Design and implementation of web-based genotyping system of Newcastle disease viruse

Yongzhong Cao
College of Veterinary Medicine
Yangzhou University
Yangzhou, China
e-mail: yzucao@yzu.edu.cn

Jianfen Zhang
Information Engineering College
Yangzhou University
Yangzhou, China
e-mail: zhangjf@yzu.edu.cn

Wanghong Lu
Information Engineering College
Yangzhou University
Yangzhou, China
e-mail: duobiao800521@163.com

Xiufan Liu
College of Veterinary Medicine
Yangzhou University
Yangzhou, China
e-mail: xfliu@yzu.edu.cn

Abstract   According to the genotyping method of Newcastle disease virus (NDV), a web-based genotyping system is developed by using Java Server Pages (JSP) technology and MySQL server database technology. Users input the sequence file in FASTA or GenBank format, or pasted either the nucleotide or amino acid sequences of the F gene or the complete genome sequence of NDV into given textbox, then the system was performed to acquire the information of enzyme restriction sites and specific amino acid sites after analyzing the sequence automatically. The sites information was then compared with the genotyping standards and the genotype of NDV strains which belong to class II division was determined and displayed subsequently. In the validation test, representative sequences of genotypes I-IX of NDV strains were determined individually, and the results of genotyping were completely consistent with the genotypes identified manually. The system is available online, without charge, at http://ndv.yzu.edu.cn/index.html .

Keywords: Newcastle Disease Virus; Genotyping system; JavaScript Page Language

I.   INTRODUCTION

Newcastle disease viruse (NDV), member of the genus Avulavirus within the Paramyxoviridae family, causes an economically important disease in all species of birds worldwide. NDV has a nonsegment, single-stranded RNA genome with envelope and has at least three genome lengths: 15186, 15192, and 15198 nucleotides. The genome is composed of six genes that encode the corresponding six structural proteins from the 3-end to the 5-end: nucleoprotein (NP), phosphoprotein (P), matrix (M), fusion (F), hemagglutinin-neuraminidase (HN), and the RNA polymerase (L) [1-3]. In recent years, many molecular biology and bioinformatics methods have been used for tracing NDV origins and spread. One of the methods is genotyping of different NDV strains. Phylogenetically, NDV has been classified into two major divisions, class I and class II. Because of important relationship with the poultry industry, class II viruses have been studied comprehensively. Class II NDV strains isolated in the past of four main pandemics have been classified into different genetic lineages and sublineages by restriction site mapping and sequence analysis of the F gene: genotype I contains avirulent viruses primarily from water fowls but also from chickens; genotypes II, III and IV contain viruses responsible for the first pandemic which started in the mid-1920s and subsided in the late

1950s; genotypes V and VI, emerged during the second pandemic between the 1960s and 1970s; in which subtype VIb viruses were responsible for the third pandemic of pigeon origin during the 1980s; novel genotypes and subtypes of VIIa, VIIb and VIII appeared in late 1980s and 1990s in the Far East , Europe and South Africa. The genotype IX viruses were found only in China. Subtype VIIc, VIId and VIIe viruses, which were considered to be the main pathogens for the fourth pandemic, were responsible for disease outbreaks in chicken and goose flocks and circulated predominantly in southern China in late 1990s and 2000s. The method of NDV genotyping is mainly according to the restriction enzyme (RE) site for *Hinf*I, *BstO*I and *Rsa*I across a region consisting of 75% of the F gene(334-1682) and the deduced amino acid sequences (residues 1-130) of the F gene[4-10].

Presently, phylogenetic tree construction based on the F gene fragment is adopted in the field of NDV genotyping as of the consistent results. But the drawback of this approach appears due to its complicated process. Researchers need to select several referenced F gene sequences containing all kinds of genotypes, convert them to a data file with specific format identified by phylogenetic tree constructing software. Huge works in preparing referenced genotyping sequences should be done. Meanwhile, a variety of algorithms are involved in phylogenetic tree constructing, such as maximum parsimony method, distance method, maximum likelihood method. Different algorithm may leads to different phylogenetic tree. Sometimes it is difficult to choose a suitable algorithm for researchers. In addition, it will take a long time to operate a series of software. Correctness of every step may influence the genotyping result directly.

