNG1, Ming-Qing XU2,1, and Lin HE2,1*

1Bi-X Life Science Research Center, Shanghai Jiao Tong University, Shanghai 200030, China; 2Institute for Nutritional science, the Chinese Academy of Sciences, Shanghai 200031, China)

Abstract Stepping acquired immunodeficiency syndrome (AIDS), severe acute respiratory syndrome (SARS) as another type of disease has been threatening mankind since late last year. Many scientists worldwide are making great efforts to study the etiology of this disease with different approaches. 13 species of SARS virus have been sequenced. However, most people still largely rely on the traditional methods with some disadvantages. In this work, we used Wolfram approach to study the relationship among SARS viruses and between SARS viruses and other types of viruses, the effect of variations on the whole genome and the advantages in the analysis of SARS based on this novel approach. As a result, the similarities between SARS viruses and other coronaviruses are not really higher than those between SARS viruses and non-coronaviruses.

Key words genome sequence; SARS; visualization; Wolfram approach

In this work, we tried to understand the pathogenesis of SARS, the world’s threat [1–14] using a complete novel approach [15–20], or Wolfram approach which was systematically described in the book entitled “A New Kind of Science” in 2001 and has drawn extensive attention in the world [15]. In contrast with the traditional methods of DNA sequence comparison, Wolfram approach was based on the concept that simple rules are able to produce highly complicated behaviour such as dynamically viewing alterations on visualization, including transposition, insertion, deletion and duplication, etc., in a whole genome scale, and even in a single base scale when the base precisely located. Furthermore, it has become possible to make progress on a remarkable range of fundamental issues of lives that have never been successfully studied by any of the existing sciences based on traditional mathematical rules, which are limited in exploring the complex behaviour in a typical biological system. For example, the evolutionary theory cannot really or completely explain the origin of complexity of biological system [15]. Research and speculation in living organisms at a molecular level that was normally neglected by Wolfram, have little success for the explanation of complexity in lives.

With Wolfram approach-based method, we explored both the simple rules and a special rule from the 256 rules suggested by Wolfram [15] to avoid traditional intuition that the behaviour must be simple if the rule for a program is simple. This is not true from the data demonstrated by both Wolfram’s work and our own work. The remarkably simple rule can actually capture the essential mechanisms responsible for complex phenomena in living organisms.

In order to gain an insight into SARS, we analyzed DNA sequences of different viruses in detail by the simple rules, initial conditions and highly complex behaviour of the final images were studied visually.

Materials and Methods

Sequences of SARS viruses and other related viruses
All studied sequences including 13 SARS viruses were downloaded from free database of National Center for Biotechnology Information (NCBI):
The genomes of SARS viruses in Fig. 1 are as follows:
- SARS BJ01, partial genome;
- SARS BJ02, partial genome;
- SARS BJ03, partial genome;
- SARS BJ04, partial genome;
- SARS CUHK-W1, complete genome;
- SARS GZ01, partial genome;
- SARS HKU-39849, complete genome;
- SARS TOR2, complete genome;
- SARS Urbani, complete genome;
- SARS coronavirus CUHK-Su10, complete genome;
- SARS coronavirus isolate SIN2774 complete genome;
- SARS coronavirus TW1, complete genome;
- SARS coronavirus, complete genome.

The HIV strains in Fig. 1(A) are as follows:
- 1, HIV-1 strain CNGL179 from China;
- 2, Human immunodeficiency virus 1;
- 3, HIV-1 isolate BK132 from Thailand;
- 4, HIV-1 isolate BK132 from Thailand;
- 5, HIV-1 strain CNGL179 from China;
- 6, Human immunodeficiency virus 1.
4, HIV-1 isolate CNHN24 subtype B-Thai from China;
5, HIV-1 isolate 011NS565.14 from India;
6, HIV-1 isolate 96GH2911;
7, HIV-1 isolate 95SN7808 from Senegal;
8, HIV-1 isolate BZ163 from Brazil;
9, HIV-1 strain 98IS002 from Israel;
10, HIV-1 strain TV012 clone 2-1_4 subtype C from South Africa;
11, HIV-1 patient WCIPR sample 1990 clone 32;
12, HIV-1;
13, HIV-1 isolate X397 from Spain;
14, HIV-1 strain 98ZA528 clone 6 from South Africa;
15, HIV-1 isolate US4 from USA;
16, HIV-1 isolate 00BW3891.6 from Botswana.

