Enhancement of *Coronaviridae* Pathotyping Using Length Polymorphism

Sheng-Lung Peng* and Yu-Wei Tsay
Department of Computer Science and Information Engineering,
National Dong Hwa University, Hualien 974, Taiwan
Email:slpeng@mail.ndhu.edu.tw

***Abstract*—*Coronaviridae* is a class of enveloped, single stranded, and positive-sense RNA viruses including genera Coronavirus and Torovirus. During the severe acute respiratory syndrome (SARS-CoV) outbreak in Asia, virus identification becomes a leading issue in recent years. Studies of SARS-CoV were diagnosed by various molecular techniques. Only a few tools are included to help the identification process. One of the detections is PCR-based approach, for the detection and identification of DNA strains – length polymorphism. Length polymorphism indicates variations in the length of sequence fragment, resulting from various genetic mutation, and used as a DNA marker to detect genetic diversity. In this paper, we formalize the problem of cleaved amplified polymorphic sequences (CAPS) subtyping issue. We propose a simple heuristic algorithm to solve the problem of CAPS subtyping. Besides, the necessary condition for optimality CAPS subtyping problem is also refined. First, fragment lengths after digestion on genome are improved from boolean vectors to integer vectors (CI-Vectors). This helps a comprehensive quantification in length polymorphism for pathotyping. Second, a simulated histogram is proposed to demonstrate the variation in DNA length polymorphism, describing a simplified view in CI-Vectors. Results showed that when genomic DNA of *Coronaviridae* strains were digested with enzymes AatII, BbeI, and SplI, most of the SARS-CoV strains can be isolated from other CoV, including bovine torovirus, equine torovirus, and human torovirus. In addition, the proposed method is applied to major structural protein of SARS-CoV in spike protein and envelope protein. Comparing to general subtyping approach, our CAPS patterns provide a promotive advantage on its capability. Each generated enzyme set is also evaluated to ensure their validation. This PCR-based approach offers a practical direction for DNA subtyping.**

***Index Terms*—CAPS, *Coronaviridae*, pathotyping, length polymorphism, set cover problem.**

I. Introduction

The cleaved amplified polymorphic sequences (CAPS) technique [8], known as PCR-RFLP markers, provides a way to utilize the DNA sequences of mapped RFLP marker and eliminating the tedious blotting. The CAPS is performed by digesting locus-specific PCR amplification with restriction enzymes, and then separating the digested DNA on agarose or polyacrylamide gels [15]. The primer pairs are synthesized depended on the sequence information available in genomic database or cDNA libraries.

Primer design is a common way in recent studies. However, it does not prove that a well-designed primer pair gives an appropriate result of digestion. Due to the bias of genotype frequencies, population heterozygosities, degree of inbreeding, population subdivision, and degree of individual relatedness, it is often challenged for the validity of CAPS markers. Therefore, the proposed method will apply to viral strains, containing potentially homozygous, heterozygous, and various genetic variations. Given that each strain has been amplified by PCR procedure completely, afterwards, a comprehensive condition of CAPS marker selection in silicon simulation will be described.

Over the past few years, a considerable number of studies have been made on CoV pathotyping, especially on SARS-CoV. Despite the early development of diagnostic tests, further progress of laboratory tests for SARS has been slower than originally expected [6]. Development of virus diagnostic tests, antiviral strategies, and vaccines for each of these pathogens is crucial to limit their impact [3]. Thus, fast and accurate diagnostic tests are needed to separate CoV. Remarkably, the Bernhard Nocht Institute developed the first diagnostic test for SARS-CoV just 11 days after WHO issued its alert about the disease [1]. They tried to take out the elusive virus with a less-specific series of PCR reactions that

can amplify the genetic material of a wide range of viruses. Despite of high sensitivity, the existing PCR tests cannot certainly rule out the presence of the SARS virus in patients [14]. On the other hand, contamination of samples in laboratories might lead to false positive results. Stringent guidelines on laboratory quality control and confirmatory testing have been issued by the WHO [23]. Moreover, some studies argue to develop and evaluate a sensitive diagnostic test for SARS-CoV by applying real-time quantitative RT-PCR technology [18]. Currently, efforts are underway to improve the sensitivity of PCR assays to increase their clinical usefulness.

In this paper, we consider the selection of restriction enzymes for the problem of CAPS subtyping [16]. The *CAPS subtyping problem* is to find an enzyme set of minimum size.

A. Materials

Due to the diversity of distinct organisms, one family may consist of several types of polymorphism, especially those contain properties of pathogenic and non-pathogenic.

