Recognition for avian influenza virus proteins based on support vector machine and linear discriminant analysis

LIANG GuiZhao1,2,3, CHEN ZeCong1,2, YANG ShanBin1,2, MEI Hu2,3, ZHOU Yuan2,3, YANG Li2,3, ZHOU Peng1,3, YANG ShengXi1, SHU Mao1,3, LIAO ChunYang1,2, WU ShiRong1,2, LI GenRong2,3, HE Liu2,3, GAO JianKun2,3, Gan MengYu1,2,3, LI DeJing1,2, CHEN GuoPing2,3, WANG GuiXue2,3, LONG Sha2,3, JING JuHua2,3, ZHENG XiaoLin2,3, ZENG Hui1,2, ZHANG QiaoXia1,2, ZHANG MengJun1,2, YANG Qi1,2, TIAN FeiFei2,3, TONG JianBo1,2, WANG JiaoNa1, LIU YongHong1,2,3, LI Bo1, QIU LiangJia2,3, CAI ShaoXi2,3, ZHAO Na1,2,3, YANG Yan1,2,3, SU Xiali2,3, SONG Jian1,2, CHEN MeiXia1,2, ZHANG XueJiao1,2, SUN JiaYing1,2, LI JingWei1,2,3, CHEN GuoHua1,2, CHEN Gang1,4, DENG Jie1,2,3, PENG ChuanYou1,2, ZHU WanPing1,2, XU Luolan1,2,3, WU YuQuan1,2, LIAO LiMin1,2, LI Zhi1,2, LI Jun1,2, LI JingWei1,2,3, HUANG ZhengHu1,2, ZHOU Ping1,2, LI ZhiLiang1,2

1 College of Chemistry and Chemical Engineering, Chongqing University, Chongqing 400044, China; 2 Key Laboratory for Biomechanics and Tissue Engineering under Ministry of Education at Chongqing University, Chongqing 400044, China; 3 College of Bioengineering, Chongqing University, Chongqing 400044, China; 4 Technology Centre for Life Sciences and Biotechnology, Singapore Polytechnic, Singapore 139651, Singapore

Total 200 properties related to structural characteristics were employed to represent structures of 400 HA coded proteins of influenza virus as training samples. Some recognition models for HA proteins of avian influenza virus (AIV) were developed using support vector machine (SVM) and linear discriminant analysis (LDA). The results obtained from LDA are as follows: the identification accuracy (Ria) for training samples is 99.8% and Ria by leave one out cross validation is 99.5%. Both Ria of 99.8% for training samples and Ria of 99.3% by leave one out cross validation are obtained using SVM model, respectively. External 200 HA proteins of influenza virus were used to validate the external predictive power of the resulting model. The external Ria for them is 95.5% by LDA and 96.5% by SVM, respectively, which shows that HA proteins of AIVs are preferably recognized by SVM and LDA, and the performances by SVM are superior to those by LDA.

Influenza viruses, including A, B, and C types, belong to the family of Orthomyxoviridae. Avian influenza is a syndrome of poultry and livestock caused by influenza A viruses1,2. Avian influenza viruses (AIVs) are classified on the basis of the antigenic properties of the hemagglutinin (HA) and the neuraminidase (NA) glycoproteins expressed on the surface of virus particles. To date, AIVs representing 16 HA and 9 NA subtypes have

Received November 27, 2005; accepted March 13, 2006
doi: 10.1007/s11426-008-0006-7

1 Corresponding author (email: zlli@163.com; zlli2662@163.com; sdqdlgz@163.com)

Supported by Foundations of National High Technology (863) Programme (Grant No. 2006AA02Z312), Innovative Group Programme for Graduates of Chongqing University, Science and Innovation Fund (Grant No. 2007711C1A0010260), National 111 Programme Introducing Talents of Discipline to Universities (Grant No. 050711106), Chongqing Municipality Basic and Applied Fundamental Science Fund (Grant No. 01-3-6), National Chunhui Project Foundation (Grant No. 99-4-4-3-7), State Key Laboratory of Chemo/Biosensing and Chemometrics Fund (Grant No.2005012), Fok-Yingtung Educational Foundation (Grant No. 98-7-6)
been detected in wild birds and poultry throughout the world. Rapid detection and recognition for the viruses are required to protect human and animal because of the changeability of AIV antigen and the risk to cause illness. Type A, B and C can be differentiated by the antigens differences obtained from agar diffuse and alexin combine experiments. But it is time consuming and painful to test all of them by experiments. In this study, total 200 properties of HA coded proteins were culled to represent structures of 400 HA proteins of AIVs, B and C influenza viruses, and recognition models for them were built by linear discriminant analysis and support vector machine. All 200 HA sequences of testing set were used to validate the external power of the models obtained. Then structural characteristics of AIVs, B and C viruses were primarily investigated. Thoughts and results in this work can provide with helpful references and foundations for study on corresponding genomics and proteomics.

1 Principles and methods

1.1 Sequences preparation and parametrization
Total 400 samples (Table 1 in supplementary materials) including 200 HA proteins of AIVs, 100 HA proteins of type B viruses and 100 HA proteins of type C viruses were randomly selected from GenBank. It has been documented\(^\text{[2]}\) that the number of basic amino acid of HA and distribution of the protease as cleavage are the main factors in breaking into cells and determining the pathogenicity. Therefore, the pathogenicity are influenced by the composition, amino acid position and the physicochemical properties and conformation characteristics of HA sequences. A variety of parameters (Table 2 in supplementary materials) including molecular weight, isoelectric point, volume, specific volume, steric descriptors, electric characters, hydrophobicity, alpha and turn propensities, beta propensity, the time and frequency of amino acid, free energy, etc. were widely collected to roundly represent the structures of HA sequences.

1.2 Linear discriminant analysis (LDA) modeling
Linear discriminant analysis (LDA)\(^\text{[3]}\) realizes the process that how objects will be classified according to many observed samples. The basic theory of it is to classify the dependence by dividing an n-dimensional descriptor space into two regions (two classes) that are separated by a hyperplane defined by a linear discriminant function as follows: 

\[ Y = a_0 + a_1X_1 + a_2X_2 + \cdots + a_nX_n, \]

where \( Y \) is the dependent variable, \( X_1, X_2, \ldots, X_n \) represents the independent variable (observed values), \( a_1, a_2, \ldots, a_n \) corresponds to weights associated with the respective independent variable, that is discriminant coefficients. Independent variable space is divided into two regions through the hyperplane, then to discriminate which region every compound belongs to. In this study, LDA was carried out by the software of SPSS13.0. In order to decrease the number of descriptors and make the model interpreted, variables entering the model were selected by a stepwise method. The selection of the variables was chosen according to \( F \) value of partial \( F \) test: The variable was accepted into the model when \( F \) value calculated was greater than 3.84; the variable was rejected when \( F \) value was less than 2.71.

1.3 Support vector machine modeling
Extensive attention and interest have been attracted due to many attractive features and promising empirical performances of support vector machine (SVM)\(^\text{[4-11]}\). Optimal interface found does not implement the minimum experimental error but minimum generalization error based on the theory of structural risk minimization. SVM can not only solve some problems such as nonlinear, high dimensional and local minimum but also avoid over-training phenomenon from other conventional machine learning. In linearly separable cases, SVM constructs a hyperplane which separates two different classes of vectors with a maximum margin. In nonlinearly separable cases, SVM maps the input variables into a high-dimensional feature space using a kernel function 

\[ K(x, y) = \Phi(x) \cdot \Phi(y) \]

which can effectively deal with many problems such as dimensional puzzledom, calculation complexity, etc. At present, familiar kernel functions are mainly: linear kernel function: 

\[ K(x, y) = x \cdot y; \]

polynomial kernel function: 

\[ K(x, y) = (a_1x \cdot x + a_2)^p; \]

radial basis kernel function: 

\[ K(x, y) = \exp(-\gamma ||x - y||^2); \]

Sigmoid kernel function: 

\[ K(x, y) = \tanh (a_1x \cdot x + a_2). \]

SVM with radial basis kernel function was applied to create recognition models of AIVs in this work. SVM was implemented using Matlab7.0.

1.4 Model evaluation and verification
Identification accuracy (\( R_u \)) was used to measure the performance of this method: 

\[ R_u = (N_t - N_e)/N_t \times 100\%, \]
where \( N_t \) is the number of all samples; \( N_e \) represents the number of misclassified samples. It is necessary to validate the external predictive capacity of a model\(^{[12]}\), so 200 HA sequences of influenza viruses (STable 3 in supplementary materials) not present in the training samples retrieved from GenBank as external test set were used to evaluate the external predictive capacity of the models built.

## 2 Results and discussion

### 2.1 Recognition results by LDA

Among the 200 variables, 24 were selected to construct the model by LDA through a stepwise manner. 24 variables, original and standardized coefficients for LDA are collected in Table 1. \( R_{ia} \) for 400 HA sequences was 99.8%, and \( R_{ia} \) of 99.5% by CV was obtained (STable 1 in supplementary materials).

Significance of 24 variables affecting identification results was analyzed according to their standardized coefficients. Table 4 shows the absolute value of standardized coefficients corresponding to the variables of the frequency of W, retention coefficient in TFA, the quality percentage of R, the quality percentage of Q, the times of N and value of theta (\( \theta \)) are greater than 5.000. So, it is predicted that these characteristics may bring significant influence on emergence of AIVs, type B viruses and type C viruses. The largest coefficient in all ones is +10.157, as the contribution of the variable related to the times of N. The acylamino group of the side chain of N is an unionized polar group, but it can form hydrogen bond with other atoms. Consequently, we speculated that there is higher tendency on forming AIVs with the increase of the times of N in a sequence. The standardization coefficient of retention coefficient in TFA which refers to a hydrophobic property is 8.289, indicating that it takes on positive influence on the creation of HA proteins of AIVs. The quality percentage of Q with the largest negative coefficient may not easily lead to HA proteins of AIVs. It can be reasoned that although the acylamino group of the side chain of Q can bring hydrogen bond with other atoms, it may be influenced by the quality percentage of other amino acids. As a result, the larger the quality percentage of other amino acids in a sequence is, the easier the development of HA proteins is.

### Table 1 The 24 variables selected, \( F \) values, original (\( b_i \)), and standardized coefficients (\( b'_i \)) for LDA model

| No. | Description of variables                               | \( F \) to remove | \( b_i \)  | \( b'_i \) |
|-----|----------------------------------------------------------|-------------------|------------|------------|
| 0   | constant                                                 | /                 | 1.067      | /          |
| 1   | frequency of W                                          | 130.209           | –0.265     | –5.292     |
| 2   | sum of geometrical distances between N and N             | 12.377            | –0.064     | –3.214     |
| 3   | retention coefficient in TFA                            | 269.913           | 334.507    | 8.289      |
| 4   | time of basic amino acids                               | 83.583            | 0.276      | 3.334      |
| 5   | times of H                                              | 11.304            | 0.091      | 0.742      |
| 6   | frequency of V                                          | 6.760             | –0.150     | –0.862     |
| 7   | times of Q                                              | 98.835            | 0.425      | 2.724      |
| 8   | times of G                                              | 58.201            | 105.144    | 2.193      |
| 9   | quality percentage of R                                 | 305.617           | –0.442     | –6.223     |
| 10  | side chain angle theta                                  | 255.701           | 1.165      | 4.032      |
| 11  | frequency of S                                          | 130.073           | 0.332      | 4.023      |
| 12  | times of I                                              | 72.358            | 0.431      | 2.774      |
| 13  | times of V                                              | 36.268            | –61.539    | –1.183     |
| 14  | quality percentage of Q                                 | 23.253            | –338.692   | –7.633     |
| 15  | times of N                                              | 40.011            | 595.566    | 10.157     |
| 16  | times of C                                              | 129.147           | 0.675      | 2.461      |
| 17  | average accessible surface area                          | 54.213            | –0.223     | –2.258     |
| 18  | times of M                                              | 147.924           | 108.587    | 1.553      |
| 19  | frequency of M                                          | 80.158            | 4.960      | 2.447      |
| 20  | quality percentage of S                                 | 62.693            | 1.422      | 2.477      |
| 21  | times of D                                              | 55.766            | –1.180     | –3.938     |
| 22  | parameter of charge transfer donor capability            | 3.842             | 1.833      | 1.747      |
| 23  | time of polar amino acids                               | 5.354             | 13.251     | 0.388      |
| 24  | value of theta (\( \theta \))                           | 138.692           | –21.630    | –5.997     |
Corresponding coefficients of the value of theta (i) and the frequency of W are less than 0.000, indicating that they are fit not for the development of HA proteins of AIVs but for the emergence of HA proteins of type B and C viruses. Especially for W, the spacial effects from the indolyl group of its side chain may result in adverse action to develop HA proteins of AIVs. Other variables are not analyzed in here.