Basic principle of Wolfram approach and the rule selected in the study

DNA sequence is currently thought to be one dimensional linear structure coded in two states: 0 and 1 (black and white seen in images), each cell in the coding sequence has immediate interaction with its neighbouring cells at left and right, respectively, in terms of “positive” power and “negative” power. The powers from either side, which are not always equal, can be scored between 0 and 1 representing for or against each other. This is the mode of power relation of one dimension sequence [Fig. 2(A)]. In this mode, only eight arrangements can be found [15] [Fig. 2(B)]. The cells in the first row and in the second row are paternal and generated, respectively [Fig. 2(C)].

If the cell and its two neighbours are all in state 0, which can be seen on the right of Fig. 2, the cell interacted with the same kind of powers from both sides keeps the original state 0 in the next step. In 001, the middle cell accepting unequal powers, which are opposite each other from both sides, still maintains the final state 0 in the next step with the two powers overlapped.

However, in 010, the powers which belong to the same kind of powers from both sides are greater than the power of the cell itself, resulting in the state 1 shifted to state 0 in the next step. By analogy, 011 will be changed to state 1, 101 to state 1, 111 remains the same state 1. However, in 100 and 110, of which both have neighbours 1 on the left and neighbour 0 on the right, the final state in the next step is the opposite of the initial state. All the eight arrangements result in very complicated behaviours of DNA sequence with two kinds of lines and the nested structure. In theory, still 255 more rules exist but the present rule provides the best solution.

Computer and image

The computer Origin 3000 (Silicon Graphics, Inc. 64 500 MHZ IP35 processors) was used throughout this study. Each sequence was run on the same simple programs to see how they behave. Eventually, one of the rules was selected for comparative study based on more than 3,000 images (200,000 Mega) showing behaviours of complexity.

Results and Discussion

Characteristics of SARS viruses

In the study of sequence homology [21–24], we compared 62 viruses including 13 SARS viruses, 16 HIV and some related viruses on the evolutionary tree, which was constructed by comparing the sequence homology between any two viruses [24–28] using Clustal (Fig. 1). Bovine coronavirus and avian infectious bronchitis virus, as SARS-related viruses, showed higher score. Previous studies [21–28] reported that SARS virus was a new kind of coronavirus similar to bovine coronavirus and avian infectious bronchitis virus. However, our work indicated that they had remarkable intrinsic differences.

All the 13 SARS viruses studied behaved quite differently (see Fig. 3). There was a very large nested structure across the beginning 10 kb regions that was rarely seen in all other studied viruses except those with higher homology. Furthermore, four smaller nested structures (approx. 2 kb long on average) were found within approx. 12–25 kb in SARS viruses but not in most other studied viruses (Fig. 4). Therefore, it is rational to think that the nested structure is the typical feature of SARS viruses containing some special bio-information absent in other viruses, which may be involved in the development of SARS virus by producing replicase or viral protein in this region as suggested previously [21] (Fig. 5).

Pair wise comparison in the traditional way also exhibited similarity of sequence among porcine transmissible gastroen, human coronavirus 229E, and SARS viruses but...
no fundamental similarity was found in equine rhinovirus (Fig. 1). However, the whole genome of equine rhinovirus (ER) had similar regulating feature of SARS virus (Fig. 4), although ER genome contained only 7734 bp as compared to more than 29 kb of SARS genome [21–24]. Using traditional method, SARS virus was found to have high homology with porcine transmissible gastroen and human coronavirus 229E, but we found that SARS virus and equine rhinovirus had similar intrinsic image and specific nested structure (Fig. 4). ER genome is much smaller than SARS virus genome, so it is unlikely that ER was mutated directly to SARS virus but we could not exclude the possibility that ER is the ancestor of SARS virus. At the same time, SARS virus and human coronavirus 229E are very different in behaviour. Another interesting finding is that the beginning genome sequences after the first “atgcce” of other coronaviruses have very similar behaviour. Furthermore, the behaviours of the last 2 kb region of each species are also similar.

Possible origin of SARS virus

SARS-GZ was likely to be formed on the basis of SARS-CUHK because its 33 bp in the beginning region, about 300 bp in the ending region [including 24 poly (A)] and 520 bp between 21,300 and 21,820 bp do not exist in SARS-GZ. However, seven pieces of 7–13 bp showing 100% homology, and dozen pieces of 10–30 bp showing 70%–99% homology, certain sequences in the 520 bp in SARS-CUHK virus DNA were found to be inserted into
the regions within 3.7–4.2 kb or 26.5–26.7 kb, and at least in 6 other positions in SARS-GZ genome (Fig. 6). Thus,
Fig. 5  The nested structure and both Replicases 1A and 1B of SARS viruses found in the same region.