In order to overcome the shortcomings of phylogeny-based method, a web-based genotyping tool was developed by using Java Server Pages (JSP) and MySQL database technology according to the NDV genotyping method based on F gene RE site and deduced amino acid sequences (residues 1-130). When a standard GenBank or Fasta format sequence file is submitted, or the NDV complete genome sequence or F gene sequence is pasted into the textbox directly, all process of genotyping will be completed automatically, including separation of F gene, acquirement information of restriction enzyme sites and deduced amino acid specific sites and comparison with genotyping standard. Finally, the result of genotyping and relevant information will be offered directly.

II. SYSTEM DESIGN

A. *system architecture*

By using browser/server (B/S) structure, three-tier model was adopted: data layer, business logic layer and presentation layer. Data layer mainly undertakes the storage and management of genotyping standard data. Business logic layer mainly undertakes the information collection regarding RE sites and specific amino acid sites and information comparison with genotyping standard database. Presentation layer undertakes data input and output.

As designed for this three-tier model, the data and its application are separated so as to enhance the stability and extensibility of this system. Data layer includes genotyping standard data, RE sites data and specific amino acid sites data (showed in tab1 and tab2). Genotyping standard data is accessed through Java DataBase Connectivity (JDBC) technology. Both RE sites data and specific amino acid sites data is analyzed and acquired by using JavaScript language. Business logic layer includes nucleic acid-based genotyping service and F protein-based genotyping service. They process nucleic acid sequence or amino acid sequence respectively. The result to web server is sent by using JSP technology. The presentation layer is of web client. Users can accomplish genotyping process through browser [13-16]. The system architecture is showed in **Fig .1**
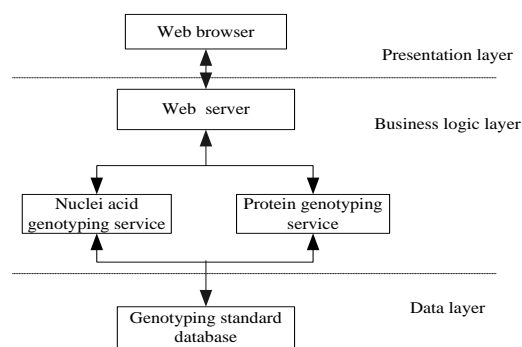
Figure 1.   system architecture

B. *System function*

As is shown in **Fig .2**, the system consists of five function modules. The upper module is input/output module, which realizes the function either submission of sequence or output of genotyping result. The second is sequence processing module. As one of important parts, it primarily deals with the length judgment of submitted sequence, F gene separation and its coding district sequence (CDS) translation, acquirement information of RE sites and specific amino acid sites. The third is the module of genotyping in terms of nucleic acid sequence. It takes function mainly to compare RE sites with genotyping standard. The fourth is the module of genotyping based on amino acid sequence of F protein, which achieves the function of comparison between specific amino acid sites and genotyping standard. The last is standard processing module which processes addition, modification and deletion of standard. Because of its frequent variation, new genotype of NDV may be identified, therefore, the genotyping standard needs to be modified in time. In database, binary code is adopted to express RE sites standard, 0 represents the absence of RE sites, and 1 represents the presence of RE sites.
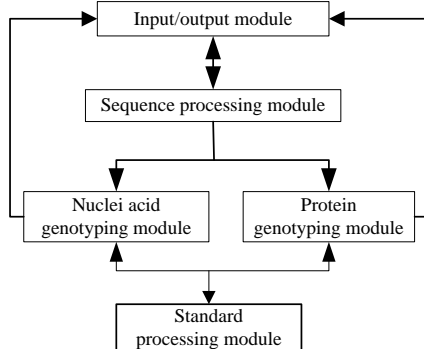
Figure 2.   system function module diagram

III. SYSTEM REALIZATION

A. Establishment of genotyping standard database

According to the published literature [5-12], the standard of NDV genotyping was confirmed. It mainly based on the distribution pattern of restriction enzyme cleavage sites on the F gene and deduced amino acids of the F gene variable region.

