

## Evolutionary implications of Avian Infectious Bronchitis Virus (AIBV) analysis

Peng Shi<sup>1,2,\*</sup>, Li Yu<sup>2,\*</sup>, Yun-xin Fu<sup>3</sup>, Jing-Fei Huang<sup>1</sup>, Ke-Qin Zhang<sup>2</sup>, Ya-ping Zhang<sup>1,2</sup>

<sup>1</sup>Laboratory of Cellular and Molecular Evolution, and Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; <sup>2</sup>Laboratory for Conservation and Utilization of Bio-resource, Yunnan University, Kunming 650091, China; <sup>3</sup>Human Genetics Center, University of Texas at Houston, Houston, Texas 77030, USA

For developing efficient vaccines, it is essential to identify which amino acid changes are most important to the survival of the virus. We investigate the amino acid substitution features in the Avian Infectious Bronchitis Virus (AIBV) antigenic domain of a vaccine serotype (DE072) and a virulent viral strain (GA98) to better understand adaptive evolution of AIBV. In addition, the SARS *Coronavirus* (SARS-CoV) was also analyzed in the same way. It is interesting to find that extreme comparability exists between AIBV and SARS in amino acid substitution pattern. It suggests that amino acid changes that result in overall shift of residue charge and polarity should be paid special attention to during the development of vaccines.

*Cell Research* (2006) 16:323-327. doi:10.1038/sj.cr.7310041; published online 16 March 2006

**Keywords:** Avian Infectious Bronchitis Virus, SARS Coronavirus, positive selection, adaptive evolution, vaccine development

### Introduction

Avian infectious bronchitis virus (AIBV), a typical species of the genus *Coronavirus* [1], can induce a highly contagious disease in chickens. The AIBV genome encodes four structural proteins: the envelope protein (E), the spike protein (S), the membrane protein (M) and the nucleocapsid protein (N) [2, 3]. Among them, the S protein is the major envelope glycoprotein of AIBV, which has been believed to associate with attachment [4] and virus-neutralizing antibody induction [5], as the same case in many other *Coronavirus* viruses.

Extensive antigenic and genetic variation is a distinct feature of AIBV. Antigenic variants are generally related to the appearance of positively selected, single point mutations in the antigenic domain of the viral proteins. These mutations lead to the alteration of virulence and the escape of the viruses from host defenses [6]. Thus, even though the vaccines have been extensively used [7], outbreaks of the disease continue. In present study, we are the first to investigate the amino acid substitutions features in the AIBV antigenic domain of a vaccine serotype (DE072) and a virulent viral strain (GA98) to better understand adaptive evolution of AIBV. More importantly, we compared our results from AIBV in a similar analysis of wellknown severe acute respiratory syndrome *Coronavirus* (SARS-CoV), identifying the locations where amino acid changes are likely important to the survival of the virus.

### Materials and Methods

#### Data mining

AIBV S protein sequences of DE072 vaccine and virulent viral strain GA98, which was thought to have arisen as a

\*These authors contributed equally to this work.