By comparison, we found the nested structure are mainly located in the beginning 21.5 region of SARS virus sequence, which is just in the position for both Replicases Fig. 1(A) and (B), an important “weapon” for SARS viruses to maintain their amounts necessarily for causing the disease.

Fig. 6  Comparisons between SARS-CUHK and SARS-GZ

In the upper three images, the green colour stands for the entire genome of SARS-CUHK and the red colour for the genome of SARS-GZ. (A) In SARS-GZ (red) there is a 520 bp of deletion indicated by the gap but other parts matched well. (B) and (C), In the second and third images with partial magnification show small pieces of DNA inserted into at least five positions of SARS-GZ (red). (D) The tiny difference at base level can be detected from the magnified original images, the one in green stands for SARS-CUHK.
the first case of SARS reported in Guang Dong province was probably caused by SARS virus originated in Hong Kong. The evolutionary position of SARS-CUHK was found to be more stable than that of SARS-GZ, supporting the view that SARS-CUHK appeared earlier than SARS-GZ. Furthermore, the eight SARS viruses other than SARS-GZ, -BJ03, -BJ04, -SIN2774, and -Coronavirus are more stable. The three strains from Hong Kong belong to three different subgroups, this fact also supports the above view (Fig. 1).

Advantages in the study of SARS with Wolfram approach-based method

The array in Fig. 7 is only consisted of two elements, white and black cells (the smallest unit in coding sequence), representing power and counter-power. The first (original) line composed of the cells in the array refers to the coding sequence of DNA. Under the same rule, every cell changes its state in each following step. If one cell is replaced with another cell, for example, after a white cell turns to a black cell, an “earthquake” in the following steps can take place. In the example of \( p_{53} \) (Fig. 8), only one base change (\( A \rightarrow T \) at 186 bp position) results in the continuing abnormal expression of genetic information (white lines) which should be terminated.

On the other hand, other genetic information, which should not be expressed in the late stage of life development, is continually expressed. Hence, despite that high similarity of sequences in a certain region between two species can sometimes be found, it does not necessarily mean that their final behaviours should be the same. Instead, they can be greatly different. This may be explained by the fact that the similar sequence is also regulated by the complex network of power interactions among the whole genome. On the contrary, sequences obtained in the traditional way show less similarity but the distribution of interaction powers of bases in the network of power interactions among the whole genome can be very similar in behaviours, such as the process of pathogenesis between the two species. From this work, we think that SARS viruses are likely to catch a part of genetic information from other non-coronaviruses to increase their special power of causing the disease as a result of the partial similarities of behaviours between SARS and other non-coronaviruses like HIV.

In the images of power (Fig. 7), the thick and thin lines have been divided into two types, the white one inclining towards left and the black one towards right with 45 degree angle. If the equal numbers of white and black lines meet each other, it will result in an interim balanced state.

![The power and counter-power shown with white cell or white line and black cell or black line, respectively](image)

(A) Interaction between “positive” and “negative” powers. It needs a rule installed for gaining a final balanced power. (B) The power in black beats the power in white, which is partially driven away. (C) and (D), The biological system is in a balanced state under the circumstance the power in white does not exist any more. (E) While interaction between the power in white and power in black is equivalent in local area, a “completely balanced state” has been reached but can only be formed in partial area in genetic material. (F) The crossover of “positive” and “negative” powers at genome level no longer exists after the state balanced to finally direct the trend of development of organism.
Homo sapiens tumor suppressor protein p53 (p53) gene analyzed with Wolfram approach-based method. The pairwise images with same size stand for sequence of p53 gene. A1, The normal one, in which the red arrow, marked with 1, indicates white line toward right representing the expressed genetic information; marked with 2, indicates the black line toward left representing the genetic information regulated at the late phase of cell development. A2, When a base A (00) at 186 bp position is altered into T (11), the white line with 1 disappears but a black line with 3 appears. The black line marked with 2 is out of control of cell development at the late phase. Note: (B1) and (B2) are of local magnification of (A1) and (A2).

between the two powers (Fig. 7). However, in most circumstances the two lines eventually disappear because of the continence of the powers. While the number of black lines is much more than the number of white lines, the power represented by black lines will remain because of its dominant role. The remaining power, which is approximately equal to the power after the white line has been deducted from the black line, gets weaker, and vice versa. Nevertheless, the white line or the black line will not always go on if the counter-powers meet again in the following step. Both lines will disappear to reach a balanced state if they are equal. Otherwise, the line with more power will dominate the situation with the deducted power. Through their “encounter” again and again, at last only one type of line will exist with a final stably balanced state. Based on the interaction of power and counter-power, the “encounter” between the powers in white towards the left and the power in black towards the right in several locations may finally result in the formation of the nested structure, which is presumed to be the important state of life filled with complex regulation network.