Coronaviridae is a class of enveloped strand of linear positive-sense RNA viruses including generas of coronavirus and torovirus. Properties of coronavirus are 27-33 kb, characterized by enveloped, and peplomer-bearing particles. It contains the following structural protein: peplomer glycoprotein in S, transmembrane glycoprotein in M, nucleocapsid phosphoprotein in N, and some viruses have peplomers with hernagglutinin plus acetylesterase activity in HE [24]. Torovirus is a genus of viruses containing single-stranded, positive-sense RNA [12]. It has been affirmed its enteric infections in horses (Berne virus), cattle (Breda virus), swine, and humans. The virions have a typical pleomorphic morphology and virion organization is characterized as the Coronaviridae family [5]. Besides, coronaviruses are gathered into three groups – Groups I and Groups II contain mammalian viruses, and Group III contains only avian viruses [22]. Appearing of a solar corona demonstrates the reason why it is denominated as coronavirus.

For the present, it may be useful to observe at some of the more important features of SARS-CoV. During the recent SARS (severe acute respiratory syndrome) outbreak, there is an evidence showing that the etiologic agent of SARS is a new coronavirus [19]. Further, the bat-SARS-CoV may be classified into an additional group of SARS-CoV [9]. The molecular and biological characteristics of SARS-CoV strains were demonstrated in recent studies. Most of studies focus on the classification of SARS-CoV sequences among Homo sapiens and other animals [22]. There is a considerable evidence revealing that the coronaviruses have a history of host-shifting [10], [21]. In the majority of its genomic regions, SARS-CoV is closely related to one of the groups among these regions [11]. Moreover, there are still some studies underway to evaluate whether specific parameters that may distinguish SARS-CoV from others [20].

The number of coronaviruses we consider is 299, including 130 SARS-CoV and 169 coronaviruses. These 169 coronaviruses are classified into three groups. Group I contains 44 coronaviruses including Human coronavirus 229E (HCoV-229E), Human coronavirus NL63 (HCoV-NL63), Canine coronavirus (Canine-BGF10), and Transmissible gastroenteritis virus (TGEV), Feline coronavirus (FIPV), Porcine epidemic diarrhea virus (PEDV), New Haven coronavirus (NL63 or NL), and so on. Group II contains 78 coronaviruses. They contain Human coronavirus OC43 (HCoV-OC43), Bovine coronavirus (Bovine-CoV), Murine hepatitis virus (MHV), Porcine hemagglutinating encephalomyelitis virus (HEV), and so on. Group III is classified mainly for Turkey coronavirus (Turkey-CoV), Avian infectious bronchitis virus (Avian-CoV), and infectious bronchitis virus (IBV), and so on. The Torovirus contains Breda virus, Berne virus, and Porcine torovirus (strain P10).

Before going to the next section, the concept of restriction endonuclease should be notified. Restriction enzymes are classified biochemically into various types. While recognition sequences vary widely, many of them are palindromic. It means the sequence on one strand reads the same in the opposite direction on the complementary strand. Type I and Type III systems, both the methylase and restriction activities are carried out by a single large enzyme complex. Although these enzymes recognize specific DNA sequences, the sites of actual cleavage are at variable distances from these recognition sites. The vast majority of known re-

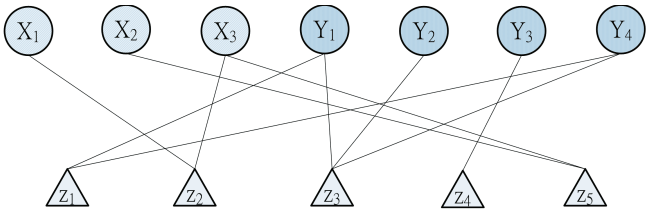


Fig. 1. An illustration of CAPS subtyping problem.

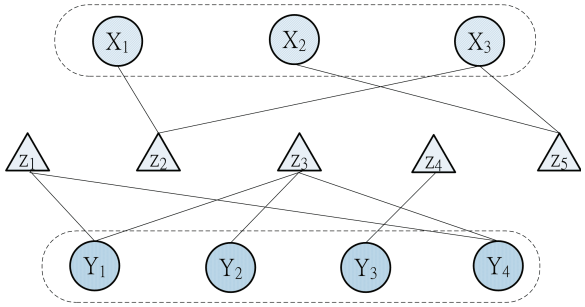


Fig. 2. Two possible sets for the CAPS subtyping problem with the input shown in Fig.1.

striction enzymes are type II, and it finds the most usages as laboratory tools.