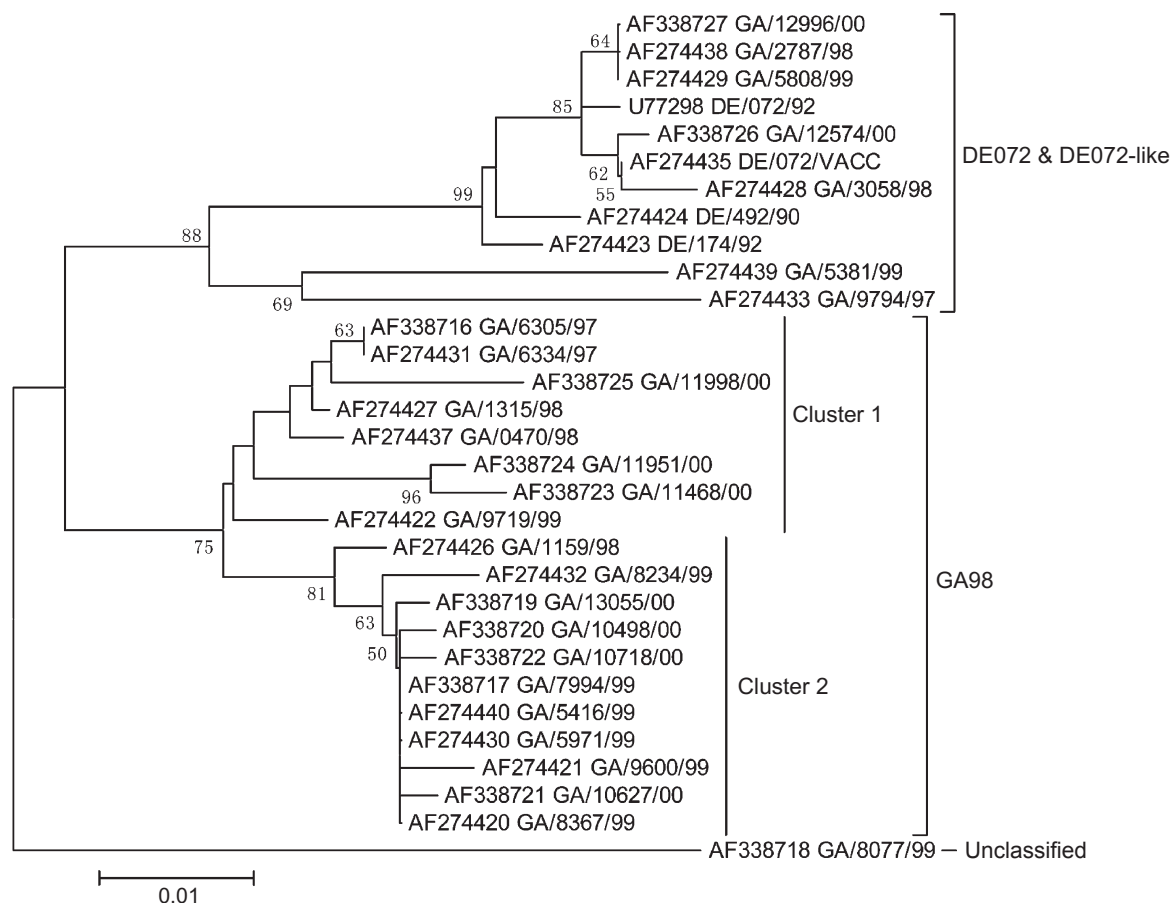
Correspondence: Ya-ping Zhang

Laboratory of Molecular Biology of Domestic Animals, and Cellular and Molecular Evolution, Kunming Institute of Zoology, Kunming 650223, China

Tel: +86-871-5190761; Fax: +86-871-5195430;

E-mail: zhangyp1@263.net.cn

Received 2 Sep 2005; revised 3 Nov 2005; accepted 23 Dec 2005; published online 16 March 2006



**Figure 1** Phylogenetic tree of 31 AIBV sequences. After removal of the ambiguous sites, tree is constructed using the NJ method. Bootstrap percentages higher than 50 are shown above the branch. The scale bar shows 0.01 nucleotide substitutions per site.

result of the escaping from the immunity of DE072 vaccine [6], were collected from literature [6]. After excluding those questionable and short sequences, there are 31 AIBV sequences in total for our analyses, including 4 DE072 and 27 GA98 S protein sequences. Their name and accession numbers in Genbank are listed in Figure 1.