In analysis of the whole genome behaviour with 10,000 fold reduction of the original size, we found that behaviours of all studied organisms can be classified into three categories: the first one mainly composed of white lines inclining towards left in 45 degree as most common viruses show; the second one mainly composed of black lines towards right in 45 degree as HIV exhibits; and the last one composed of the typical nested structure as SARS viruses possess. Interestingly, in this study the three kinds of viruses happen to have these three types of behaviours (Fig. 4), suggesting that the behaviours of HIV and SARS are specific and rare. The typical feature of Wolfram approach is the conversion of “A” “G” “C” “T” into two states, white and black, in accordance with the characteristics that opposite sides coexist in all substances and also with the theory of “Yin-Yang” balance (Fig. 7).

All genetic materials are made up of thousands and hundreds of bases of AGCT or AGCU, but there must be an inner uncertain mode of base organization that each species follows in the utilization of these four bases, so different viruses show different behaviours even under the same rule. Though sometimes the result obtained by both the traditional and the new methods are similar, one of the important differences between them is due to the fact that the judgment with the traditional way always relies on homology or the variation of a single base, etc., which we think is not complete or objective. Instead, Wolfram approach pays attention to the interaction power of regulation between any adjacent bases, reflecting a major and basic power determining the development and evolution of organism along its own specific track. In fact, the interaction power between bases can be divided into two categories: promotion and continence with each other (Fig.
which can be much more effective than single base alterations in the development of organisms. In addition, each base receives powers from neighbouring bases and exerts power to neighbouring bases at the same time, indicating that the role of each base in a sequence is variable. When all powers from different bases are gathered together, a network of huge power interactions with overlap of powers will be formed, but the balance will be easily broken by an alteration of sequence occurs because of its precision and fragility. As we know, sequence alteration sometimes can be ignored, but sometimes can not because the original base may play a key role in the behaviour of the whole genome. This also explains why some mutations do not affect life too much but sometimes even a single base mutation can cause death, which we call “key (vital) base unbalance” (Fig. 8). The Wolfram approach is just a good way to answer this sort of questions.

Ignoring the power interactions between bases in the traditional analysis usually causes unsatisfactory explanation or misleads the direction of research, especially in the study of some complex phenomenon of lives. At present, researchers are trying hard to produce drugs against SARS. For example, human rhinovirus 3 Cpro inhibitors have been considered for SARS therapy [29]. However, equine rhinovirus (ER) has more similar regulating characteristics of SARS as shown in the analysis by Wolfram approach-based method (Fig. 4). Therefore, ER-based inhibitor could be better than human rhinovirus-based inhibitor for fighting SARS.

In analysis with Wolfram approach, SARS virus shows an image of much more outstanding and complicated behaviours under one of 256 rules (Fig. 8). It suggests that SARS should have a specific bio-information with its own mode of base organization, which is not necessary to be the same as HIV that can follow another type of mode of base organization (Fig. 4). According to the current thinking, however, HIV, which is still a kind of virus to cause huge troubles in man, should have a large number of similar complex information in the image but, instead, it looks more special than what we have expected. This may be also a reason why people are facing big difficulties to answer many scientific questions using the traditional way, which people have noted is actually limited in use.

Our result is different from those reported recently in other papers [30–31] which believe SARS is one species of coronavirus genus, but our image results shows that SARS could be a new kind of virus, which may be far away form other coronaviruses in family-relation distance. This can be due to the fact that traditional sequence analysis ignores the difference between any two bases. In contrary, Wolfram approach is sensitive even to the spot mutation of a single base, and our result shows the inner characteristics and future trend of the whole genome. As a result, the similarities between SARS viruses and other coronaviruses are not really higher than those between SARS viruses and other non-coronaviruses. Even in the result obtained by Guan et al.[32], the difference between the sequences of SARS viruses and coronaviruses from Masked palm civet can be found. Therefore, more work remains to be done to clarify the cause in the future.

Briefly, we are the first one to apply Wolfram approach in the analysis of SARS virus at the molecular and sequence level with many advantages, particular in the magnification of tiny changes in DNA sequence for both detailed and overall analysis in the whole genome scale. If the analysis is only based on the comparison of sequences in the traditional way, the scope of research will be limited. In this work, we studied and discussed the origin of SARS with Wolfram approach but the etiology of SARS investigated in this study may need more work to further confirm.