II. Methods

With proper restriction enzyme chosen in an incorporating CAPS digestion, a specific selection method in CAPS may offer a more accurate and rapid identification scheme than conventional one [13]. In this section, we propose an algorithm for finding a particular enzyme set to differentiate two populations. Our problem and method can be stated in the following subsection.

A. Problem formulation

Let $G = \{g_1, g_2, \ldots, g_n\}$ be a set of considered genomes and $Z = \{z_1, z_2, \ldots, z_k\}$ be a set of possible enzymes. Let $\{X, Y\}$ be a partition of G, *i.e.*, $X \cup Y = G$ and $X \cap Y = \emptyset$. For convenience, we use $G = X + Y$ to denote that $\{X, Y\}$ is a partition of G. Given the collection of genome $G = X + Y$ and enzyme set Z, the *CAPS subtyping problem* is to find a minimum subset $U \subseteq Z$ such that X can be digested by U but Y cannot, or *vice versa*.

The CAPS subtyping problem can be treated as a graph problem shown in Fig.1. Let $X = \{x_1, x_2, x_3\}$ and $Y = \{y_1, y_2, y_3, y_4\}$. Let enzyme set $Z = \{z_1, z_2, z_3, z_4, z_5\}$. If z_i can digest x_j (y_j), then there is an edge between z_i and x_j (y_j). It is easy to check that enzyme set $U = \{z_2, z_5\}$ can digest X but cannot digest Y. We say that U is a covering set of X. The difference of covering set between X and Y causes various digestion in CAPS procedure. The *CAPS subtyping problem* is to find an enzyme set of minimum size that only digests (covers) X or Y. Fig.2 shows two possible solutions for $X + Y$. It is easy to see that $\{z_2, z_5\}$ is the one with minimum size. By letting $Y = \emptyset$, the problem becomes the set cover problem where X is a set, each z_i is a subset of X, and Z is a set of subsets of X. Thus the set cover problem is a special case of the CAPS subtyping problem. Therefore, it is not hard to show that the *CAPS subtyping problem* is NP-hard.

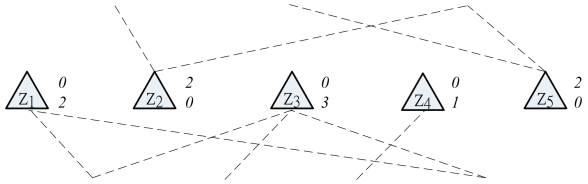


Fig. 3. An example of CAPS subtyping problem using heuristic approach.

Intuitively, there are two stages for the CAPS subtyping problem. The first stage is to find a cover set U_x for X from Z. The second one is to find a cover set U_y for Y from Z. Then we select the minimum one (U_x or U_y) as the solution. Since the set cover problem is NP-hard, it does not be expected to have a polynomial-time algorithm for it [4], [2]. Moreover, it is possible that no such set exists. Thus we have to find a set with minimum side effect, *e.g.*, a cover set for X but only digests elements of Y as few as possible, or *vice versa*.

B. An Approximation Algorithm

Each enzyme $z_i \in Z$ estimates its degrees in X and Y. Let a_i (respectively, b_i) be the degree of z_i in X (respectively, Y). Let N_X (respectively, N_Y) be the degree array of Z in X (respectively, Y). As the example in Fig.3, $N_X = [0, 2, 0, 0, 2]$ and $N_Y = [2, 0, 3, 1, 0]$. Consequently, the difference between N_X and N_Y can be used in various scoring approach, *e.g.*, Euclidean distance, one of the most common uses of distance. We simply use $N_X - N_Y$ to express their disparity. For example, let D be

Algorithm 1 An Approximation Algorithm

Input: two genome sets X and Y, an enzyme set Z with its corresponding degree arrays N_X and N_Y.
Output: an enzyme set U
$C = X$, $D = N_X - N_Y$, $U = \emptyset$;
for $k = 1$ to 2 **do**
  while C is not covered **do**
    Let j be the index with the max value and unselected in D;
    $U = U \cup \{z_j\}$;
    mark $D[j]$ as selected element;
    update z_j's neighbors as covered;
  end while
  $U_1 = U$, $D = N_Y - N_X$, $U = \emptyset$, $C = Y$;
end for
if $|U_1| < |U|$ **then**
  $U = U_1$
end if
return U

the difference of N_X and N_Y. In Fig.3, $D_{XY} = [-2, 2, -3, -1, 2]$ and $D_{YX} = [2, -2, 3, 1, -2]$. By using the difference arrays, we propose a greedy algorithm to find an enzyme set U from Z to differentiate X and Y. **Algorithm 1** shows the detail.