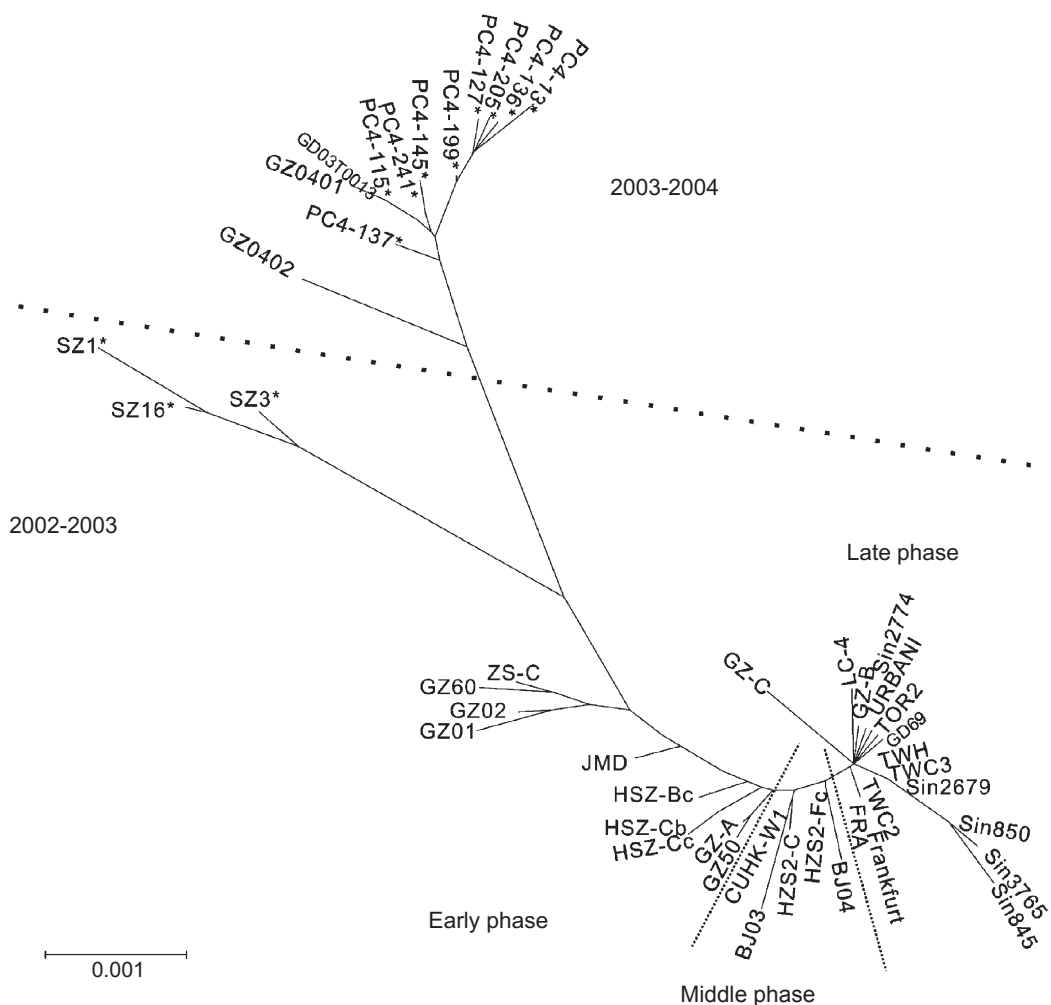
In addition, eighty six sequences from the first epidemic, which include 4 CoV sequences of the palm civets (*Paguma Larvata*) and 82 SARS-CoV sequences, representing three phases of first SARS outbreak [8], as well as thirteen sequences from the second epidemic were also retrieved from literature [8, 9]. The phylogenetic relationships between these sequences are presented in Figure 2.

#### Evolutionary analyses

The amino acid sequences were first aligned by CLUSTAL W [10]. Then, the nucleotide sequences were aligned according to the amino acid sequence alignment matrix and used in the phylogenetic reconstruction. We

conducted phylogenetic analysis using program MEGA2.1 [11]. The reliability of the resulting trees was evaluated by the bootstrap method [12] with 1 000 replications.

To identify putative amino acids that are subjected to the positive selection, the likelihood method of Yang *et al.* [13] was used with the PAML package [14] to estimate the numbers of synonymous substitutions per synonymous site ( $d_s$ ) and the number of nonsynonymous substitutions per nonsynonymous site ( $d_n$ ). First, the likelihood ratio test (LRT) was employed to test if positive selection exists by comparing a null model with a more general one. The null model does not allow for sites with  $\omega$  ( $d_n/d_s$ ) > 1, while the more general one does. Here the LRT compares M7 (null model) with M8 (general model) for the presence of sites under positive selection. M7 assumes that  $\omega$  ratios were distributed among sites by a beta distribution and M8 adds a discrete  $\omega$  class to M7. In general, positive selection can be inferred if  $\omega$  value estimated under M8 is greater than 1. Second, the Bayesian theorem, which calculates



**Figure 2** Phylogenetic tree of SARS sequences examined in this study. After removal of the identical sequences, 46 sequences were used to construct the tree using the NJ method. The sequences from palm civets are marked by star.

the posterior probability that a site has  $\omega > 1$ , is used to identify residues under positive selection when the LRT suggests their presence.