According to the original equation, the centroids were 3.729 and −3.729 for HA group of AIVs and HA group of non-AIVs, respectively. The selected 24 variables of a sequence were introduced into the original equation. As a result, it was HA proteins of AIVs if discriminant value was more than 0.000, otherwise, it was HA group of non-AIVs. Finally, total \( R_{ia} \) of 95.5\% for 200 sequences (STable 3 in supplementary materials) indicates that the LDA model cherishes favorable external identification capacity.

2.2 Recognition results by SVM

Parameters are crucial to SVM regression. As for SVM with radial basis kernel function, they are regularized coefficient \( C \) and the kernel style \( K \) and its corresponding parameters. \( C \) is a regulative parameter, which is used to control the tradeoff maximizing the margin and minimizing the training error between maximal interval and minimum training error. Insufficient stress will be placed on fitting the training data if \( C \) is too small, contrarily, fitting will be beyond if \( C \) is too large. The \( \gamma \) controls amplitude of Gaussian function, that is, generality capability of SVM. In the present study, SVM parameters were confirmed by \( R_{ia} \) of external validation using single factor alternation method.

Based on the SVM model developed by 200 original variables, \( R_{ia} \) for 400 training samples and 200 testing samples were both relatively low. So, 16 principal components whose corresponding eigenvalues were larger than 1.0 were derived from 200 original variables by principal component analysis\(^ {13}\). These 16 principal components explained 98.2\% variance of the original variable matrix. Scores of the 16 principal components were treated as the input of SVM. Firstly, \( \gamma \) value is set as 0.031, then \( C \) is 5, 10, 30, 50, 100, 300, 600 and 1000, respectively. Plot of \( \gamma \) versus \( R_{ia} \) shows (Figure 2) that \( R_{ia} \) is 96.5\% when \( \gamma \) is 0.031. So, parameters of SVM are as follows: \( C = 100 \), the kernel function: \( K(x, x_i) = \exp(-0.031 ||x - x_i||^2) \). \( R_{ia} \) for 400 HA sequences is 99.8\% which is equivalent to that of LDA. \( R_{ia} \) of CV (99.3\%) is a bit less than that of LDA (99.5\%), but \( R_{ia} \) (96.5\%) for 200 external samples is higher than that of LDA (95.5\%).

3 Conclusion

Rapid identification for AIVs is very important for taking effective methods to control its pandemics\(^ {14-18}\). 200 local and whole parameters of influenza viruses were culled to recognize their structural characteristics. Then, recognition models for AIVs were developed by LDA and SVM. The results indicate that two models can preferably discriminate HA proteins of AIV from HA proteins of non-AIV. By comparison, the performances of SVM are better than those of LDA. The recognition
models for AIVs can help researchers investigate the differences between AIV proteins and non-AIV proteins, such as coding proteins related to severe acute respira-
tory syndrome coronavirus (SARS-CoV)\textsuperscript{[10]}, which can provide useful references for correlated genomics and proteomics.

1 Gan M H. Avian Influenza (the second edition). Beijing: Chinese Agriculture Press, 2002. 15—79
2 Garten W, Klenk H D. Understanding influenza virus pathogenicity. Trends in Microbiol, 1999, 7(3): 99 — 100
3 Kachigan S K. Statistical analysis. New York: Radius Press, 1986: 264—285
4 Burbidge R, Trotter M, Buxton B. Drug design by machine learning: Support vector machines for pharmaceutical data analysis. Comp and Chem, 2001, 26: 5 — 14
5 Hua S, Sun Z. A novel method of protein secondary structure prediction with high segment overlap measure: Support vector machine approach. J Mol Biol, 2001, 308: 397—407
6 Hong S J, Weiss S M. Advances in predictive models for data mining. Pattern Recogn Lett, 2001, 22: 55 — 61
7 Belousov A I, Verzakov S A, Frese J von. A flexible classification approach with optimal generalization performance: support vector machines. Chemom Intell Lab Syst, 2002, 64: 15 — 25
8 Gao J B, Gunn S R, Harris C J. SVM regression through variational methods and its sequential implementation. Neurocomputing, 2003, 55: 151 — 167
9 Cortes C, Vapnik V. Support vector networks. Mach Learn, 1995, 20: 273 — 293
10 Flach P A. On the state of the art in machine learning: A personal review. Artif Intell, 2001, 131: 199—222
11 Sánchez A V D. Advanced support vector machines and kernel methods. Neurocomputing, 2003, 55: 5 — 20
12 Tropsha A, Gramatica P, Gombar V K. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. QSAR Comb Sci, 2003, 22: 69—77
13 Kim D, Lee I-B. Process monitoring based on probabilistic PCA. Chemom Intell Lab Syst, 2003, 67: 109—123
14 Guan Y, Poon L L M, Cheung C Y. H5N1 influenza: A protein pandemic threat. Proc Natl Acad Sci, 2004, 101(21): 8156—8161
15 Chen H, Deng G, Li Z, Tian G, Li Y, Jiao P, Zhang L, Liu Z, Webster R G, Yu K. The evolution of H5N1 influenza viruses in ducks in southern China, Proc Natl Acad Sci USA, 2004, 101: 10452—10457
16 Guan Y, Shortridge K F, Krauss S, et al. Molecular characterization of H9N2 influenza viruses: Were they the donors of the “internal” genes of H5N1 viruses in Hong Kong? Proc Natl Acad Sci USA, 1999; 96: 9363—9367
17 Chen H, Smith G J D, Li K S. Establishment of multiple sublineages of H5N1 influenza virus in Asia: Implications for pandemic control. Proc Natl Acad Sci USA, 2006, 103: 2845—2850
18 Lin Y P, Shaw M, Gregory V. Avian-to-human transmission of H9N2 subtype influenza A viruses: Relationship between H9N2 and H5N1 human isolates. Proc Natl Acad Sci USA, 2000, 97: 9654—9658
19 Mei H, Sun L L, Zhou Y, Xing Q, Li Z. Identification of coding proteins related to SARS-CoV. Chin Sci Bull, 2004, 49(19): 2037—2040
| No. | Strains  | Accession number | Type | LDA | LDAcv | SVM | SVMcv |
|-----|----------|------------------|------|-----|-------|-----|-------|
| 1   | A/avian/NY/53726/00(H5N2) | AAP57601 | +    | +   | +    | +   | +     |
| 2   | A/avian/NY/31588-2/00(H5N2) | AAP57596 | +    | +   | +    | +   | +     |
| 3   | A/avian/NY/53726/00(H5N2) | AAP57577 | +    | +   | +    | +   | +     |
| 4   | A/avian/NY/31588-3/00(H5N2) | AAP57573 | +    | +   | +    | +   | +     |
| 5   | A/avian/NY/31588-2/00(H5N2) | AAP57572 | +    | +   | +    | +   | +     |
| 6   | A/avian/NY/53726/00(H5N2) | AAP57548 | +    | +   | +    | +   | +     |
| 7   | A/avian/NY/31588-3/00(H5N2) | AAP57544 | +    | +   | +    | +   | +     |
| 8   | A/avian/NY/31588-2/00(H5N2) | AAP57543 | +    | +   | +    | +   | +     |
| 9   | A/avian/China/417-428/98(H9N2) | AAG53055 | +    | +   | +    | +   | +     |
| 10  | A/avian/Ireland/PV46B/93(H9N3) | AAG53046 | +    | +   | +    | +   | +     |
| 11  | A/avian/NY/17150-7/2000(H6N2) | ABA12261 | +    | +   | +    | +   | +     |
| 12  | A/avian/NY/17150-7/2000(H6N2) | ABA12260 | +    | +   | +    | +   | -     |
| 13  | A/avian/NY/31588-3/00(H5N2) | AAP57597 | +    | +   | +    | +   | +     |
| 14  | A/avian/NY/81746-5/00(H7N2) | AAP21438 | +    | +   | +    | +   | +     |
| 15  | A/avian/NY/76247-3/00(H7N2) | AAP21437 | +    | +   | +    | +   | +     |
| 16  | A/avian/NY/118353-1/01(H7N2) | AAP21436 | +    | +   | +    | +   | +     |
| 17  | A/avian/NY/73063-6/00(H7N2) | AAO86952 | +    | +   | +    | +   | +     |
| 18  | A/avian/NY/7729-6/00(H7N2) | AAO86909 | +    | +   | +    | +   | +     |
| 19  | A/avian/NY/81746-5/00(H7N2) | AAO86908 | +    | +   | +    | +   | +     |
| 20  | A/avian/NY/76247-3/00(H7N2) | AAO86907 | +    | +   | +    | +   | +     |
| 21  | A/avian/NY/74211-2/00(H7N2) | AAO86906 | +    | +   | +    | +   | +     |
| 22  | A/avian/NY/70411-12/00(H7N2) | AAO86905 | +    | +   | +    | +   | +     |
| 23  | A/avian/NY/118353-1/01(H7N2) | AAO86904 | +    | +   | +    | +   | +     |
| 24  | A/avian/NY/53726/00(H5N2) | AAP71999 | +    | +   | +    | +   | +     |
| 25  | A/avian/NY/31588-3/00(H5N2) | AAP71995 | +    | +   | +    | +   | +     |
| 26  | A/Hong Kong/482/97(H5N1) | AAK49273 | +    | +   | +    | +   | +     |
| 27  | A/Hong Kong/481/97(H5N1) | AAK49272 | +    | +   | +    | +   | +     |
| 28  | A/Hong Kong/1074/99(H9N2) | AAK49271 | +    | +   | +    | +   | +     |
| 29  | A/Hong Kong/1073/99(H9N2) | AAK49270 | +    | +   | +    | +   | +     |
| 30  | A/Hong Kong/516/97(H5N1) | AAK49269 | +    | +   | +    | +   | +     |
| 31  | A/Hong Kong/516/97(H5N1) | AAK49268 | +    | +   | +    | +   | +     |
| 32  | A/Hong Kong/514/97(H5N1) | AAK49267 | +    | +   | +    | +   | +     |
| 33  | A/Hong Kong/514/97(H5N1) | AAK49266 | +    | +   | +    | +   | +     |
| 34  | A/Hong Kong/507/97(H5N1) | AAK49265 | +    | +   | +    | +   | +     |
| 35  | A/Hong Kong/507/97(H5N1) | AAK49264 | +    | +   | +    | +   | +     |
| 36  | A/Hong Kong/503/97(H5N1) | AAK49263 | +    | +   | +    | +   | +     |
| 37  | A/Hong Kong/503/97(H5N1) | AAK49262 | +    | +   | +    | +   | +     |
| 38  | A/Hong Kong/491/97(H5N1) | AAK49261 | +    | +   | +    | +   | +     |
| 39  | A/Hong Kong/491/97(H5N1) | AAK49260 | +    | +   | +    | +   | +     |
| 40  | A/Hong Kong/491/97(H5N1) | AAK49259 | +    | +   | +    | +   | +     |
| 41  | A/Hong Kong/488/97(H5N1) | AAK49258 | +    | +   | +    | +   | +     |
| 42  | A/Hong Kong/488/97(H5N1) | AAK49257 | +    | +   | +    | +   | +     |
| 43  | A/Hong Kong/488/97(H5N1) | AAK49256 | +    | +   | +    | +   | +     |
| 44  | A/Hong Kong/485/97(H5N1) | AAK49255 | +    | +   | +    | +   | +     |
| 45  | A/Hong Kong/485/97(H5N1) | AAK49254 | +    | +   | +    | +   | +     |
| 46  | A/Hong Kong/485/97(H5N1) | AAK49253 | +    | +   | +    | +   | +     |
| 47  | A/Hong Kong/542/97(H5N1) | AAK49250 | +    | +   | +    | +   | +     |
| 48  | A/Hong Kong/542/97(H5N1) | AAK49249 | +    | +   | +    | +   | +     |
| 49  | A/Hong Kong/538/97(H5N1) | AAK49248 | +    | +   | +    | +   | +     |
| 50  | A/Hong Kong/538/97(H5N1) | AAK49247 | +    | +   | +    | +   | +     |
| 51  | A/turkey/Italy/125/89(H9N2) | AAG53044 | +    | +   | +    | +   | +     |
| 52  | A/turkey/Pennsylvania/11086-2/88(H9N2) | AAG53043 | +    | +   | +    | +   | +     |
| 53  | A/turkey/Minnesota/46719-5/88(H9N2) | AAG53042 | +    | +   | +    | +   | +     |
| 54  | A/duck/NewYork/17542-4/86(H9N1) | AAG53041 | +    | +   | +    | +   | +     |
| 55  | A/duck/India/31g/86(H9N2) | AAG53040 | +    | +   | +    | +   | +     |