C. CAPS Rating Criteria

In the *CAPS subtyping problem*, it is a time-consuming work to determine a proper enzyme set, even running gel electrophoresis with naked eye observing. Each gel map is produced according to the lengths of fragments after digested by an enzyme set. Hence, a quantitative criterion is needed to determine which enzyme set is applicable for simplified gel maps. In brief, we need to normalize each gel map and check their similarity by given enzyme set. The rating criterion is given as follows.

Let V be a set of CAPS fragment lengths on some genome g generated by enzyme set U, and *CAPS Integer Pattern Vector of V with respect to U* (abbreviated as *CI-Vector*(U, V)) be a boolean vector of CAPS pattern V by enzyme set U. And L be the degree of fineness according to CAPS fragments. From the point of view of macrography, CAPS gel patterns must be centralized at some specific areas, or simply distributed at few locations. As the fineness L goes smaller, the size of CAPS pattern vector may raise. For example, suppose a full-length DNA virus contains 10k bp. After PCR amplified and restriction procedure, digested patterns are put into zone electrophoresis, ranging from 0 to 10k bp. Let $L = 0.2$, the scale of *CI-Vector* is allowed to start at 0 to 2k, 2k to 4k, 4k to 6k, 6k to 8k, and 8k to 10k. Then the obtained fragment are placed into each *CI-Vector* tag orderly. If strain g can be digested into fragment set V with fragment lengths of 0.3k, 0.7k, 2k, 3k, and 4k by U, then *CI-Vector*$(U, V) = [2, 3, 0, 0, 0]$. In this manner, the distance between two CAPS patterns can be measured by some distance function, *e.g.*, Euclidean distance, one of the most common used distance function. The shorter distance of *CI-Vector*, the similar gel maps they might have.

In respect of fineness L, the default value is set to 0.15% based on input genome size. That is, if length of genome g is 10,000 bp, the scale of *CI-Vector* may regulates to 15 bp. Besides, parameters of CAPS enzyme selection can be adjusted by demand, including numbers of genome sets, category of enzyme set, agarose gel, polyacrylamide gel, or pulsed field gel electrophoresis (PFGE) simulation. Typically, PCR amplification of DNA sequence involves 20-30 cycles, these repeated cycles are usually done automatically. We give a simulated environment of PCR procedure, ranking each PCR amplified segment in 2^{10} to 2^{12} cycle.

Suppose X and Y are two relative species requiring to quantify genetic relationships, Π_X and Π_Y measure the nucleotide diversity among several sequences in a given region of genome within a population, and V_{XY} measures population divergence based on the degree of sequence variation [7], [17]. Eventually, the measure of genetic diversity gives a criterion of our proposed rule in enzyme selection.

III. Results

There are three groups of coronavirus (CoV): Groups I and Groups II contain mammalian viruses, whereas Group III contains only avian viruses. Within each group, the CoV are classified into distinct strain by host ranges. Therefore, our purpose is to find a minimal set of CAPS enzymes that can identify SARS-CoV from other CoV.

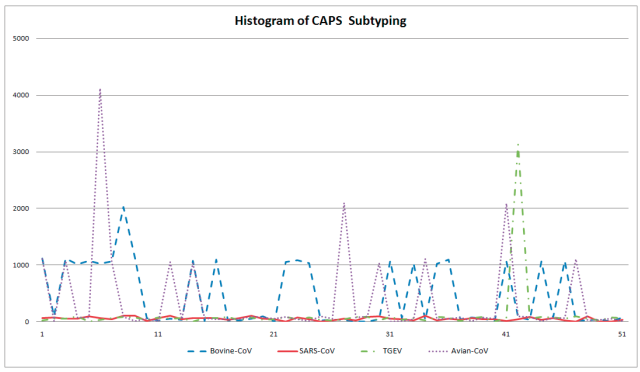
TABLE I
COV SUBTYPING BYC VARIOUS ENZYME SETS