### Results and Discussion

In this study, we analyzed the spike protein (S) of DE072 vaccine and GA98 strain to identify putative amino acids that are subjected to positive selection and contribute to the evasion from the host immunity. The results revealed that inferred amino acid substitutions predominantly occurred within the regions of conformational virus-neutralizing epitopes previously recognized by monoclonal antibodies analyses [15], indicating that these substitutions may be responsible for the emergence of antigenic variations and that they may be crucial for the evasion of GA98 serotype

from DE072 immunity. Specifically, we found that 6 of 9 amino acid changes that were under positive selection in GA98 strain tended to switch from polar to non-polar or from neutral to charged residues (Table 1). For example, positions 60, 67 and 121 exhibit the amino acid transformation from neutral (A, G, Q) in DE072 to charged (E, R, K) in GA98 population, and positions 23, 123 and 278 change from polar (E, T, E) in DE072 to non-polar (V, L, A) in GA98 population. Therefore, it appears that the shift of residue charges in S protein was especially important in determining the potential epitopes for antibody, as also has been demonstrated in other viruses[16]. Finally, the phylogenetic analysis showed that GA98 strain sequences can be divided into two clusters, and at some inferred sites, different amino acids substitutions were observed between the two clusters of GA98 strain (Table 1). It is possible that

**Table 1** LRTs of positive selection for the SARS and AIBV

Viruses/epidemics	<i>n</i>	<i>L<sub>s</sub></i>	$\frac{2\Delta l}{M7\text{vs.}M8}$	Parameters estimated under M8	Positively selected sites <sup>a</sup>
SARS First epidemics (2002-2003) <sup>b</sup>	86	3753	20.12***	$p_0 = 0.9253$ $p_1 = 0.0747$ $\omega = 9.57$	227K/N 261K/T 479K/N 607P/S 701L/S 743A/T 754V/A 894A/T
SARS Second epidemics (2003-2004)	13	3765	10.5*	$p_0 = 0.9289$ $p_1 = 0.0711$ $\omega = 4.98$	479R/N 609L/A 613E/D 765V/A
AIBV <sup>c</sup>	31	1602	71.58***	$p_0 = 0.9588$ $p_1 = 0.0412$ $\omega = 13.55$	23E/V 60A/E 67G/R 121Q/K 123T/L 144V 278E/A 282K <u>388K</u>

Note: <sup>a</sup> The sites with posterior probabilities > 99% under M8 estimated are listed. The sites are indexed by their position in the alignment.

<sup>b</sup> Only those positive selection sites that distinguished animal and human viruses are presented. Amino acids from palm civet are presented before slash, whereas those from human are listed after slash. <sup>c</sup> The sites that distinguished two clusters of GA98 strain are underlined. Amino acids from GA98 strains are listed before slash and those from DE072 after slash. \*Significant at 5% level. \*\*\* Significant at 0.01% level.

these substitutions are associated with strain virulence differences of two GA98 serotype clusters.

It is interesting to see whether the substitution features identified to be important for the survival of GA98 AIBV are also characteristic of SARS-CoV. Human SARS epidemics are generally believed to originate from the animals with the palm civet as the primary suspect [8, 9, 17]. The comparison between the CoV sequences of the palm civets and SARS-CoV sequences allows inference of the importance of amino acid changes since the divergence of the virus [9]. After the first epidemic of SARS occurring during 2002-2003, scattered new cases were reported in Guangzhou during 2003-2004. Phylogenetic analysis has suggested that the new epidemics may be caused by independent viral invasion from animal to human [17]. Thus, we analyzed the sequences of animal and human viruses from two epidemics separately. In the first epidemic, we found 8 of the amino acids changes under positive selection distinguishing animal and human viruses (Table 1). The first three occur in the S1 subunit and all exhibit an interesting amino acid transformation from negative charge (K) to neutral residues (N or T). Four of the remaining five revealed the alterations of amino acid polarity from non-polar to polar residues, more specifically, from P, L, A and A in CoV of the palm civets to S, S, R/T and T in SARS-CoV, respectively. These changes may influence the conformation of S protein and consequently the power of antibody binding. Similarly, among the 4 amino acid changes under positive selection in the second epidemic we identified two charged residue shifts (Table 1). Importantly, one of these two positively selected sites (479) was observed not only in this epidemic but also in the first epidemic, with the same trends of alterations of amino acid charge from negative charge (K, R) to neutral residues (N or T). Interestingly,

this amino acid has also been suggested to be under positive selection in Song *et al.* (2005) by analyzing SNV (single-nucleotide variation) in different populations. Moreover, this amino acid is located in the region (residues 318–510), where it can efficiently bind angiotensin-converting enzyme 2, a functional receptor for SARS-Cov [18]. Since the CoV of palm civets and SARS-CoV are analogous to the DE072 vaccine and GA98 strain AIBV, our analysis suggests that significant changes of residue charge and polarity in critical proteins have enabled these viruses to establish themselves, although the direction of changes are different in the two situations.