In the genotyping standard database, two sections are covered as nucleic acid-base database and amino acid-based database. The structure of nucleic acid-based database is designed as shown in **Table I.,** which includes

genotype ID, the information of RE sites (*Hinf*I site, *BstO*I site, *Rsa*I site). In order to facilitate sequences analysis, the figures 0 and 1 are used to represent the absence and presence of every RE site respectively. The structure of amino acid-based database is designed as shown in **Table II.,** which includes genotype ID and all specific amino acid sites information.

Table I. Structure of nucleotide genotyping table

Field	Type	Reference
first	bigint	Genotyping ID
second …	char	*Hinf* I site information
sixth …	char	*BstO* I site information
eleventh … sixteenth	char	*Rsa* I site information

Table II. Structure of protein genotyping table

Field	Type	Reference
first	bigint	genotype ID
second	char	the information of
…	…	twenty-six amino
twenty-seventh	char	acid character sites

B. Algorithm of genotyping

The genotyping process is divided to two steps.

(i) Information acquirement.

The process includes the separation of F gene sequence from submitted sequence, acquirement of RE sites information, translation CDS of the F gene, and acquirement of specific amino acid sites information.

(ii) Comparison with genotyping standard.

After completing step one, comparison of the obtained information with genotyping standard in database and determination of the genotype of the submitted sequence is involved in step two. Matching strategy item by item is used in the process of comparison between the obtained information from submitted sequence and genotyping standard. The matching degree is introduced in the light of existence of imperfect matching, which represents the consistency between the sites information and genotyping standard. When the obtained site information is compared with genotyping standard of all genotypes, a group of

data will be produced and sort from high to low by bubble sort algorithm. The greater the value of matching-degree, the closer the genotype determined.

C. Computer environment of system implementation

The implementation of this system is based on the following computer environment.

Operating system: Windows 2000/XP.

Application server: Tomcat 5.0.

Database system: MySQL 6.0.

Development platform: Myeclipse 6.5[17].

D. Running case

The system function is mainly realized through input and output interface. The browser was used as user interface. Users do not need any special computer skills just inputting a given IP address (http://ndv.yzu.edu.cn /index.html) and following the instruction.

Take the NDV strain Mukteswar for example. The complete genome is 15186nt in length. Through manual analysis, we can find that there are *Hinf* I cleavage sites at position 885nt and 883nt, *BstO* I cleavage sites at position 1116nt and 1601nt, and *Rsa* I cleavage sites at position 540nt and 1625nt. Furthermore, 26 specific amino acid sites with protein genotyping standard compared one by one. Amino acid mutations are found at position of site 4 (K→ P), site 16 (I→T), site 17 (T→I), site 104 (G→E), site 107 (S→T) , site 115 (K→R) and site 124 (G→S). These results indicate that Mukteswar can be identified as genotype III. Meanwhile, the results from our genotyping system (Figure 3) show that the three RE sites (*Hinf* I , *BstO* I , *Rsa* I) of F gene are "0110,00101,100001" and the specific amino acid sites of F protein are "RSIPLTIILLVTREVTSRRQRRF IIS". Therefore, the genotype of NDV strain Mukteswar is "III", which is exactly consistent with the result from manual analysis.

IV. CONCLUSION

A web-based genotyping system of NDV is developed based on the criterion of genotyping relying mainly on NDV F gene. By the practice validated, the advantages of the system are as follows. First, it compares the submitted sequence with genotyping standard directly. It overcame the inconvenience of phylogeny-based method and realized accurate genotyping of NDV. Second, the whole genotyping process is completed by the computer except

nucleic acid or amino acid sequence submission, so the time of genotyping has been shortened greatly. In addition, the operation of this tool is very easy to use. The further development of this system may be used to find the site mutation of NDV genome. The system is available online at http://ndv.yzu.edu.cn/index.html .