Species	Cat	Num	V^a_{XY}	Π^a_Y	V^b_{XY}	Π^b_Y	V^c_{XY}	Π^c_Y	V^d_{XY}	Π^d_Y
Group I										
Human-CoV	C	5	∞	0.00000	0.25470	0.00139	0.36018	0.01829	0.25548	0.00000
Feline-CoV	C	5	∞	0.10469	0.25723	0.01373	0.30070	0.03950	0.27680	0.03186
Human-CoV 229E	C	2	∞	0.00000	0.27621	0.00000	0.30525	0.00000	0.37131	0.00425
Bat-CoV	C	10	∞	0.15921	0.26975	0.05295	0.33137	0.05582	0.30336	0.04640
Porcine-CoV	C	3	∞	0.00000	0.27444	0.00000	0.32465	0.00552	0.33137	0.00137
TGEV	C	10	∞	0.05255	0.22394	0.01249	0.28525	0.00954	0.28646	0.00108
Ferret enteric	P	3	∞	0.34067	0.43326	0.07806	0.54389	0.17925	0.52137	0.10221
Rocky-Mt. CoV	P	6	∞	0.25925	0.51001	0.25470	0.72258	0.15377	0.64769	0.08886
Group II										
Bovine-CoV	C	17	∞	0.01415	0.24613	0.01324	0.30356	0.00500	0.28525	0.03745
Murine-CoV	C	6	∞	0.10278	0.27652	0.02377	0.29384	0.03161	0.28304	0.00525
Human-CoV OC43	C	5	∞	0.00000	0.25519	0.01021	0.39810	0.00283	0.37152	0.03306
HKU1	C	20	∞	0.06078	0.28555	0.01689	0.28799	0.01488	0.31929	0.00525
HKU4	C	4	∞	0.00000	0.24787	0.00000	0.30634	0.00185	0.34559	0.00136
HKU5	C	5	∞	0.02055	0.25894	0.01115	0.27455	0.00818	0.33038	0.00868
HKU9	C	5	∞	0.15109	0.26260	0.02384	0.32987	0.05837	0.35226	0.00455
Puffinosis-CoV	P	2	∞	0.14842	0.35127	0.05408	0.42358	0.03423	0.45019	0.03510
Rat-CoV	P	10	∞	0.27597	0.39021	0.09176	0.55274	0.09150	0.51109	0.12561
Equine-CoV	P	2	∞	0.21183	0.47835	0.26260	0.55695	0.11054	0.51389	0.02085
Group III										
Avian-CoV	C	20	∞	0.23041	0.27657	0.04376	0.31048	0.07482	0.30073	0.02142
Turkey-CoV	C	20	∞	0.00000	0.25583	0.02601	0.31679	0.04739	0.33064	0.03588
IBV	P	10	∞	0.00000	0.36649	0.05670	0.60240	0.02683	0.64366	0.00000
CoV-SW1	C	2	∞	0.00000	0.31253	0.03163	0.28864	0.01624	0.30807	0.00000
Others										
Bat-CoV	P	20	∞	0.00000	0.20984	0.00573	0.61557	0.09496	0.61293	0.11713
Feline-CoV UU	C	12	∞	0.00000	0.29014	0.01795	0.31093	0.04392	0.33064	0.03588
Bovine-ToV	P	10	∞	0.00000	0.25470	0.00139	0.36018	0.01829	0.68803	0.11539
Equine-ToV	P	2	∞	0.00000	0.39753	0.01840	0.30071	0.00000	0.27355	0.00000
Human-ToV	P	2	∞	0.23407	0.41446	0.06887	0.82420	0.11083	0.59014	0.05724
Porcine-ToV	P	20	∞	0.23768	0.47254	0.00139	0.53561	0.11028	0.63316	0.11392

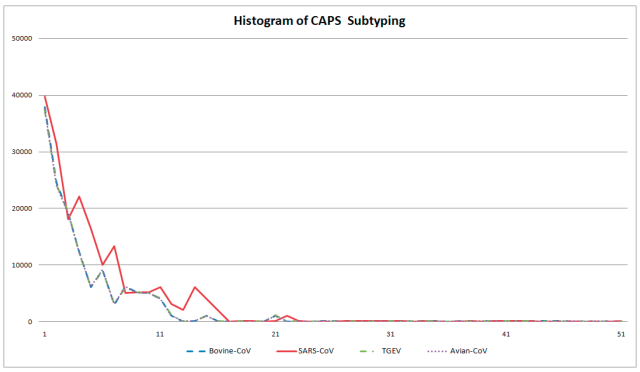
In this table, **Species** indicates each categories of complete genome. **Num** shows numbers of strains in each CoV. V_{XY} measures the diversity between various CoV and SARS-CoV. Π_Y is the intrapopulaton diversity in each CoV by given enzyme set. **Cat** means the category of given CoV, we abbreviates it as **C** if it is a complete genome, as **P** if it is a partial one.