*(To be continued on the next page)*
| No. | Strains | Accession number | Type | LDA | LDA<sub>cv</sub> | SVM | SVM<sub>cv</sub> |
|-----|---------|-----------------|------|-----|----------------|-----|-------------|
| 56  | A/chicken/Italy/169/85(H9N2) | AAG53039 | +    | +   | +              | +   | +           |
| 57  | A/turkey/Italy/245/84(H9N2)  | AAG53038 | +    | +   | +              | +   | +           |
| 58  | A/turkey/Italy/VR86/83(H9N2) | AAG53037 | +    | +   | +              | +   | +           |
| 59  | A/duck/Hong Kong/D91/76(H9N2) | AAG53036 | +    | +   | +              | +   | +           |
| 60  | A/turkey/Scotland/70(H9N7)    | AAG53035 | +    | +   | +              | +   | +           |
| 61  | A/CNIC/149/98(H3N2)          | AAK49204 | +    | +   | +              | +   | +           |
| 62  | A/CNIC/146/98(H3N2)          | AAK49203 | +    | +   | +              | +   | +           |
| 63  | A/CNIC/145/98(H3N2)          | AAK49202 | +    | +   | +              | +   | +           |
| 64  | A/CNIC/130/98(H3N2)          | AAK49201 | +    | +   | +              | +   | +           |
| 65  | A/CNIC/125/98(H3N2)          | AAK49200 | +    | +   | +              | +   | +           |
| 66  | A/CNIC/121/98(H3N2)          | AAK49199 | +    | +   | +              | +   | +           |
| 67  | A/CNIC/109/98(H3N2)          | AAK49198 | +    | +   | +              | +   | +           |
| 68  | A/CNIC/97/98(H3N2)           | AAK49197 | +    | +   | +              | +   | +           |
| 69  | A/CNIC/96/98(H3N2)           | AAK49196 | +    | +   | +              | +   | +           |
| 70  | A/CNIC/52/98(H3N2)           | AAK49195 | +    | +   | +              | +   | +           |
| 71  | A/CNIC/3/98(H3N2)            | AAK49194 | +    | +   | +              | +   | +           |
| 72  | A/teal/Taiwan/WB2-37-2TPFE2/98(H7N1) | AAG10680 | +    | +   | +              | +   | +           |
| 73  | A/chicken/Italy/1081/99(H7N1) | AAG10679 | +    | +   | +              | +   | +           |
| 74  | A/peregrine falcon/U.A.E./188/2384/98(H7N3) | AAG10678 | +    | +   | +              | +   | +           |
| 75  | A/ostrich/South Africa/M320/96(H7N7) | AAG10677 | +    | -   | -              | +   | +           |
| 76  | A/Kalij Pheasant/Bangkok/Thailand/CU-18/04(H5N1) | AAZ9998 | +    | +   | +              | +   | +           |
| 77  | A/chicken/Saraburi/Thailand/CU-17/04(H5N1) | AAZ9997 | +    | +   | +              | +   | +           |
| 78  | A/white peafowl/Bangkok/Thailand/CU-16/04(H5N1) | AAZ9996 | +    | +   | +              | +   | +           |
| 79  | A/crow/Bangkok/Thailand/CU-15/04(H5N1) | AAZ9995 | +    | +   | +              | +   | +           |
| 80  | A/chicken/Nakhon Pathom/Thailand/CU-14/04(H5N1) | AAZ9994 | +    | +   | +              | +   | +           |
| 81  | A/chicken/Nakhon Sawan/Thailand/CU-13/04(H5N1) | AAZ9993 | +    | +   | +              | +   | +           |
| 82  | A/chicken/Nakhon Sawan/Thailand/CU-12/04(H5N1) | AAZ9992 | +    | +   | +              | +   | +           |
| 83  | A/chicken/Chachoengsao/Thailand/CU-11/04(H5N1) | AAZ9991 | +    | +   | +              | +   | +           |
| 84  | A/chicken/Chachoengsao/Thailand/CU-10/04(H5N1) | AAZ9990 | +    | +   | +              | +   | +           |
| 85  | A/chicken/Suphanburi/Thailand/CU-9/04(H5N1) | AAZ9989 | +    | +   | +              | +   | +           |
| 86  | A/chicken/Prachinburi/Thailand/CU-8/04(H5N1) | AAZ9988 | +    | +   | +              | +   | +           |
| 87  | A/chicken/Chonburi/Thailand/CU-7/04(H5N1) | AAZ9987 | +    | +   | +              | +   | +           |
| 88  | A/chicken/Bangkok/Thailand/CU-6/04(H5N1) | AAZ9986 | +    | +   | +              | +   | +           |
| 89  | A/chicken/Bangkok/Thailand/CU-3/04(H5N1) | AAZ9983 | +    | +   | +              | +   | +           |
| 90  | A/chicken/Suphanburi/Thailand/CU-1/04(H5N1) | AAZ9982 | +    | +   | +              | +   | +           |
| 91  | A/Mynas/Ranong/Thailand/CU-209/04(H5N1) | AAZ9981 | +    | +   | +              | +   | +           |
| 92  | A/sparrow/Phang-Nga/Thailand/CU-203/04(H5N1) | AAZ9980 | +    | +   | +              | +   | +           |
| 93  | A/pigeon/Samut Prakan/Thailand/CU-202/04(H5N1) | AAZ9979 | +    | +   | +              | +   | +           |
| 94  | A/chicken/Prachinburi/Thailand/CU-104/04(H5N1) | AAZ9978 | +    | +   | +              | +   | +           |
| 95  | A/duck/Saraburi/Thailand/CU-74/04(H5N1) | AAZ9977 | +    | +   | +              | +   | +           |
| 96  | A/chicken/Chonburi/Thailand/CU-73/04(H5N1) | AAZ9976 | +    | +   | +              | +   | +           |
| 97  | A/chicken/Nakhon Pathom/Thailand/CU-71/04(H5N1) | AAZ9975 | +    | +   | +              | +   | +           |
| 98  | A/chicken/Ratchaburi/Thailand/CU-68/04(H5N1) | AAZ9974 | +    | +   | +              | +   | +           |
| 99  | A/chicken/Nakhon Sawan/Thailand/CU-39/04(H5N1) | AAZ9973 | +    | +   | +              | +   | +           |
| 100 | A/chicken/Lopburi/Thailand/CU-38/04(H5N1) | AAZ9972 | +    | +   | +              | +   | +           |
| 101 | A/duck/Korea/S10/03(H3N2)    | AAW78083 | +    | +   | +              | +   | +           |
| 102 | A/duck/Korea/S9/03(H3N2)     | AAW78082 | +    | +   | +              | +   | +           |
| 103 | A/duck/Korea/S8/03(H3N2)     | AAW78081 | +    | +   | +              | +   | +           |
| 104 | A/duck/Korea/S7/03(H3N2)     | AAW78080 | +    | +   | +              | +   | +           |
| 105 | A/chicken/Korea/S6/03(H3N2)  | AAW78079 | +    | +   | +              | +   | +           |
| 106 | A/chicken/Korea/S18/03(H9N2) | AAW78078 | +    | +   | +              | +   | +           |
| 107 | A/chicken/Korea/S16/03(H9N2) | AAW78077 | +    | +   | +              | +   | +           |
| 108 | A/chicken/Korea/S15/03(H9N2) | AAW78076 | +    | +   | +              | +   | +           |
| 109 | A/dove/Korea/S14/03(H9N2)    | AAW78075 | +    | +   | +              | +   | +           |
| 110 | A/duck/Korea/S13/03(H9N2)    | AAW78074 | +    | +   | +              | +   | +           |