REFERENCES

[1] Fauquet CM, Fargette D: International Committee on Taxonomy of Viruses and the 3,142 unassigned species. Virol. J . 2005, 2: 64

[2] Huang Y, Wan HQ, Liu HQ, Wu YT, Liu XF. Genomic sequence of Newcastle disease virus isolated from an outbreak in geese: a novel six nucleotide insertion in non-coding region of the nucleoprotein gene. Arch Virol . 2004, 149(7):1445-1457.

[3] Czegládi A, Ujvári D, Somogyi E, Wehmann E, Werner O, Lomniczi B: Third genome size category of avian paramyxovirus serotype 1 (Newcastle disease virus) and evolutionary implications. Virus Res. 2006, 120:36–48.

[4] Aldous EW, Alexander DJ: Detection and differentiation of Newcastle disease virus (avian paramyxovirus type1). Avian Pathol . 2001, 30(2): 117–128.

[5] Ballagi-Pordany A, Wehmann E, Herczeg J, Belak S, Lomniczi B: Identification and grouping of Newcastle disease virus strains by restriction site analysis of a region from the F gene. Arch Virol. 1996, 141:243-261.

[6] Herezeg J, Wehamann E, Bargg RR, Travassos DPM, Hadjiev G, Werner O, Lomniczi B: Two novel genetic groups (VIIb and VIII) responsible for recent Newcastle disease outbreaks in Southern Europe. Arch Virol. 1999: 144: 2087–2099.

[7] Herczeg J, Pascucci S, Massi P, Luini M, Selli L, Capua I, Lomniczi B : A longitudinal study of velogenic Newcastle disease virus genotypes isolated in Italy between 1960 and 2000. Avian Pathol. 2001: 30: 163–168.

[8] Li Y, Wang ZL, Jiang YH, Chang LK, Wang J : Characterization of newly emerging Newcastle disease virus isolates from the People's Republic of China and Taiwan. J Clin Microbiol. 2002, 39: 3512–3519 .

[9] Liang R, Cao DJ, Li JQ, Chen J, Guo X, Zhuang FF, Duan MX: Newcastle disease outbreaks in western China were caused by the genotypes VIIa and VIII. Vet Microbiol . 2002, 87: 193–203.

[10] Lomniczi B, Wehmann E, Herczeg J, Ballagi-Pordany A, Kaleta EF, Werner O, Meulemans G, Jorgensen PH, Mante AP, Gielkens ALJ, Capua I, Damoser J: Newcastle disease outbreaks in recent years in Western Europe were caused by a (VI) and a novel genotype (VII). Arch Virol. 1998, 143: 49–64.

[11] Xiufan Liu, Hongquan Wan, Xiaoxia Ni, Yantao Wu, Wenbao Liu: Pathotypical and genotypical characterization of strains of Newcastle disease virus isolated from outbreaks in chicken and goose flocks in some regions of China during 1985-2001. Arch Virol. 2003,148:1387-1403.

[12] Yantao Wu, Xuexia Ni, Hongquan Wan, Wenbo Liu, Xiufan Liu: Molecular epidemiological study on Newcastle Disease Virus originated from different animals in China. Chinese Journal of Virology (In Chinese) . 2002, 18(3):264-269.

[13] Vivek Chopra, Amit Bakore, Ben Galbraith, Sing Li, Chanoch Wiggers: Professional Apache Tomcat 5. Wiley Publishing Inc. Indianapolis, Indiana. 2004: 293-320.

[14] Hans Bergsten: JavaServer Pages. O'Reilly Media. 3rd Edition. 2003.

[15] JavaScriptSource Staff: Core JavaScript Reference 1.5. 2006. [http://www.javascriptsource.com/]

[16] Ricky Setyawan: Best Practices for Web, Mobile & Cloud Applications. 2013. [http://www.mysql.com/]

[17] http://www.myeclipseide.cn/