A. Results on Coronaviridae

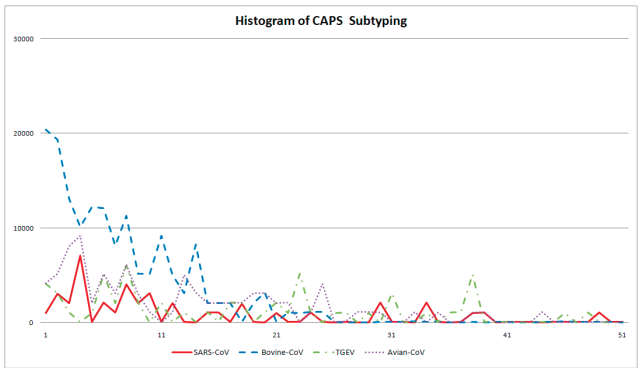
In our experiment results, three of specific enzymes (AatII:GACGTC, BbeI:GGCGCC, SplI:CGTACG) are capable to differentiate SARS-CoV from other CoV. In other word, let $X = \{\text{SARS}\}$ and Y be the set of others CoV. In TABLE I, three enzyme sets demonstrate each population diversity among SARS-CoV (X) and CoV (Y). In the first enzyme set by our proposed method, V^a_{XY} showed a frequent variation between CoV and SARS-CoV. In the equation of diversity measurement, proportion of shared alleles cannot be found, this reason lead to the effect of high divergence among CoV and SARS-CoV based on AatII, BbeI, and SplI. As mentioned to the nucleotide diversity



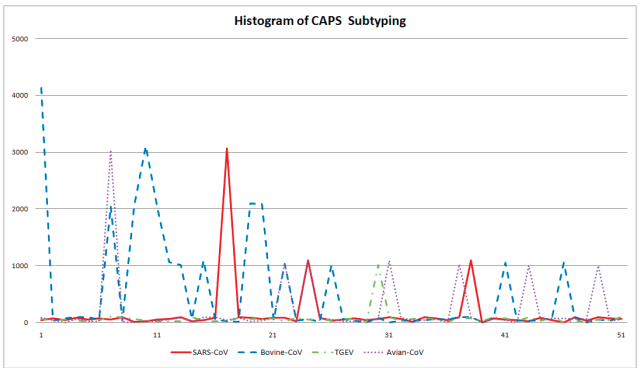
(a) specific enzyme set: AatII, BbeI, and SplI



(b) randomly selected enzyme set: BspEI, MseI, and SpeI



(c) frequently used enzyme set: XbaI, HaeIII, and ClaI



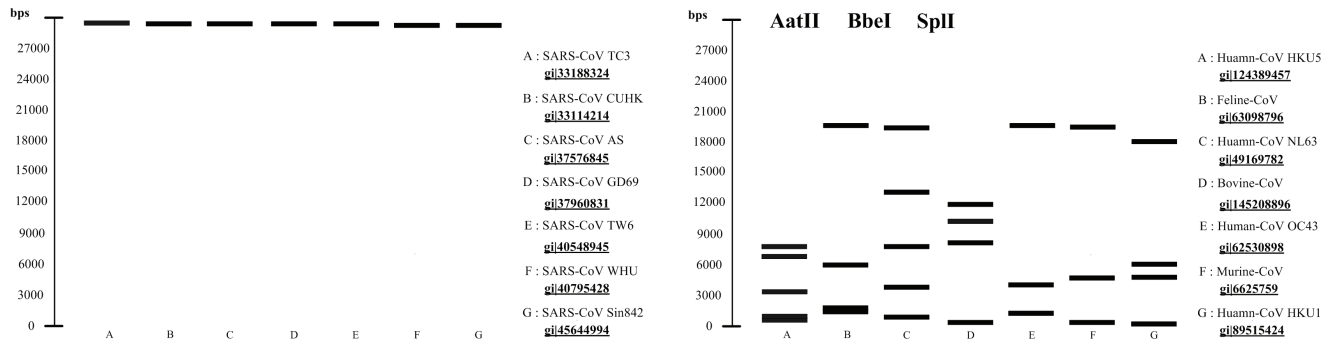
(d) greedy select enzyme set: AgeI, BspDI, MnlI, SalI, BamHI

Fig. 4. The simulated patterns of PCR procedure in length polymorphism.

Π_Y^a , the number of restriction fragments displayed similar results in each CoV, indicating the recognition sites of given restriction enzyme.