*(To be continued on the next page)*
| No. | Strains                                      | Accession number | Type | LDA/DA | SVM/SVM |
|-----|---------------------------------------------|------------------|------|--------|---------|
| 111 | A/chicken/Korea/S12/03(H9N2)                | AAW78073         | +    | +      | +       |
| 112 | A/chicken/Korea/S5/03(H9N2)                | AAW78072         | +    | +      | +       |
| 113 | A/chicken/Korea/S4/03(H9N2)                | AAW78071         | +    | +      | +       |
| 114 | A/silky chicken/Korea/S3/03(H9N2)          | AAW78070         | +    | +      | +       |
| 115 | A/duck/Korea/S17/03(H6N1)                  | AAW78069         | +    | +      | +       |
| 116 | A/dove/Korea/S11/03(H3N2)                  | AAW78068         | +    | +      | +       |
| 117 | A/duck/Korea/S10/03(H3N2)                  | AAW78067         | +    | +      | +       |
| 118 | A/duck/Korea/S9/03(H3N2)                   | AAW78066         | +    | +      | +       |
| 119 | A/duck/Korea/S8/03(H3N2)                   | AAW78065         | +    | +      | +       |
| 120 | A/duck/Korea/S7/03(H3N2)                   | AAW78064         | +    | +      | +       |
| 121 | A/chicken/Korea/S6/03(H3N2)                | AAW78063         | +    | +      | +       |
| 122 | A/chicken/Korea/S18/03(H9N2)               | AAW78062         | +    | +      | +       |
| 123 | A/chicken/Korea/S16/03(H9N2)               | AAW78061         | +    | +      | +       |
| 124 | A/chicken/Korea/S15/03(H9N2)               | AAW78060         | +    | +      | +       |
| 125 | A/dove/Korea/S14/03(H9N2)                  | AAW78059         | +    | +      | +       |
| 126 | A/chicken/Chile/176822/02(H7N3)            | AAQ77402         | +    | +      | +       |
| 127 | A/turkey/Italy/220158/2002(H7N3)           | CAF33023         | +    | +      | +       |
| 128 | A/turkey/Italy/220158/2002(H7N3)           | CAF33022         | +    | +      | +       |
| 129 | A/turkey/Italy/220158/2002(H7N3)           | CAF33020         | +    | +      | +       |
| 130 | A/turkey/Italy/214845/2002(H7N3)           | CAF33017         | +    | +      | +       |
| 131 | A/turkey/Italy/4603/1999(H7N1)             | CAF33013         | +    | +      | +       |
| 132 | A/turkey/Italy/4603/1999(H7N1)             | CAF33012         | +    | +      | +       |
| 133 | A/turkey/Italy/214845/2002(H7N3)           | CAF33011         | +    | +      | +       |
| 134 | A/turkey/Italy/214845/2002(H7N3)           | CAF33010         | +    | +      | +       |
| 135 | A/chicken/Mudanjiang/0823/2000(H9N2)        | AAS18237         | +    | +      | +       |
| 136 | A/chicken/Mudanjiang/0823/2000(H9N2)        | AAS18236         | +    | +      | +       |
| 137 | A/dove/China/E319-2/03(H5N1)               | AAR99633         | +    | +      | +       |
| 138 | A/dove/China/E319-2/03(H5N1)               | AAR99632         | +    | +      | +       |
| 139 | A/dove/China/E319-2/03(H5N1)               | AAR99631         | +    | +      | -       |
| 140 | A/dove/China/E319-2/03(H5N1)               | AAR99630         | +    | +      | +       |
| 141 | A/dove/China/E319-2/03(H5N1)               | AAR99629         | +    | +      | +       |
| 142 | A/dove/China/E319-2/03(H5N1)               | AAR99628         | +    | +      | +       |
| 143 | A/dove/China/E319-2/03(H5N1)               | AAR99627         | +    | +      | +       |
| 144 | A/dove/China/E319-2/03(H5N1)               | AAR99626         | +    | +      | +       |
| 145 | A/dove/China/E319-2/03(H5N1)               | AAR99625         | +    | +      | +       |
| 146 | A/dove/China/E319-2/03(H5N1)               | AAR99624         | +    | +      | +       |
| 147 | A/chicken/Mudanjiang/0823/00 (H9N2)         | AAR98872         | +    | +      | +       |
| 148 | A/Mallard/WI/944/82)                       | AAB39639         | +    | +      | +       |
| 149 | A/chicken/Germany/R28/03(H7N7)             | CAF04486         | +    | +      | +       |
| 150 | A/chicken/Germany/R28/03(H7N7)             | CAF04466         | +    | +      | +       |
| 151 | A/chicken/Hong Kong/830.2/01 (H5N1)        | AAO52871         | +    | +      | +       |
| 152 | A/chicken/Hong Kong/829.2/01 (H5N1)        | AAO52870         | +    | +      | +       |
| 153 | A/chicken/Hong Kong/822.1/01 (H5N1)        | AAO52869         | +    | +      | +       |
| 154 | A/chicken/Hong Kong/715.5/01 (H5N1)        | AAO52868         | +    | +      | +       |
| 155 | A/chicken/Hong Kong/SF219/01 (H5N1)        | AAO52867         | +    | +      | +       |
| 156 | A/pigeon/Hong Kong/SF215/01 (H5N1)         | AAO52866         | +    | +      | +       |
| 157 | A/quail/Hong Kong/SF203/01 (H5N1)          | AAO52865         | +    | +      | +       |
| 158 | A/silky chicken/Hong Kong/SF189/01 (H5N1)  | AAO52864         | +    | +      | +       |
| 159 | A/chicken/Hong Kong/FY150/01 (H5N1)        | AAO52862         | +    | +      | +       |
| 160 | A/chicken/Hong Kong/YU563/01 (H5N1)        | AAO52861         | +    | +      | +       |
| 161 | A/chicken/Hong Kong/YU562/01 (H5N1)        | AAO52860         | +    | +      | +       |
| 162 | A/chicken/Hong Kong/FY77/01 (H5N1)         | AAO52859         | +    | +      | +       |
| 163 | A/chicken/California/0139/2001(H6N2)       | AAO35354         | +    | +      | +       |
| 164 | A/chicken/California/650/2000(H6N2)        | AAO35359         | +    | +      | +       |
| 165 | A/chicken/California/9420/2001(H6N2)       | AAO35358         | +    | +      | +       |
| No. | Strains                        | Accession number | Type_1 | LDA | LDA_α | SVM | SVM_α |
|-----|--------------------------------|------------------|--------|-----|-------|-----|-------|
| 166 | A/chicken/California/9420/2001(H6N2) | AAO33537         | +      | +   | +     | +   | +     |
| 167 | A/chicken/California/9174/2001(H6N2) | AAO33536         | +      | +   | +     | +   | +     |
| 168 | A/chicken/California/9174/2001(H6N2) | AAO33535         | +      | +   | +     | +   | +     |
| 169 | A/chicken/California/1316/2001(H6N2) | AAO33534         | +      | +   | +     | +   | +     |
| 170 | A/chicken/California/1316/2001(H6N2) | AAO33533         | +      | +   | +     | +   | +     |
| 171 | A/chicken/California/8892/2001(H6N2) | AAO33532         | +      | +   | +     | +   | +     |
| 172 | A/chicken/California/8892/2001(H6N2) | AAO33531         | +      | +   | +     | +   | +     |
| 173 | A/chicken/California/6643/2001(H6N2) | AAO33530         | +      | +   | +     | +   | +     |
| 174 | A/chicken/California/6643/2001(H6N2) | AAO33529         | +      | +   | +     | +   | +     |
| 175 | A/chicken/California/0139/2001(H6N2) | AAO33528         | +      | +   | +     | +   | +     |
| 176 | A/chicken/California/0139/2001(H6N2) | AAO33527         | +      | +   | +     | +   | +     |
| 177 | A/chicken/California/1002/2000(H6N2) | AAO33526         | +      | +   | +     | +   | +     |
| 178 | A/chicken/California/1002/2000(H6N2) | AAO33525         | +      | +   | +     | +   | +     |
| 179 | A/chicken/California/650/2000(H6N2) | AAO33524         | +      | +   | +     | +   | +     |
| 180 | A/chicken/California/650/2000(H6N2) | AAO33523         | +      | +   | +     | +   | +     |
| 181 | A/chicken/California/465/2000(H6N2) | AAO33522         | +      | +   | +     | +   | +     |
| 182 | A/chicken/California/465/2000(H6N2) | AAO33521         | +      | +   | +     | +   | +     |
| 183 | A/chicken/California/431/2000(H6N2) | AAO33520         | +      | +   | +     | +   | +     |
| 184 | A/chicken/California/431/2000(H6N2) | AAO33519         | +      | +   | +     | +   | +     |
| 185 | A/chicken/California/9420/2001(H6N2) | AAO33518         | +      | +   | +     | +   | +     |
| 186 | A/chicken/California/9420/2001(H6N2) | AAO33517         | +      | +   | +     | +   | +     |
| 187 | A/chicken/California/9174/2001(H6N2) | AAO33516         | +      | +   | +     | +   | +     |
| 188 | A/chicken/California/9174/2001(H6N2) | AAO33515         | +      | +   | +     | +   | +     |
| 189 | A/chicken/California/8892/2001(H6N2) | AAO33514         | +      | +   | +     | +   | +     |
| 190 | A/chicken/California/8892/2001(H6N2) | AAO33513         | +      | +   | +     | +   | +     |
| 191 | A/chicken/California/6643/2001(H6N2) | AAO33512         | +      | +   | +     | +   | +     |
| 192 | A/chicken/California/6643/2001(H6N2) | AAO33511         | +      | +   | +     | +   | +     |
| 193 | A/chicken/California/1316/2001(H6N2) | AAO33510         | +      | +   | +     | +   | +     |
| 194 | A/chicken/California/1316/2001(H6N2) | AAO33509         | +      | +   | +     | +   | +     |
| 195 | A/chicken/California/0139/2001(H6N2) | AAO33508         | +      | +   | +     | +   | +     |
| 196 | A/chicken/California/0139/2001(H6N2) | AAO33507         | +      | +   | +     | +   | +     |
| 197 | A/chicken/California/1002/2000(H6N2) | AAO33506         | +      | +   | +     | +   | +     |
| 198 | A/chicken/California/1002/2000(H6N2) | AAO33505         | +      | +   | +     | +   | +     |
| 199 | A/chicken/California/650/2000(H6N2) | AAO33504         | +      | +   | +     | +   | +     |
| 200 | A/chicken/California/650/2000(H6N2) | AAO33503         | +      | +   | +     | +   | +     |
| 201 | B/Memphis/12/97-MA             | AAP22120         | −      | −   | −     | −   | −     |
| 202 | B/Memphis/12/97-MA             | AAP22119         | −      | −   | −     | −   | −     |
| 203 | B/Memphis/12/97-MA             | AAP22118         | −      | −   | −     | −   | −     |
| 204 | B/Memphis/12/97-MA             | AAP22117         | −      | −   | −     | −   | −     |
| 205 | B/Memphis/12/97-MA             | AAP22116         | −      | −   | −     | −   | −     |
| 206 | B/Memphis/12/97-MA             | AAP22115         | −      | −   | −     | −   | −     |
| 207 | B/clinical isolate SA116 Philippines/2002 | AAY21787       | −      | −   | −     | −   | −     |
| 208 | B/clinical isolate SA115 Philippines/2002 | AAY21786       | −      | −   | −     | −   | −     |
| 209 | B/clinical isolate SA114 Philippines/2002 | AAY21785       | −      | −   | −     | −   | −     |
| 210 | B/clinical isolate SA113 Philippines/2002 | AAY21784       | −      | −   | −     | −   | −     |
| 211 | B/clinical isolate SA112 Philippines/2002 | AAY21783       | −      | −   | −     | −   | −     |
| 212 | B/clinical isolate SA110 Philippines/2002 | AAY21782       | −      | −   | −     | −   | −     |
| 213 | B/clinical isolate SA109 Philippines/2002 | AAY21781       | −      | −   | −     | −   | −     |
| 214 | B/clinical isolate SA108 Philippines/2002 | AAY21780       | −      | −   | −     | −   | −     |
| 215 | B/clinical isolate SA107 Philippines/2002 | AAY21779       | −      | −   | −     | −   | −     |
| 216 | B/clinical isolate SA106 Philippines/2002 | AAY21778       | −      | −   | −     | −   | −     |
| 217 | B/clinical isolate SA105 Philippines/2002 | AAY21777       | −      | −   | −     | −   | −     |
| 218 | B/clinical isolate SA104 Philippines/2002 | AAY21776       | −      | −   | −     | −   | −     |
| 219 | B/clinical isolate SA103 Philippines/2002 | AAY21775       | −      | −   | −     | −   | −     |
| 220 | B/clinical isolate SA102 Philippines/2002 | AAY21774       | −      | −   | −     | −   | −     |

(Continued)

(To be continued on the next page)
| No.  | Strains                               | Accession number | TypeLDA | LDA  | LDA* | SVM  | SVM* |
|------|---------------------------------------|------------------|---------|------|------|------|------|
| 221  | B/clinical isolate SA101 Philippines/2002 | AAY21773         |         |      |      |      |      |
| 222  | B/clinical isolate SA100 Philippines/2002 | AAY21772         |         |      |      |      |      |
| 223  | B/clinical isolate SA99 Philippines/2002 | AAY21771         |         |      |      |      |      |
| 224  | B/clinical isolate SA98 Philippines/2002 | AAY21770         |         |      |      |      |      |
| 225  | B/clinical isolate SA97 Philippines/2002 | AAY21769         |         |      |      |      |      |
| 226  | B/clinical isolate SA96 Thailand/2002  | AAY21768         |         |      |      |      |      |
| 227  | B/clinical isolate SA95 Philippines/2002 | AAY21767         |         |      |      |      |      |
| 228  | B/clinical isolate SA94 Thailand/2002  | AAY21766         |         |      |      |      |      |
| 229  | B/clinical isolate SA93 Thailand/2002  | AAY21765         |         |      |      |      |      |
| 230  | B/clinical isolate SA92 Thailand/2002  | AAY21764         |         |      |      |      |      |
| 231  | B/clinical isolate SA91 Thailand/2002  | AAY21763         |         |      |      |      |      |
| 232  | B/clinical isolate SA90 Thailand/2002  | AAY21762         |         |      |      |      |      |
| 233  | B/clinical isolate SA89 Thailand/2002  | AAY21761         |         |      |      |      |      |
| 234  | B/clinical isolate SA88 Thailand/2002  | AAY21760         |         |      |      |      |      |
| 235  | B/clinical isolate SA87 Thailand/2002  | AAY21759         |         |      |      |      |      |
| 236  | B/clinical isolate SA86 Thailand/2002  | AAY21758         |         |      |      |      |      |
| 237  | B/clinical isolate SA85 Thailand/2002  | AAY21757         |         |      |      |      |      |
| 238  | B/clinical isolate SA84 Philippines/2002 | AAY21756         |         |      |      |      |      |
| 239  | B/clinical isolate SA83 Philippines/2002 | AAY21755         |         |      |      |      |      |
| 240  | B/clinical isolate SA82 Philippines/2002 | AAY21754         |         |      |      |      |      |
| 241  | B/clinical isolate SA81 Philippines/2002 | AAY21753         |         |      |      |      |      |
| 242  | B/clinical isolate SA80 Philippines/2002 | AAY21752         |         |      |      |      |      |
| 243  | B/clinical isolate SA79 Philippines/2002 | AAY21751         |         |      |      |      |      |
| 244  | B/clinical isolate SA78 Philippines/2002 | AAY21750         |         |      |      |      |      |
| 245  | B/clinical isolate SA77 Philippines/2002 | AAY21749         |         |      |      |      |      |
| 246  | B/clinical isolate SA76 Philippines/2002 | AAY21748         |         |      |      |      |      |
| 247  | B/clinical isolate SA75 Philippines/2002 | AAY21747         |         |      |      |      |      |
| 248  | B/clinical isolate SA74 Philippines/2002 | AAY21746         |         |      |      |      |      |
| 249  | B/clinical isolate SA73 Philippines/2002 | AAY21745         |         |      |      |      |      |
| 250  | B/clinical isolate SA72 Philippines/2002 | AAY21744         |         |      |      |      |      |
| 251  | B/Yamagata/K320/2001                   | BAD29880         |         |      |      |      |      |
| 252  | B/Yamagata/K270/2001                   | BAD29879         |         |      |      |      |      |
| 253  | B/Yamagata/1311/2003                   | BAD29878         |         |      |      |      |      |
| 254  | B/Yamagata/1311/2003                   | BAD29877         |         |      |      |      |      |
| 255  | B/Yamagata/1246/2003                   | BAD29876         |         |      |      |      |      |
| 256  | B/Yamagata/1246/2003                   | BAD29875         |         |      |      |      |      |
| 257  | B/Yamagata/115/2003                    | BAD29874         |         |      |      |      |      |
| 258  | B/Yamagata/115/2003                    | BAD29873         |         |      |      |      |      |
| 259  | B/Yamagata/222/2002                    | BAD29872         |         |      |      |      |      |
| 260  | B/Yamagata/222/2002                    | BAD29871         |         |      |      |      |      |
| 261  | B/Yamagata/K542/2001                   | BAD29870         |         |      |      |      |      |
| 262  | B/Yamagata/K542/2001                   | BAD29869         |         |      |      |      |      |
| 263  | B/Yamagata/K535/2001                   | BAD29868         |         |      |      |      |      |
| 264  | B/Yamagata/K535/2001                   | BAD29867         |         |      |      |      |      |
| 265  | B/Yamagata/K521/2001                   | BAD29866         |         |      |      |      |      |
| 266  | B/Yamagata/K521/2001                   | BAD29865         |         |      |      |      |      |
| 267  | B/Yamagata/K520/2001                   | BAD29864         |         |      |      |      |      |
| 268  | B/Yamagata/K520/2001                   | BAD29863         |         |      |      |      |      |
| 269  | B/Yamagata/K519/2001                   | BAD29862         |         |      |      |      |      |
| 270  | B/Yamagata/K519/2001                   | BAD29861         |         |      |      |      |      |
| 271  | B/Yamagata/K515/2001                   | BAD29860         |         |      |      |      |      |
| 272  | B/Yamagata/K515/2001                   | BAD29859         |         |      |      |      |      |
| 273  | B/Yamagata/K508/2001                   | BAD29858         |         |      |      |      |      |
| 274  | B/Yamagata/K508/2001                   | BAD29857         |         |      |      |      |      |
| 275  | B/Yamagata/K501/2001                   | BAD29856         |         |      |      |      |      |