In addition, as was found in the AIBV case, amino acids substitutions were found to be different among three clusters corresponding to three phases of the first SARS outbreak. For instance, two substitutions (75 T→R and 311 G→R) were only found in the early phase of Guangzhou and Zhongshan lineages [8], but not in the subsequent phases. Similarly, unique substitutions to the middle and late phases, respectively, were also present. The evidence suggests that these particular amino acid substitutions may affect strain virulence and infectivity. Hence, it can be followed that different genotypes of viruses must be taken into account when we adopt attenuated virus as SARS vaccine, similar to the common practice of administering influenza vaccine [19].

In conclusion, the implication of this study is that during the development of vaccines special attention needs to be paid to those amino acid changes that have resulted in overall shift of residue charge and polarity. Therefore, the current findings are expected not only to shed new light on the thorough understanding of avian infectious bronchitis viral evolution, but provide suggestive information in the drug and vaccination development of SARS as well, whether the

vaccine is based on the attenuated virus or DNA.

## Acknowledgements

This work was supported by special grants from Yunnan Province. We thank Wendy Grus, University of Michigan, for helpful comments on the manuscript.

## References:

- 1 King DJ, Cavanagh D. Infectious bronchitis. In: Calnek BW, Barnes HJ, Reid WM, Yoder HW. Eds. Diseases of poultry, 9<sup>th</sup> ed Iowa State University Press: Ames 1991:471-484.
- 2 Collisson EW, Parr RL, Wang L, Williams AK. An overview of the molecular characteristics of avian infectious bronchitis virus. *Poult Sci Rev* 1992; **4**:41-55.
- 3 Cavanagh D, Naqil SA. Infectious bronchitis. In: Calnek BW, Barnes HJ, Beard CW, McDonald YM. Eds. Diseases of poultry, 10<sup>th</sup> ed Iowa State University Press: Ames 1997:511-526.
- 4 Cavanagh D, Davis PJ. Coronavirus IBV: removal of spike glycopolypeptide S1 by urea abolishes infectivity and haemagglutination but not attachment to cells. *J Gen Virol* 1986; **67**:1443-1448.
- 5 Cavanagh D, Davis PJ, Mockett AP. Amino acids within hypervariable region 1 of avian coronavirus IBV (Massachusetts serotype) spike glycoprotein are associated with neutralization epitopes. *Virus Res* 1988; **11**:141-150.
- 6 Lee CW, Jackwood MW. Origin and evolution of Georgia 98 (GA98), a new serotype of avian infectious bronchitis virus. *Virus Res* 2001; **80**:33-39.
- 7 Pei J, Briles WE, Collisson EW. Memory T cells protect chicks from acute infectious bronchitis virus infection. *Virology* 2003; **306**:376-384.
- 8 Chinese SMEC. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science* 2004; **303**:1666-1669.
- 9 Song HD, Tu CC, Zhang GW, *et al.* Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc Natl Acad Sci U S A* 2005; **102**:2430-2435.
- 10 Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; **22**:4673-4680.
- 11 Kumar S, Tamura K, Jakobsen IB, Nei M. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 2001; **17**:1244-1245.
- 12 Felsenstein. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985; **39**:783-791.
- 13 Yang Z, Nielsen R, Goldman N, Pedersen AM. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 2000; **155**:431-449.
- 14 Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 1997; **13**:555-556.
- 15 Karaca K, Naqi S, Gelb J Jr. Production and characterization of monoclonal antibodies to three infectious bronchitis virus serotypes. *Avian Dis* 1992; **36**:903-915.
- 16 Hopp TP, Woods KR. Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci U S A* 1981; **78**:3824-3828.
- 17 Guan Y, Zheng BJ, He YQ, *et al.* Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* 2003; **302**:276-278.
- 18 Wong SK, Li W, Moore MJ, Choe H, Farzan M. A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. *J Biol Chem* 2004; **279**:3197-3201.
- 19 Fitch WM, Bush RM, Bender CA, Cox NJ. Long term trends in the evolution of H(3) HA1 human influenza type A. *Proc Natl Acad Sci U S A* 1997; **94**:7712-7718.

Edited by Yufang Shi