For a comparison with other frequent used enzymes, we randomly select another group of restriction enzymes. TABLE I also shows these enzymes fail to identify SARS-CoV from CoV, representing our selected restriction enzymes have high specificity. Obviously, the nucleotide differentiation based on V_{XY}^b (BspEI:TCCGGA, MseI:TTAA, and SpeI:ACTAGT) gives an unacceptable outcome of their nucleotide diversity, even related to intrapopulation sequence diversity Π_Y^b. In V_{XY}^a (AatII, BbeI, and SplI), result of population diversity is incompatible with V_{XY}^b (BspEI, MseI, and SpeI), due to the nucleotide differentiation on variant restriction sites. It should be noticed that the trend of population diversity within each species is similar even digesting with dissimilar enzyme set. For example, restriction diversity Π_Y in Bat-CoV **Group I** with different digestion always retains maximal. In conclusion, the proposed CAPS marker selection gives an appealing issue on CoV subtyping. Moreover, collection of frequently used enzymes also gives a similar result in V_{XY}^c. Enzyme set (XbaI:TCTAGA, HaeIII:GGCC, and ClaI:ATCGAT) also contrasts sharply with our proposed rule. This argument admits of no refutation. It gives a good account in the rule of thumb which is not applicable to enzyme selection. Since the *CAPS subtyping problem* is a hard problem, a habitual greedy approach can be applied to this problem. However, more noteworthy is that such enzyme set may not be easily found in most cases. V_{XY}^d (AgeI:ACCGGT, BspDI:ACTGAT, MnlI:CCTC, SalI:GTCGAT, and BamHI:GGATCC) gives a good evidence for this event.

In view of pathotyping in *Coronaviridae* **Group I**, it seems that CoVs do not be associated with enteric and respiratory diseases of SARS-CoV in human. Fig.4 can be schematized as follows: The histogram indicates the amplified strains of G after digestion. Here, X-axis is the scale of CI-Vector, and Y-axis is the amplified collection of given genomes. As shown in Fig.4(a), the SARS-CoV (red) illus-



(a) specific enzyme set: AatII, BbeI, SplI (b) randomly selected enzyme set: BspEI, MseI, and SpeI

Fig. 5. The simulated gel maps of our proposed method on various CoV.

trates a completely different distribution in its length polymorphism applying our algorithm (V^a_{XY}). The difference between SARS-CoV and others CoV can be easily recognized. As mentioned to the randomly selected enzyme set (V^b_{XY}) or frequently used enzyme set (V^c_{XY}), none of them can perform as ours. In addition, the traditional greedy approach (V^d_{XY}) looks barely recognizable between SARS-CoV and others CoV. Fig.5 shows some polyacrylamide gel simulation of various types of full-length RNA *Coronaviridae* digested by our enzyme set (AatII, BbeI, and SplI). The major difference on gel maps of CAPS patterns is the plenty of mass fragments.

B. Program and Environment

We have implemented a simulation procedure described in previous sections, and testing with *Coronaviridae* sequence datasets. The environment is running under 2Ghz PC with 512MB of main memory with Linux-2.6.11-1.1369. The implementation is temporarily written using Bash-3.00.16(1) and Octave-3.0.0 since the web-based interface of CAPS enzyme selection is under construction by Gnuplot.

IV. Conclusion

By studying several researches on the limitation of PCR-based marker, a method for CAPS enzyme selection is proposed to improve genomic subtyping. In this paper, we formalize the problem of cleaved amplified polymorphic sequences (CAPS) subtyping issue. In addition, it is a generalization of traditional set cover problem. Thus, it is an NP-hard problem. For the purpose of serviceability, this paper develops a simple heuristic algorithm for the problem of CAPS subtyping. The key of CAPS analysis is on the selection of DNA strain, including non-coding region, CDS, cDNA, partial genome, or full-length strains. Due to the SNP or numerous reasons of nucleotide variations, parameters of CAPS enzyme selection can be adjusted by demand, including number of genome sets, category of enzyme set, agarose gel, polyacrylamide gel, or pulsed field gel electrophoresis (PFGE) simulation. In the pathotyping of CoVs, results also give an encouraged validation for proposed method. In the future, the concept of CAPS enzyme selection may be extended to several ways on DNA fingerprint. cDNA-CAPS techniques may also be improved by our method for finding specific and applicable enzyme set. Phylogenetic relation is derived by CAPS features of species. It is expected that our method may be used by biologists in genetics, plant discrimination, forensics and other fields.

References

[1] A Abbott, SARS testing: First past the post, *Nature* **423** (2003) 114–114.

[2] TH Cormen and CE Leiserson, RL Rivest, and C Stein, *Introduction to Algorithms*, 2003, Flying Publisher.

[3] RA Fouchier, GF Rimmelzwaan, T Kuiken, and AD Osterhaus, Newer respiratory virus infections: human metapneumovirus, avian influenza virus, and human coronaviruses, *Curr Opin Infect Dis.* **18** (2005) 141-146.

[4] MR Garey and DS Johnson, *Computers and Intractability; A Guide to the Theory of NP-Completeness*, 1990, W. H. Freeman & Co.

[5] DC Hirsh, NJ MacLachlan and RL Walker, *Veterinary microbiology*, 2004, Blackwell Publishing.