(To be continued on the next page)
| No. | Strains                  | Accession number | Typedef | LDA | LDA<sub>nc</sub> | SVM | SVM<sub>nc</sub> |
|-----|--------------------------|------------------|---------|-----|------------------|-----|-----------------|
| 276 | B/Yamagata/K501/2001     | BAD29855         | −       | −   | −                | −   | −               |
| 277 | B/Yamagata/K500/2001     | BAD29854         | −       | −   | −                | −   | −               |
| 278 | B/Yamagata/K500/2001     | BAD29853         | −       | −   | −                | −   | −               |
| 279 | B/Yamagata/K490/2001     | BAD29852         | −       | −   | −                | −   | −               |
| 280 | B/Yamagata/K490/2001     | BAD29851         | −       | −   | −                | −   | −               |
| 281 | B/Yamagata/K461/2001     | BAD29850         | −       | −   | −                | −   | −               |
| 282 | B/Yamagata/K461/2001     | BAD29849         | −       | −   | −                | −   | −               |
| 283 | B/Yamagata/K386/2001     | BAD29848         | −       | −   | −                | −   | −               |
| 284 | B/Yamagata/K386/2001     | BAD29847         | −       | −   | −                | −   | −               |
| 285 | B/Yamagata/K320/2001     | BAD29846         | −       | −   | −                | −   | −               |
| 286 | B/Yamagata/K320/2001     | BAD29845         | −       | −   | −                | −   | −               |
| 287 | B/Yamagata/K270/2001     | BAD29844         | −       | −   | −                | −   | −               |
| 288 | B/Yamagata/K270/2001     | BAD29843         | −       | −   | −                | −   | −               |
| 289 | B/Yamagata/1311/2003     | BAD29842         | −       | −   | −                | −   | −               |
| 290 | B/Yamagata/1246/2003     | BAD29841         | −       | −   | −                | −   | −               |
| 291 | B/Yamagata/115/2003      | BAD29840         | −       | −   | −                | −   | −               |
| 292 | B/Yamagata/222/2002      | BAD29839         | −       | −   | −                | −   | −               |
| 293 | B/Yamagata/K542/2001     | BAD29838         | −       | −   | −                | −   | −               |
| 294 | B/Yamagata/K535/2001     | BAD29837         | −       | −   | −                | −   | −               |
| 295 | B/Yamagata/K521/2001     | BAD29836         | −       | −   | −                | −   | −               |
| 296 | B/Yamagata/K520/2001     | BAD29835         | −       | −   | −                | −   | −               |
| 297 | B/Yamagata/K519/2001     | BAD29834         | −       | −   | −                | −   | −               |
| 298 | B/Yamagata/K515/2001     | BAD29833         | −       | −   | −                | −   | −               |
| 299 | B/Yamagata/K508/2001     | BAD29832         | −       | −   | −                | −   | −               |
| 300 | B/Yamagata/K501/2001     | BAD29831         | −       | −   | −                | −   | −               |
| 301 | CM2 protein [Influenza C virus] | YP. 089658 | −       | −   | −                | −   | −               |
| 302 | matrix protein [Influenza C virus] | YP. 089657 | −       | −   | −                | −   | −               |
| 303 | nucleoprotein [Influenza C virus] | YP. 089656 | −       | −   | −                | −   | −               |
| 304 | hemagglutinin-esterase precursor [Influenza C virus] | YP. 089655 | −       | −   | −                | −   | −               |
| 305 | polymerase 3 [Influenza C virus] | YP. 089654 | −       | −   | −                | −   | −               |
| 306 | polymerase 1 [Influenza C virus] | YP. 089653 | −       | −   | −                | −   | −               |
| 307 | polymerase 2 [Influenza C virus] | YP. 089652 | −       | −   | −                | −   | −               |
| 308 | NS1 [Influenza C virus]  | YP. 089651       | −       | −   | −                | −   | −               |
| 309 | NS2 [Influenza C virus]  | YP. 089650       | −       | −   | −                | −   | −               |
| 310 | hemagglutinin esterase [Influenza C virus] | BAA06094 | −       | −   | −                | −   | −               |
| 311 | hemagglutinin esterase [Influenza C virus] | BAA06093 | −       | −   | −                | −   | −               |
| 312 | hemagglutinin esterase [Influenza C virus] | BAA06092 | −       | −   | −                | −   | −               |
| 313 | hemagglutinin esterase [Influenza C virus] | BAA06096 | −       | −   | −                | −   | −               |
| 314 | hemagglutinin esterase [Influenza C virus] | BAA06095 | −       | −   | −                | −   | −               |
| 315 | NS1 [Influenza C virus]  | BAB12101         | −       | −   | −                | −   | −               |
| 316 | NS2 [Influenza C virus]  | BAB12100         | −       | −   | −                | −   | −               |
| 317 | NS1 [Influenza C virus]  | BAB12099         | −       | −   | −                | −   | −               |
| 318 | NS2 [Influenza C virus]  | BAB12098         | −       | −   | −                | −   | −               |
| 319 | NS1 [Influenza C virus]  | BAB12097         | −       | −   | −                | −   | −               |
| 320 | NS2 [Influenza C virus]  | BAB12096         | −       | −   | −                | −   | −               |
| 321 | NS1 [Influenza C virus]  | BAB12095         | −       | −   | −                | −   | −               |
| 322 | NS2 [Influenza C virus]  | BAB12094         | −       | −   | −                | −   | −               |
| 323 | NS1 [Influenza C virus]  | BAB12093         | −       | −   | −                | −   | −               |
| 324 | NS2 [Influenza C virus]  | BAB12092         | −       | −   | −                | −   | −               |
| 325 | NS1 [Influenza C virus]  | BAB12091         | −       | −   | −                | −   | −               |
| 326 | NS2 [Influenza C virus]  | BAB12090         | −       | −   | −                | −   | −               |
| 327 | NS1 [Influenza C virus]  | BAB12089         | −       | −   | −                | −   | −               |
| 328 | NS2 [Influenza C virus]  | BAB12088         | −       | −   | −                | −   | −               |
| 329 | NS1 [Influenza C virus]  | BAB12087         | −       | −   | −                | −   | −               |
| 330 | NS2 [Influenza C virus]  | BAB12086         | −       | −   | −                | −   | −               |
| No. | Strains                  | Accession number | Type <sub>int</sub> | LDA | LDA<sub>ax</sub> | SVM | SVM<sub>ax</sub> |
|-----|-------------------------|------------------|---------------------|-----|-----------------|-----|-----------------|
| 331 | NS1 [Influenza C virus] | BAB12085         |                     |     |                 |     |                 |
| 332 | NS2 [Influenza C virus] | BAB12084         |                     |     |                 |     |                 |
| 333 | NS1 [Influenza C virus] | BAB12083         |                     |     |                 |     |                 |
| 334 | NS2 [Influenza C virus] | BAB12082         |                     |     |                 |     |                 |
| 335 | NS1 [Influenza C virus] | BAB12081         |                     |     |                 |     |                 |
| 336 | NS2 [Influenza C virus] | BAB12080         |                     |     |                 |     |                 |
| 337 | NS1 [Influenza C virus] | BAB12079         |                     |     |                 |     |                 |
| 338 | NS2 [Influenza C virus] | BAB12078         |                     |     |                 |     |                 |
| 339 | NS1 [Influenza C virus] | BAB12077         |                     |     |                 |     |                 |
| 340 | NS2 [Influenza C virus] | BAB12076         |                     |     |                 |     |                 |
| 341 | NS1 [Influenza C virus] | BAB12075         |                     |     |                 |     |                 |
| 342 | NS2 [Influenza C virus] | BAB12074         |                     |     |                 |     |                 |
| 343 | NS1 [Influenza C virus] | BAB12073         |                     |     |                 |     |                 |
| 344 | NS2 [Influenza C virus] | BAB12072         |                     |     |                 |     |                 |
| 345 | NS1 [Influenza C virus] | BAB12071         |                     |     |                 |     |                 |
| 346 | NS2 [Influenza C virus] | BAB12070         |                     |     |                 |     |                 |
| 347 | NS1 [Influenza C virus] | BAB12069         |                     |     |                 |     |                 |
| 348 | NS2 [Influenza C virus] | BAB12068         |                     |     |                 |     |                 |
| 349 | NS1 [Influenza C virus] | BAB12067         |                     |     |                 |     |                 |
| 350 | NS2 [Influenza C virus] | BAB12066         |                     |     |                 |     |                 |
| 351 | C/Saitama/3/2000        | BAD37028         |                     |     |                 |     |                 |
| 352 | C/Saitama/2/2000        | BAD37027         |                     |     |                 |     |                 |
| 353 | C/Saitama/1/2000        | BAD37026         |                     |     |                 |     |                 |
| 354 | C/Miyagi/5/2000         | BAD37025         |                     |     |                 |     |                 |
| 355 | C/Miyagi/4/2000         | BAD37023         |                     |     |                 |     |                 |
| 356 | C/Miyagi/3/2000         | BAD37022         |                     |     |                 |     |                 |
| 357 | C/Miyagi/2/2000         | BAD37021         |                     |     |                 |     |                 |
| 358 | C/Yamagata/8/2000       | BAD37020         |                     |     |                 |     |                 |
| 359 | C/Yamagata/6/2000       | BAD37019         |                     |     |                 |     |                 |
| 360 | C/Yamagata/3/2000       | BAD37018         |                     |     |                 |     |                 |
| 361 | C/Yamagata/2/2000       | BAD37017         |                     |     |                 |     |                 |
| 362 | C/Yamagata/9/2000       | BAD37016         |                     |     |                 |     |                 |
| 363 | C/Yamagata/2/99         | BAD37015         |                     |     |                 |     |                 |
| 364 | C/Greece/79             | BAD37014         |                     |     |                 |     |                 |
| 365 | C/Kansas/1/79           | BAD37013         |                     |     |                 |     |                 |
| 366 | C/California/78         | BAD37012         |                     |     |                 |     |                 |
| 367 | C/NewJersey/76          | BAD37011         |                     |     |                 |     |                 |
| 368 | C/Hiroshima/251/2000    | BAD37010         |                     |     |                 |     |                 |
| 369 | C/Hiroshima/250/2000    | BAD37009         |                     |     |                 |     |                 |
| 370 | C/Hiroshima/248/2000    | BAD37008         |                     |     |                 |     |                 |
| 371 | C/Hiroshima/247/2000    | BAD37007         |                     |     |                 |     |                 |
| 372 | C/Hiroshima/246/2000    | BAD37006         |                     |     |                 |     |                 |
| 373 | C/Hiroshima/249/2000    | BAD37005         |                     |     |                 |     |                 |
| 374 | C/Hiroshima/252/99      | BAD37004         |                     |     |                 |     |                 |
| 375 | C/Hiroshima/290/99      | BAD37003         |                     |     |                 |     |                 |
| 376 | C/Saitama/3/2000        | BAD37002         |                     |     |                 |     |                 |
| 377 | C/Saitama/2/2000        | BAD37001         |                     |     |                 |     |                 |
| 378 | C/Saitama/1/2000        | BAD37000         |                     |     |                 |     |                 |
| 379 | C/Miyagi/5/2000         | BAD36999         |                     |     |                 |     |                 |
| 380 | C/Miyagi/4/2000         | BAD36998         |                     |     |                 |     |                 |
| 381 | C/Miyagi/3/2000         | BAD36997         |                     |     |                 |     |                 |
| 382 | C/Miyagi/2/2000         | BAD36996         |                     |     |                 |     |                 |
| 383 | C/Yamagata/8/2000       | BAD36995         |                     |     |                 |     |                 |
| 384 | C/Yamagata/6/2000       | BAD36994         |                     |     |                 |     |                 |
| 385 | C/Yamagata/3/2000       | BAD36993         |                     |     |                 |     |                 |