References

1 Stephenson J. Global Impact of SARS. JAMA, 2003, 289: 2349
2 Mackay B. SARS: “A domino effect through entire system”. CMAJ, 2003, 168: 1308
3 Chandler C. SARS attacks, China shudders. Fortune, 2003, 147: 32
4 Lemennick MD, Park A. The truth about SARS. Time, 2003, 161: 48–53
5 Shure N. SARS hit home. US News World Rep. 2003, 134: 38–42
6 Kondo W. Canadians still stung by WHO’s SARS travel advisory. Lancet, 2003, 361: 1624
7 Shohridge KF. SARS exposed, pandemic influenza lurks. Lancet, 2003, 361: 1649
8 Wenzel RP, Edmond MB. Managing SARS amidst uncertainty. N Engl J Med, 2003, 348: 1947–1948
9 Parry J. Containment of SARS depends on how it is handled in China. BMJ, 2003, 326: 1004
10 Parry J. WHO warns that death rate from SARS could reach 10%. BMJ, 2003, 326: 999
11 Cameron PA, Rainer TH, de Villiers Smit P. The SARS epidemic: Lessons for Australia. Med J Aust, 2003, 178: 478–479
12 Clark J. Fear of SARS thwarts medical education in Toronto. BMJ, 2003, 326: 784
13 Gerberding JL. Faster... but fast enough? Responding to the epidemic of severe acute respiratory syndrome. N Engl J Med, 2003, 348: 2030–2031
14 Cyranoski D. Taiwan left isolated in fight against SARS. Nature, 2003, 422: 652
15 Wolfram S. A new kind of science (Wolfram Media, Inc., Champaign, 2001)
16 Giles J. What kind of science is this? Nature, 2002, 417: 216–218
17 Casti JL. Science is a computer program. Nature, 2002, 417: 381–382
18 Hayes B. The world according to Wolfram. American Scientist, 2002, 90: 308–312
19 Mitchell M. Is the universe a universal Computer? Science, 2002, published online, http://www.sciencemag.org/cgi/content/full/298/5591/65
20 Lemonick ND. How everything works. Time, 2002, 159(20): 67, published online, http://www.ph.sophia.ac.jp/~boes-ken/E-Wolfram-Automata.pdf
21 Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YS, Khattra J et al. The genome sequence of the SARS-associated coronavirus. Science, 2003, published online, http://www.sciencemag.org/cgi/rapidpdf/1085953v1.pdf

22 London OD. Two strains of the SARS virus sequenced. BMJ, 2003, 326: 999

23 Enserink M, Vogel G. Infectious diseases. Hungry for details, scientists zoom in on SARS genomes. Science, 2003, 300: 715–717

24 2003, Published online, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=nucleotide&cmd=search&term=SARS

25 Wang Y, Ma WL, Song YB, Xiao WW, Zhang B, Huang H, Wang HM et al. Gene sequence analysis of SARS-associated coronavirus by nested RT-PCR. Di Yi Jun Yi Da Xue Xue Bao, 2003, 23: 421–423

26 Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, Peñaranda S et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science, 2003. published online, http://www.sciencemag.org/cgi/rapidpdf/1085952v1.pdf

27 Yang J, Wang ZH, Chen JJ, Hou JL. Clinical detection of polymerase gene of SARS-associated coronavirus. Di Yi Jun Yi Da Xue Xue Bao, 2003, 23: 424–427

28 Holmes KV. SARS-associated coronavirus. N Engl J Med, 2003, 348: 1948–1951

29 Anand K, Ziebuhr J, Wadhwani P, Mesters JR, Hilgenfeld R. Coronavirus main protease (3CLpro) structure: Basis for design of anti-SARS drugs. Science, 2003, published online, http://www.sciencemag.org/cgi/rapidpdf/1085658v1.pdf

30 Qi Z, Hu Y, Li W, Chen YJ, Zhang ZH, Sun SW, Lu HC et al. Phylogeny of SARS-CoV as inferred from complete genome comparison. Chinese Science Bulletin, 2003, 12: 1175–1179

31 Gao L, Qi J, Wei HB, Sun YG, Hao BL. Molecular phylogeny of coronavirus-es including human SARS-CoV. Chinese Science Bulletin, 2003, 12: 1170–1175

32 Guan Y, Zhang BJ, He Y, Lin XL, Zhang ZX, Cheung CL, Luo SW et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in Southern China. Science, 2003, 302: 276–278