[6] BS Kamps and C Hoffmann, *SARS Reference*, 2003, Flying Publisher.

[7] A Karp, GP Isaac, and DS Ingram, *Molecular Tools for Screening Biodiversity: Plants and Animals*, 1998, Springer.

[8] SA Kharitonov and PJ Barnes, A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers, **Plant J. 4** (1993) 403–410.

[9] S Lau, P Woo, K Li, Y. Huang, HW Tsoi, B Wong, S Wong, SY Leung, KH Chan, KY Yuen, Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats, *Proceedings of the National Academy of Sciences of the United States of America* **102** (2005) 14040–14045.

[10] PH Liu, SL Peng, CY Tang, YW Tsay, and JTL Wang, Classifying SARS-CoV Genomes Based on Their Unique Sequences, *6th International Symposium on Computational Biology and Genome Informatics (CBGI2005)*, 1315–1318, 2005.

[11] G Magiorkinis, E Magiorkinis, D Paraskevis, AM Vandamme, RM Van, V Moulton, and A Hatzakis, Phylogenetic analysis of the full-length SARS-CoV sequences: Evidence for phylogenetic discordance in three genomic regions, *Journal of Medical Virology* **74** (2004) 369–372.

[12] H Marian, F Tom, S Linda, S Willy, W Marianne, and W Gerald, A New Family of Vertebrate Viruses: Toroviridae, *Intervirology* **27** (1987) 17–24.

[13] WL McDonald, BN Fry, and MA Deighton, Identification of Streptococcus spp. causing bovine mastitis by PCR-RFLP of 16S-23S ribosomal DNA, *Vet Microbiol.* **111** (2005) 241-246.

[14] K McIntosh, The SARS coronavirus: rapid diagnostics in the limelight, *Clin Chem.* **49** (2003) 845-846.

[15] SD Michaels and RM Amasino, A robust method for detecting single-nucleotide changes as polymorphic markers by PCR, *Plant J.* **14** (1998) 381-385.

[16] SL Peng, YW Tsay, CS Lin, and CY Tang, CAPS Genomic Subtyping on *Orthomyxoviridae*, to appear in the Proceedings of International Conference on Intelligent Computing, 2009.

[17] SL Peng and YW Tsay, Verification of Pathotyping by Quasispecies Model, to appear in the Proceedings of International Conference on Intelligent Computing, 2009.

[18] L Poon, KH Chan, OK Wong, WC Yam, KY Yuen, Y Guan, D Lo, and J Peiris, Early diagnosis of SARS Coronavirus infection by real time RT-PCR, *Journal of Clinical Virology* **28** (2003) 233–238.

[19] L Poon, Y Guan, JM Nicholls, KY Yuen, and JSM Peiris, The aetiology, origins, and diagnosis of severe acute respiratory syndrome, *The Lancet Infectious Diseases* **4** (2004) 663–671.

[20] SM Poutanen, DE Low, B Henry, S Finkelstein, D Rose, K Green, R Tellier, R Draker, D Adachi, M Ayers, AK Chan, DM Skowronski, I Salit, AE Simor, AS Slutsky, PW Doyle, M Krajden, M Petric, RC Brunham, and AJ McGeer, Identification of Severe Acute Respiratory Syndrome in Canada, *The New England J. of Medicine* **348** (2003) Low-1–Low-11.

[21] JS Rest and DP Mindell, SARS associated coronavirus has a recombinant polymerase and coronaviruses have a history of host-shifting, *Infection, Genetics and Evolution* **3** (2003) 219–225.

[22] PA Rota, MS Oberste, SS Monroe, WA Nix, R Campagnoli, JP Icenogle, S Penaranda, B Bankamp, K Maher, MH Chen, S Tong, A Tamin, L Lowe, M Frace, JL DeRisi, Q Chen, D Wang, DD Erdman, TCT Peret, C Burns, TG Ksiazek, PE Rollin, A Sanchez, S Liffick, B Holloway, J Limor, K McCaustland, M Olsen-Rasmussen, R Fouchier, S Gunther, ADME Osterhaus, C Drosten, MA Pallansch, LJ Anderson, and WJ Bellini, Characterization of a Novel Coronavirus Associated with Severe Acute Respiratory Syndrome, *Science* **300** (2003) 1394–1399.

[23] *Use of laboratory methods for SARS diagnosis (WHO)*, http://www.who.int/csr/sars/labmethods/en/

[24] *Coronaviridae family and its properties [ONLINE]*, http://maxshouse.com/virology/coronaviridae_table_of_contents.htm