(Continued)

(To be continued on the next page)
| No. | Strains            | Accession number | Type_{init} | LDA | LDA_{cv} | SVM | SVM_{cv} |
|-----|--------------------|------------------|------------|-----|----------|-----|----------|
| 386 | C/Yamagata/2/2000  | BAD36992         | -          | -   | -        | -   | -        |
| 387 | C/Yamagata/9/2000  | BAD36991         | -          | -   | -        | -   | -        |
| 388 | C/Yamagata/2/99    | BAD36990         | -          | -   | -        | -   | -        |
| 389 | C/Greece/79        | BAD36989         | -          | -   | -        | -   | -        |
| 390 | C/Kansas/1/79      | BAD36988         | -          | -   | -        | -   | -        |
| 391 | C/California/78    | BAD36987         | -          | -   | -        | -   | -        |
| 392 | C/NewJersey/76     | BAD36986         | -          | -   | -        | -   | -        |
| 393 | C/Hiroshima/251/2000 | BAD36985   | -          | -   | -        | -   | -        |
| 394 | C/Hiroshima/250/2000 | BAD36984  | -          | -   | -        | -   | -        |
| 395 | C/Hiroshima/248/2000 | BAD36983   | -          | -   | -        | -   | -        |
| 396 | C/Hiroshima/247/2000 | BAD36982  | -          | -   | -        | -   | -        |
| 397 | C/Hiroshima/246/2000 | BAD36981  | -          | -   | -        | -   | -        |
| 398 | C/Hiroshima/249/2000 | BAD36980   | -          | -   | -        | -   | -        |
| 399 | C/Hiroshima/252/99 | BAD36979        | -          | -   | -        | -   | -        |
| 400 | C/Hiroshima/290/99 | BAD36978        | -          | -   | -        | -   | -        |

\( \text{a) } ^{+}\text{ and } ^{-}\text{ represent AIVs and the non-AIVs, respectively. LDA, SVM, LDA_{cv}, and SVM_{cv} represent the method with simulative modeling, leave-one-out cross validation for LDA and SVM, respectively.} \)

\begin{table}
\centering
\begin{tabular}{|l|l|l|l|}
\hline
No. & Description of variables & Resources \\
\hline
1 & molecular weight & \\
2 & times of charged amino acids & \\
3 & frequency of charged amino acids & \\
4 & quality percentage of charged amino acids & \\
5 & times of acidic amino acids & \\
6 & frequency of acidic amino acids & \\
7 & quality percentage of acidic amino acid & \\
8 & times of basic amino acids & \\
9 & frequency of basic amino acids & \\
10 & quality percentage of basic amino acids & \\
11 & times of polar amino acids & \\
12 & frequency of polar amino acids & \\
13 & quality percentage of polar amino acids & \\
14 & times of hydrophobic amino acids & \\
15 & frequency of hydrophobic amino acids & \\
16 & quality percentage of hydrophobic amino acids & \\
17 & times of A & \\
18 & frequency of A & \\
19 & quality percentage of A & \\
20 & times of R & \\
21 & frequency of R & \\
22 & quality percentage of R & \\
23 & times of N & \\
24 & frequency of N & \\
25 & quality percentage of N & \\
26 & times of D & \\
27 & frequency of D & \\
28 & quality percentage of D & \\
29 & times of C & \\
30 & frequency of C & \\
31 & quality percentage of C & \\
32 & times of Q & \\
33 & frequency of Q & \\
34 & quality percentage of Q & \\
35 & times of E & \\
\hline
\end{tabular}
\end{table}

(To be continued on the next page)
| No. | Description of variables                                      | Resources |
|-----|--------------------------------------------------------------|-----------|
| 36  | frequency of E                                              |           |
| 37  | quality percentage of E                                     |           |
| 38  | times of G                                                  |           |
| 39  | frequency of G                                              |           |
| 40  | quality percentage of G                                     |           |
| 41  | times of H                                                  |           |
| 42  | frequency of H                                              |           |
| 43  | quality percentage of H                                     |           |
| 44  | times of I                                                  |           |
| 45  | frequency of I                                              |           |
| 46  | quality percentage of I                                     |           |
| 47  | times of L                                                  |           |
| 48  | frequency of L                                              |           |
| 49  | quality percentage of L                                     |           |
| 50  | times of K                                                  |           |
| 51  | frequency of K                                              |           |
| 52  | quality percentage of K                                     |           |
| 53  | times of M                                                  |           |
| 54  | frequency of M                                              |           |
| 55  | quality percentage of M                                     |           |
| 56  | times of F                                                  |           |
| 57  | frequency of F                                              |           |
| 58  | quality percentage of F                                     |           |
| 59  | times of P                                                  |           |
| 60  | frequency of P                                              |           |
| 61  | quality percentage of P                                     |           |
| 62  | times of S                                                  |           |
| 63  | frequency of S                                              |           |
| 64  | quality percentage of S                                     |           |
| 65  | times of T                                                  |           |
| 66  | frequency of T                                              |           |
| 67  | quality percentage of T                                     |           |
| 68  | time of W                                                   |           |
| 69  | frequency of W                                              |           |
| 70  | quality percentage of W                                     |           |
| 71  | times of Y                                                  |           |
| 72  | frequency of Y                                              |           |
| 73  | quality percentage of Y                                     |           |
| 74  | times of V                                                  |           |
| 75  | frequency of V                                              |           |
| 76  | quality percentage of V                                     |           |
| 77  | Retention coefficient in TFA                                 | BROC820101¹¹ |
| 78  | Free energy of solution in water                            | CHAM820102 |
| 79  | Solvation free energy                                       | EISD860101 |
| 80  | Melting point                                               | FASG760102 |
| 81  | Number of hydrogen bond donors                              | FAUJ880109 |
| 82  | Number of full nonbonding orbitals                          | FAUJ880110 |
| 83  | Partition energy                                            | GUYH850101 |
| 84  | Hydration number                                            | HOPA770101 |
| 85  | Retention coefficient in HPLC, pH7.4                        | MEEJ800101 |
| 86  | Retention coefficient in HPLC, pH2.1                        | MEEJ800102 |
| 87  | Partition coefficient in thin-layer chromatography           | PLIV810101 |
| 88  | Retention coefficient at pH 2                               | GUOD860101 |
| 89  | $R_f$ for 1-N-(4-nitrobenzofurazono)-amino acids in ethyl acetate/pyridine/water | –²³ |
| 90  | $\Delta G$ of transfer from organic solvent to water        | –          |
| No. | Description of variables                                                                 | Resources          |
|-----|------------------------------------------------------------------------------------------|--------------------|
| 91  | Hydration potential or free energy of transfer from vapor phase to water                  |                    |
| 92  | $R_s$, salt chromatography                                                                |                    |
| 93  | log $D_r$, partition coefficient at pH 7.1 for acetylamide derivatives of amino acids in octanol/water |                    |
| 94  | $dG = RT \ln f$, $f$ = fraction buried/accessible amino acids in 22 proteins              |                    |
| 95  | Average volume of buried residue                                                          | CHOC750101         |
| 96  | Residue accessible surface area in tripeptide                                            | CHOC760101         |
| 97  | Graph shape index                                                                        | FAUJ880101         |
| 98  | Normalized van der Waals volume                                                          | FAUJ880103         |
| 99  | STERIMOL length of the side chain                                                        | FAUJ880104         |
| 100 | STERIMOL minimum width of the side chain                                                 | FAUJ880105         |
| 101 | STERIMOL maximum width of the side chain                                                 | FAUJ880106         |
| 102 | Average accessible surface area                                                          | JANJ780101         |
| 103 | Distance between C-alpha and centroid of side chain                                       | LEVM760102         |
| 104 | Side chain angle theta                                                                   | LEVM760103         |
| 105 | Side chain torsion angle phi                                                             | LEVM760104         |
| 106 | Radius of gyration of side chain                                                         | LEVM760105         |
| 107 | van der Waals parameter $R_0$                                                            | LEVM760106         |
| 108 | van der Waals parameter epsilon                                                          | LEVM760107         |
| 109 | Refractivity                                                                             | MCMT640101         |
| 110 | Value of theta ($i$)                                                                     | RACS820113         |
| 111 | Substituent van der Waals volume                                                         |                    |
| 112 | Alpha-CH chemical shifts                                                                 | ANDN920101         |
| 113 | Alpha-NH chemical shifts                                                                 | BUNA790101         |
| 114 | A parameter of charge transfer capability                                                | CHAM830107         |
| 115 | A parameter of charge transfer donor capability                                          | CHAM830108         |
| 116 | N M R chemical shift of alpha-carbon                                                     | FAUJ880107         |
| 117 | Localized electrical effect                                                              | FAUJ880108         |
| 118 | Positive charge                                                                          | FAUJ880111         |
| 119 | Negative charge                                                                         | FAUJ880112         |
| 120 | Polarity                                                                                | GRAR740102         |
| 121 | Net charge                                                                              | KLEP840101         |
| 122 | Amphilicity index                                                                        | MITS020101         |
| 123 | Isoelectric point                                                                        | ZIMJ680104         |
| 124 | Electron-ion interaction potential values                                                | COSI940101         |
| 125 | pKNH$_2$(NH$_2$ on C-alpha)                                                               | FASG760104         |
| 126 | pKCOOH(COOH on C-alpha)                                                                   | FASG760105         |
| 127 | Depth                                                                                   | c)                  |
| 128 | log$P$ value                                                                            | b)                  |
| 129 | Inorganic(I) value                                                                       | b)                  |
| 130 | Organic(O) value                                                                         | b)                  |
| 131 | Residue electronegativity                                                                | b)                  |
| 132 | Polarizability                                                                          | e)                  |
| 133 | Hydrophobicity                                                                          | e)                  |
| 134 | Helix probability                                                                        | e)                  |
| 135 | Sheet probability                                                                        | e)                  |
| 136 | ECI                                                                                    | f)                  |
| 137 | ISA                                                                                    | f)                  |
| 138 | Specific volume $c$/mL·g$^{-1}$                                                           | g)                  |
| 139 | Hydrophobic parameter 1 ($K$-$D$)                                                        | h)                  |

(To be continued on the next page)
| No. | Description of variables                                                                 | Resources |
|-----|------------------------------------------------------------------------------------------|-----------|
| 140 | hydrophobic parameter 2                                                                  | h)        |
| 141 | hydrophobic parameter 3                                                                  | h)        |
| 142 | hydrophobic parameter 4 \((F-P)\)                                                        | h)        |
| 143 | Volume \(\text{Å}^3\)                                                                   | f)        |
| 144 | Sum of atomic van der Waals volumes (scaled on carbon atom)                              | i)        |
| 145 | Sum of atomic Sanderson electronegativities (scaled on carbon atom)                      | /         |
| 146 | Sum of atomic polarizabilities (scaled on carbon atom)                                  | /         |
| 147 | Sum of Kier-Hall electrotopological states                                               | /         |
| 148 | Mean atomic van der Waals volume (scaled on carbon atom)                                | /         |
| 149 | Mean atomic Sanderson electronegativity (scaled on carbon atom)                          | /         |
| 150 | Mean atomic polarizability (scaled on carbon atom)                                      | /         |
| 151 | Mean electrotopological state                                                            | /         |
| 152 | Aromatic ratio                                                                          | /         |
| 153 | Total structure connectivity index                                                       | /         |
| 154 | Polarity number                                                                         | /         |
| 155 | Wiener W index                                                                          | /         |
| 156 | All-path Wiener index                                                                   | /         |
| 157 | Balaban distance connectivity index                                                       | /         |
| 158 | Maximal electrotopological negative variation                                            | /         |
| 159 | Maximal electrotopological positive variation                                            | /         |
| 160 | Molecular electrotopological variation                                                   | /         |
| 161 | E-state topological parameter                                                            | /         |
| 162 | Kier symmetry index                                                                     | /         |
| 163 | Kier flexibility index                                                                   | /         |
| 164 | Unipolarity                                                                             | /         |
| 165 | Balaban centric index                                                                   | /         |
| 166 | Lopping centric index                                                                   | /         |
| 167 | Radial centric information index                                                         | /         |
| 168 | Modified Randic connectivity index                                                       | /         |
| 169 | Information index on molecular size                                                     | /         |
| 170 | Graph distance complexity index (log)                                                    | /         |
| 171 | 3D-Wiener index                                                                         | /         |
| 172 | 3D-Balaban index                                                                        | /         |
| 173 | 3D-Harary index                                                                         | /         |
| 174 | Average distance/distance degree                                                        | /         |
| 175 | Gravitational index G1                                                                   | /         |
| 176 | Radius of gyration (mass weighted)                                                      | /         |
| 177 | Span \(R\)                                                                              | /         |
| 178 | Molecular eccentricity                                                                  | /         |
| 179 | Spherosity                                                                              | /         |
| 180 | Asphericity                                                                             | /         |
| 181 | Folding degree index                                                                    | /         |
| 182 | 3D Petitjean shape index                                                                 | /         |
| 183 | Absolute eigenvalue sum on geometry matrix                                               | /         |
| 184 | Aromaticity index                                                                       | /         |
| 185 | HOMA total                                                                              | /         |
| 186 | Sum of geometrical distances between N...N                                              | /         |
| 187 | Sum of geometrical distances between N...O                                              | /         |

(To be continued on the next page)
No. | Description of variables | Resources
---|--------------------------|--------------
188 | Sum of geometrical distances between O..O | / 
189 | Number of total primary C(sp3) | / 
190 | Number of total secondary C(sp3) | / 
191 | H attached to heteroatom | / 
192 | Phenol/enol/ carboxyl OH | / 
193 | The number of O== | / 
194 | The number of R—SH | / 
195 | Unsaturation index | / 
196 | Hydrophilic factor | / 
197 | Ghose-Crippen molar refractivity | / 
198 | Fragment-based polar surface area | / 
199 | Moriguchi octanol-water partition coefficient (logP) | / 
200 | Ghose-Crippen octanol-water partition coefficient (logP) | / 

a) Kawashima S, Kanehisa M. AAindex: amino acid index database. Nucleic Acids Res, 2000, 28: 374. Tomii K, Kanehisa M. Analysis of amino acid indices and mutation matrices for sequence comparison and structure prediction of proteins. Protein Eng, 1996, 9: 27—36. Nakai K, Kidera A, Kanehisa M. Cluster analysis of amino acid indices for prediction of protein structure and function. Protein Eng, 1988, 2: 93—100
b) Hellberg S, Sjöström M, Skagerberg B, Word S. Peptide quantitative structure-activity relationships, a multivariate approach. J Med Chem, 1987, 30: 1126—1135
c) Pintar A, Carugo O, Pongor S. Atom depth in protein structure and function. Trends in Biochem Sci, 2003, 28(11): 593—597
d) Wu H Q. Chemical property calculation through javascript and applications in QSAR. Molecules, 1999, 4: 16—27
e) Meiler J, Müller M, Zeidler A, Schmaeschke F. Generation and evaluation of dimension-reduced amino acid parameter representations by artificial neural networks. J Mol Model, 2001(7): 360—369
f) Collantes E R, Dunn III W J. Amino acid side chain descriptors for quantitative structure-activity relationship studies of peptide analogues. J Med Chem, 1995, 38: 2705—2713
g) Creighton T E. Protein Molecule, New York: WH Freeman and Co, 1982
h) Thorton J, Taylor W R. Structure Prediction. In: Findlay J B C, Geisow M J eds. Protein Aequencing. Oxford: IRL Press, 1989
i) Riesi M (President and Chief Executive Officer), Todeschini Ed, Dragon & Mobydigs evaluation versions of software developed by the Milano Chemometrics and QSAR Research Group, Web site: http://www.talete.mi.it/
| No. | Strains | Accession number | Type<sub>orig</sub> | LDA<sub>ext</sub> | SVM<sub>ext</sub> |
|-----|---------|------------------|-------------------|----------------|----------------|
| 1   | A/Hong Kong/532/97(H5N1) | AAK49246         | +                 | +              | +              |
| 2   | A/Hong Kong/532/97(H5N1) | AAK49245         | +                 | +              | +              |
| 3   | A/Hong Kong/486/97(H5N1) | AAK49244         | +                 | +              | +              |
| 4   | A/Hong Kong/486/97(H5N1) | AAK49243         | +                 | +              | +              |
| 5   | A/Hong Kong/483/97(H5N1) | AAK49242         | +                 | +              | +              |
| 6   | A/Hong Kong/483/97(H5N1) | AAK49241         | +                 | +              | +              |
| 7   | A/Hong Kong/482/97(H5N1) | AAK49240         | +                 | +              | +              |
| 8   | A/Hong Kong/482/97(H5N1) | AAK49239         | +                 | +              | +              |
| 9   | A/Hong Kong/481/97(H5N1) | AAK49238         | +                 | +              | +              |
| 10  | A/Hong Kong/481/97(H5N1) | AAK49237         | +                 | +              | +              |
| 11  | A/Hong Kong/1074/99(H9N2) | AAK49236         | +                 | +              | +              |
| 12  | A/Hong Kong/1074/99(H9N2) | AAK49235         | +                 | +              | +              |
| 13  | A/Hong Kong/1073/99(H9N2) | AAK49234         | +                 | +              | +              |
| 14  | A/Hong Kong/1073/99(H9N2) | AAK49233         | +                 | +              | +              |
| 15  | A/chicken/Puebla/8623-607/94(H5N2) | AAR91547    | +                 | –              | –              |
| 16  | A/chicken/Puebla/8623-607/94(H5N2) | AAR91539    | +                 | +              | +              |
| 17  | A/chicken/Puebla/8623-607/94(H5N2) | AAR91531    | +                 | +              | +              |
| 18  | A/chicken/Puebla/8623-607/94(H5N2) | AAR91523    | +                 | +              | +              |
| 19  | A/chicken/Puebla/8623-607/94(H5N2) | AAR91515    | +                 | –              | –              |
| 20  | A/chicken/Puebla/8623-607/94(H5N2) | AAR91507    | +                 | –              | –              |
| 21  | A/chicken/Mexico/3181-7/94 (H5N2) | AAR91506    | +                 | –              | –              |
| 22  | A/chicken/British Columbia/CN6/04(H7N3) | AAS68158   | +                 | +              | +              |
| 23  | A/chicken/Guatemala/45511-1/00 (H5N2) | AAR88833   | +                 | +              | +              |
| 24  | A/chicken/China(H9N2) | AAT70836      | +                 | +              | +              |
| 25  | A/chicken/Mudanjiang/0823/00 (H9N2) | AAF51706   | +                 | +              | +              |
| 26  | A/parrot/Northern Ireland/VF-73-67/73(H7N1) | AAG10676   | +                 | +              | +              |
| 27  | A/parakeet/Netherlands/267497/94(H7N1) | AAG10675   | +                 | +              | +              |
| 28  | A/macaw/England/626/80(H7N7) | AAG10674   | +                 | +              | +              |
| 29  | A/conure/England/766/94(H7N1) | AAG10673   | +                 | +              | +              |
| 30  | A/gull/Italy/692-2/93(H7N2) | AAG10672   | +                 | +              | +              |
| 31  | A/turkey/England/647/77(H7N7) | AAG10671   | +                 | +              | +              |
| 32  | A/turkey/Northern Ireland/VF-1545 CS/98(H7N7) | AAG10670   | +                 | +              | +              |
| 33  | A/turkey/England/192-328/79(H7N3) | AAG10669   | +                 | +              | +              |
| 34  | A/parrot/South Africa/1069/91(H7N1) | AAG10668   | +                 | +              | +              |
| 35  | A/parrot/England/1174/94(H7N1) | AAG10667   | +                 | +              | +              |
| 36  | A/psittacine/Italy/1/91(H7N2) | AAG10666   | +                 | +              | +              |
| 37  | A/conure/England/1234/94(H7N1) | AAG10665   | +                 | +              | +              |
| 38  | A/non-psittacine/England/Q-1985/89(H7N7) | AAG10664   | +                 | +              | +              |
| 39  | A/chicken/Ireland/1733/89(H7N7) | AAG10663   | +                 | +              | +              |
| 40  | A/turkey/England/63/97(H7N3) | AAG10662   | +                 | +              | +              |
| 41  | A/FPV/Egypt/45(H7N1) | AAG10661    | +                 | –              | –              |
| 42  | A/chicken/England/71/82(H7N1) | AAG10660   | +                 | +              | +              |
| 43  | A/turkey/Israel/Ramon/79(H7N2) | AAG10659   | +                 | +              | +              |
| 44  | A/parrot/Zimbabwe/222/96(H7N1) | AAG10658   | +                 | +              | +              |
| 45  | A/chicken/Pakistan/16/99/95(H7N3) | AAG10657   | +                 | +              | +              |
| 46  | A/chicken/Queensland/667/95(H7N3) | AAG10655   | +                 | –              | –              |
| 47  | A/chicken/Pakistan/CR2/95(H7N3) | AAG10654   | +                 | +              | +              |
| 48  | A/fairy bluebird/Singapore/F92/94(H7N1) | AAG10653   | +                 | +              | +              |
| 49  | A/common iora /Singapore/F89/95(H7N1) | AAG10652   | +                 | +              | +              |
| 50  | A/chicken/Victoria/192/92(H7N3) | AAG10651   | +                 | –              | –              |
| 51  | A/crow/Bangkok/Thailand/CU-35/04(H5N1) | AAZ29971   | +                 | +              | +              |
| 52  | A/parrot/Samut Prakan/Thailand/CU-31/04(H5N1) | AAZ29970 | +                 | +              | +              |
| 53  | A/white peafowl/Bangkok/Thailand/CU-29/04(H5N1) | AAZ29969 | +                 | +              | +              |
| 54  | A/chicken/Saraburi/Thailand/CU-27/04(H5N1) | AAZ29968 | +                 | +              | +              |
| 55  | A/rollers/Bangkok/Thailand/CU-26/04(H5N1) | AAZ29967 | +                 | +              | +              |

*(To be continued on the next page)*
| No. | Strains                                      | Accession number | Type<sub>ext</sub> | LDA<sub>ext</sub> | SVM<sub>ext</sub> |
|-----|---------------------------------------------|------------------|---------------------|-------------------|------------------|
| 56  | A/crow/Bangkok/Thailand/CU-25/04(H5N1)      | AAZ29966         | +                   | +                 | +                |
| 57  | A/chicken/Ayutthaya/Thailand/CU-24/04(H5N1) | AAZ29965         | +                   | +                 | +                |
| 58  | A/chicken/Bangkok/Thailand/CU-20/04(H5N1)   | AAZ29964         | +                   | +                 | +                |
| 59  | A/ostrich/Samut Prakan/Thailand/CU-19/04(H5N1) | AAZ29963   | +                   | +                 | +                |
| 60  | A/kalij pheasant/Bangkok/Thailand/CU-18/04(H5N1) | AAZ29962   | +                   | +                 | +                |
| 61  | A/chicken/Saraburi/Thailand/CU-17/04(H5N1)  | AAZ29961         | +                   | +                 | +                |
| 62  | A/white peafowl/Bangkok/Thailand/CU-16/04(H5N1) | AAZ29960   | +                   | +                 | +                |
| 63  | A/crow/Bangkok/Thailand/CU-15/04(H5N1)      | AAZ29959         | +                   | +                 | +                |
| 64  | A/chicken/Nakhon Pathom/Thailand/CU-14/04(H5N1) | AAZ29958   | +                   | +                 | +                |
| 65  | A/chicken/Nakhon Sawan/Thailand/CU-13/04(H5N1) | AAZ29957   | +                   | +                 | +                |
| 66  | A/chicken/Nakhon Sawan/Thailand/CU-12/04(H5N1) | AAZ29956   | +                   | +                 | +                |
| 67  | A/chicken/Chachoengsao/Thailand/CU-11/04(H5N1) | AAZ29955   | +                   | +                 | +                |
| 68  | A/chicken/Chachoengsao/Thailand/CU-10/04(H5N1) | AAZ29954   | +                   | +                 | +                |
| 69  | A/chicken/Suphanburi/Thailand/CU-9/04(H5N1)  | AAZ29953         | +                   | +                 | +                |
| 70  | A/chicken/Prachinburi/Thailand/CU-8/04(H5N1) | AAZ29952         | +                   | +                 | +                |
| 71  | A/chicken/Chonburi/Thailand/CU-7/04(H5N1)   | AAZ29951         | +                   | +                 | +                |
| 72  | A/chicken/Bangkok/Thailand/CU-6/04(H5N1)    | AAZ29950         | +                   | +                 | +                |
| 73  | A/chicken/Bangkok/Thailand/CU-3/04(H5N1)    | AAZ29947         | +                   | +                 | +                |
| 74  | A/chicken/Suphanburi/Thailand/CU-1/04(H5N1)  | AAZ29946         | +                   | +                 | +                |
| 75  | A/ostrich/HeNan/14/2002(H5N1)               | AAZ29456         | +                   | +                 | +                |
| 76  | A/duck/Korea/S13/03(H9N2)                   | AAW78058         | +                   | +                 | +                |
| 77  | A/chicken/Korea/S12/03(H9N2)                | AAW78057         | +                   | +                 | +                |
| 78  | A/chicken/Korea/S5/03(H9N2)                 | AAW78056         | +                   | +                 | +                |
| 79  | A/chicken/Korea/S4/03(H9N2)                 | AAW78055         | +                   | +                 | +                |
| 80  | A/silky chicken/Korea/S3/03(H9N2)           | AAW78054         | +                   | +                 | +                |
| 81  | A/duck/Korea/S17/03(H6N1)                   | AAW78053         | +                   | +                 | +                |
| 82  | A/dove/Korea/S11/03(H5N2)                   | AAW78052         | +                   | +                 | +                |
| 83  | A/duck/Korea/S10/03(H3N2)                   | AAW78051         | +                   | +                 | +                |
| 84  | A/duck/Korea/S9/03(H3N2)                    | AAW78050         | +                   | +                 | +                |
| 85  | A/duck/Korea/S8/03(H3N2)                    | AAW78049         | +                   | +                 | +                |
| 86  | A/duck/Korea/S7/03(H3N2)                    | AAW78048         | +                   | +                 | +                |
| 87  | A/chicken/Korea/S6/03(H3N2)                 | AAW78047         | +                   | +                 | +                |
| 88  | A/chicken/Korea/S18/03(H9N2)                | AAW78046         | +                   | +                 | +                |
| 89  | A/chicken/Korea/S16/03(H9N2)                | AAW78045         | +                   | +                 | +                |
| 90  | A/chicken/Korea/S15/03(H9N2)                | AAW78044         | +                   | +                 | +                |
| 91  | A/dove/Korea/S14/03(H9N2)                   | AAW78043         | +                   | +                 | +                |
| 92  | A/duck/Korea/S13/03(H9N2)                   | AAW78042         | +                   | +                 | +                |
| 93  | A/chicken/Korea/S12/03(H9N2)                | AAW78041         | +                   | +                 | +                |
| 94  | A/chicken/Korea/S5/03(H9N2)                 | AAW78040         | +                   | +                 | +                |
| 95  | A/chicken/Korea/S4/03(H9N2)                 | AAW78039         | +                   | +                 | +                |
| 96  | A/silky chicken/Korea/S3/03(H9N2)           | AAW78038         | +                   | +                 | +                |
| 97  | A/chicken/Viet Nam/LA-024/2004(H5N1)        | AAW59560         | +                   | –                 | –                |
| 98  | A/chicken/Viet Nam/LA-024/2004(H5N1)        | AAW59559         | +                   | +                 | +                |
| 99  | A/chicken/Viet Nam/CT-018/2004(H5N1)        | AAW59558         | +                   | +                 | +                |
| 100 | A/chicken/Viet Nam/DT-015/2004(H5N1)        | AAW59557         | +                   | –                 | –                |
| 101 | B/clinical isolate SA69 Philippines/2002    | AAY21743         | –                   | –                 | –                |
| 102 | B/clinical isolate SA68 Philippines/2002    | AAY21742         | –                   | –                 | –                |
| 103 | B/clinical isolate SA67 Philippines/2002    | AAY21741         | –                   | –                 | –                |
| 104 | B/clinical isolate SA66 Philippines/2002    | AAY21740         | –                   | –                 | –                |
| 105 | B/clinical isolate SA65 Philippines/2002    | AAY21739         | –                   | –                 | –                |
| 106 | B/clinical isolate SA64 Philippines/2002    | AAY21738         | –                   | –                 | –                |
| 107 | B/clinical isolate SA63 Philippines/2002    | AAY21737         | –                   | –                 | –                |
| 108 | B/clinical isolate SA62 Philippines/2002    | AAY21736         | –                   | –                 | –                |
| 109 | B/clinical isolate SA61 Philippines/2002    | AAY21735         | –                   | –                 | –                |
| 110 | B/clinical isolate SA60 Philippines/2002    | AAY21734         | –                   | –                 | –                |

(Continued)
| No. | Strains                      | Accession number | Type ($\text{tag}$) | LDA$_{out}$ | SVM$_{out}$ |
|-----|------------------------------|------------------|---------------------|-------------|-------------|
| 111 | B/clinical isolate SA59 Philippines/2002 | AAY21733         | −                   | −           | −           |
| 112 | B/clinical isolate SA58 Philippines/2002 | AAY21732         | −                   | −           | −           |
| 113 | B/clinical isolate SA57 Philippines/2002 | AAY21731         | −                   | −           | −           |
| 114 | B/clinical isolate SA53 Philippines/2002 | AAY21730         | −                   | −           | −           |
| 115 | B/clinical isolate SA52 Philippines/2002 | AAY21729         | −                   | −           | −           |
| 116 | B/clinical isolate SA51 Philippines/2002 | AAY21728         | −                   | −           | −           |
| 117 | B/clinical isolate SA50 Philippines/2002 | AAY21727         | −                   | −           | −           |
| 118 | B/clinical isolate SA47 Philippines/2002 | AAY21726         | −                   | −           | −           |
| 119 | B/clinical isolate SA46 Philippines/2002 | AAY21725         | −                   | −           | −           |
| 120 | B/clinical isolate SA45 Philippines/2002 | AAY21724         | −                   | −           | −           |
| 121 | B/clinical isolate SA44 Thailand/2002 | AAY21723         | −                   | −           | −           |
| 122 | B/clinical isolate SA43 Thailand/2002 | AAY21722         | −                   | −           | −           |
| 123 | B/clinical isolate SA42 Philippines/2002 | AAY21721         | −                   | −           | −           |
| 124 | B/clinical isolate SA41 Philippines/2002 | AAY21720         | −                   | −           | −           |
| 125 | B/clinical isolate SA40 Thailand/2002 | AAY21719         | −                   | −           | −           |
| 126 | B/Yamagata/K50/2001 | BAD29830         | −                   | −           | −           |
| 127 | B/Yamagata/K490/2001 | BAD29829         | −                   | −           | −           |
| 128 | B/Yamagata/K461/2001 | BAD29828         | −                   | −           | −           |
| 129 | B/Yamagata/K386/2001 | BAD29827         | −                   | −           | −           |
| 130 | B/Yamagata/K320/2001 | BAD29826         | −                   | −           | −           |
| 131 | B/Yamagata/K270/2001 | BAD29825         | −                   | −           | −           |
| 132 | B/Yamagata/1311/2003 | BAD29824         | −                   | −           | −           |
| 133 | B/Yamagata/1311/2003 | BAD29823         | −                   | −           | −           |
| 134 | B/Yamagata/1246/2003 | BAD29822         | −                   | −           | +           |
| 135 | B/Yamagata/1246/2003 | BAD29821         | −                   | −           | −           |
| 136 | B/Yamagata/115/2003 | BAD29820         | −                   | −           | +           |
| 137 | B/Yamagata/115/2003 | BAD29819         | −                   | −           | −           |
| 138 | B/Yamagata/222/2002 | BAD29818         | −                   | −           | −           |
| 139 | B/Yamagata/222/2002 | BAD29817         | −                   | −           | −           |
| 140 | B/Yamagata/K542/2001 | BAD29816         | −                   | −           | −           |
| 141 | B/Yamagata/K542/2001 | BAD29815         | −                   | −           | −           |
| 142 | B/Yamagata/K535/2001 | BAD29814         | −                   | −           | +           |
| 143 | B/Yamagata/K535/2001 | BAD29813         | −                   | −           | −           |
| 144 | B/Yamagata/K521/2001 | BAD29812         | −                   | −           | +           |
| 145 | B/Yamagata/K521/2001 | BAD29811         | −                   | −           | −           |
| 146 | B/Yamagata/K520/2001 | BAD29810         | −                   | −           | −           |
| 147 | B/Yamagata/K520/2001 | BAD29809         | −                   | −           | −           |
| 148 | B/Yamagata/K519/2001 | BAD29808         | −                   | −           | +           |
| 149 | B/Yamagata/K519/2001 | BAD29807         | −                   | −           | −           |
| 150 | B/Yamagata/K515/2001 | BAD29806         | −                   | −           | −           |
| 151 | NS1 [Influenza C virus] | BAB12065         | −                   | −           | −           |
| 152 | NS2 [Influenza C virus] | BAB12064         | −                   | −           | −           |
| 153 | NS1 [Influenza C virus] | BAB12063         | −                   | −           | −           |
| 154 | NS2 [Influenza C virus] | BAB12062         | −                   | −           | −           |
| 155 | NS1 [Influenza C virus] | BAB12061         | −                   | −           | −           |
| 156 | NS2 [Influenza C virus] | BAB12060         | −                   | −           | −           |
| 157 | NS1 [Influenza C virus] | BAB12059         | −                   | −           | −           |
| 158 | NS2 [Influenza C virus] | BAB12058         | −                   | −           | −           |
| 159 | NS1 [Influenza C virus] | BAB12057         | −                   | −           | −           |
| 160 | NS2 [Influenza C virus] | BAB12056         | −                   | −           | −           |
| 161 | NS1 [Influenza C virus] | BAB12055         | −                   | −           | −           |
| 162 | NS2 [Influenza C virus] | BAB12054         | −                   | −           | −           |
| 163 | NS1 [Influenza C virus] | BAB12053         | −                   | −           | −           |
| 164 | NS2 [Influenza C virus] | BAB12052         | −                   | −           | −           |
| 165 | NS1 [Influenza C virus] | BAB12051         | −                   | −           | −           |

(Continued)

(To be continued on the next page)
| No. | Strains                  | Accession number | Type_{orig} | LDA_{ext} | SVM_{ext} |
|-----|--------------------------|------------------|-------------|-----------|-----------|
| 166 | NS2 [Influenza C virus]  | BAB12050         | –           | –         | –         |
| 167 | NS1 [Influenza C virus]  | BAB12049         | –           | –         | –         |
| 168 | NS2 [Influenza C virus]  | BAB12048         | –           | –         | –         |
| 169 | NP [Influenza C virus]   | BAA98019         | –           | –         | –         |
| 170 | NP [Influenza C virus]   | BAA98018         | –           | –         | –         |
| 171 | PB1 [Influenza C virus]  | BAA98017         | –           | –         | –         |
| 172 | PB1 [Influenza C virus]  | BAA98016         | –           | –         | –         |
| 173 | CM2 [Influenza C virus]  | BAA98015         | –           | –         | –         |
| 174 | M1 [Influenza C virus]   | BAA98014         | –           | –         | –         |
| 175 | CM2 [Influenza C virus]  | BAA98013         | –           | –         | –         |
| 176 | C/Saitama/3/2000         | BAD36977         | –           | –         | –         |
| 177 | C/Saitama/2/2000         | BAD36976         | –           | –         | –         |
| 178 | C/Saitama/1/2000         | BAD36975         | –           | –         | –         |
| 179 | C/Miyagi/5/2000          | BAD36974         | –           | –         | –         |
| 180 | C/Miyagi/4/2000          | BAD36973         | –           | –         | –         |
| 181 | C/Miyagi/3/2000          | BAD36972         | –           | –         | –         |
| 182 | C/Miyagi/2/2000          | BAD36971         | –           | –         | –         |
| 183 | C/Yamagata/8/2000        | BAD36970         | –           | –         | –         |
| 184 | C/Yamagata/6/2000        | BAD36969         | –           | –         | –         |
| 185 | C/Yamagata/3/2000        | BAD36968         | –           | –         | –         |
| 186 | C/Yamagata/2/2000        | BAD36967         | –           | –         | –         |
| 187 | C/Yamagata/9/2000        | BAD36966         | –           | –         | –         |
| 188 | C/Yamagata/2/99          | BAD36965         | –           | –         | –         |
| 189 | C/Miyagi/4/2000          | BAD36964         | –           | –         | –         |
| 190 | C/Miyagi/3/2000          | BAD36963         | –           | –         | –         |
| 191 | C/Miyagi/2/2000          | BAD36962         | –           | –         | –         |
| 192 | C/Yamagata/8/2000        | BAD36961         | –           | –         | –         |
| 193 | C/Yamagata/6/2000        | BAD36960         | –           | –         | –         |
| 194 | C/Yamagata/3/2000        | BAD36959         | –           | –         | –         |
| 195 | C/Yamagata/2/2000        | BAD36958         | –           | –         | –         |
| 196 | C/Yamagata/9/2000        | BAD36957         | –           | –         | –         |
| 197 | C/Yamagata/2/99          | BAD36956         | –           | –         | –         |
| 198 | C/Yamagata/13/98         | BAD36955         | –           | –         | –         |
| 199 | C/Yamagata/13/98         | BAD36954         | –           | –         | –         |
| 200 | C/Yamagata/6/98          | BAD36953         | –           | –         | –         |

a) “+” and “−” represent avian influenza viruses and the non-avian influenza viruses, respectively. LDA_{ext} and SVM_{ext} represent the method with external validation for LDA and SVM, respectively.