Original Article

Molecular Characteristics and Phylogenetic Analysis of N Gene of Human Derived Rabies Virus*

CAI Liang[1], TAO XiaoYan[2], LIU YunZhi[1], ZHANG Hong[1], GAO LiDong[1], HU ShiXiong[1], LIU FuQiang[1], LI Hao[2], SHEN XinXin[2], LIU JiaHui[1], WANG ShiQing[1], and TANG Qing[2,#]

1. Hunan Center for Disease Control and Prevention, Key Laboratory of Microbial Molecular Biology of Hunan Province, Changsha 410005, Hunan, China; 2. State Key Laboratory for Molecular Virology and Genetic Engineering, Institute for Viral Disease Control and Prevention, China CDC, Beijing 102206, China

Abstract

Objective To investigate the relationship between the molecular characteristics and phylogenetic evolution of rabies N gene.

Methods Saliva samples were collected from rabies cases, and RT-PCR was used to amplify the N gene of rabies virus with the specific primers. The amplifying product of RT-PCR was cloned to pUCm-T vector and transformed into *E.coli* XL1-Blue and then the blue-white selection, PCR screening and gene sequencing were carried out to identify the positive clones. Finally, ExPASy and other bioinformatics softwares were used to analyze and predict the structure and biological characteristics of the N genome.

Results The amplification product of RT-PCR was 1 353 bp, the recombinant plasmid pUCm-T/N was constructed, the whole length of the N gene open reading frame was composed of 1 353 nucleotide residues to code 450 amino acids (20 kinds), the accession number submitted to the Genbank was HM756692, its sequence homology of nucleotides and amino acids compared with the vaccine strain CTN-1-V was 90% and 99% respectively. The evolutionary analysis showed that the isolated strain belonged to genotype I with certain geographic regionality.

Conclusion The characteristics investigation and bioinformatics analysis of Hunan0806 N gene will provide fundament data to reveal the significance of the N gene characteristics for rabies epidemiology and its prevention & control.

Key words: Rabies; N gene; Molecular Characteristics; Phylogenetic

Biomed Environ Sci, 2011; 24(4):431-437 doi:10.3967/0895-3988.2011.04.015 ISSN:0895-3988

www.besjournal.com(full text) CN: 11-2816/Q

INTRODUCTION

Rabies is an acute zoonosis that mainly attacks central nervous system after infection with rabies virus, and more than 55 000 people die of rabies every year all over the world[1]. The genome of rabies virus is about 12 kb, and totally five structural genes are separated by non-coding intergenic sequences from 3' terminus to 5' terminus, namely N, P, M, G, and L, which encode respectively the nucleoprotein, phosphoprotein, matrix protein, glycoprotein and large subunit of transcriptase[2]. N gene in the whole genome of rabies virus is highly conserved and can be efficiently

*This research was supported by the science foundation of the Health Department of Hunan Province (Grant No A2007008 and B2009-088), the National Department Public Benefit Research Foundation (Grant No 200803014)

#Correspondence should be addressed to TANG Qing, Tel: 86-10-58900895, E-mail: qtang04@sina.com

Biographical note of the first author: CAI Liang, male, born in 1980, Master of Medicine, graduated from the university of south China, the professional research fields is mainly in the basic research of arbovirus.

Received: March 8, 2011; Accepted: May 13, 2011

expressed, which is the reason why N gene is frequently used as a judging index for genotyping and population variation of rabies viruses[3]. We carried out preliminary studies on the characteristics of biology, structure and function as well as phylogenetic evolution of the nucleoprotein genes by cloning and analyzing the complete open reading frame of rabies virus N gene in this study and the results of the study is described as below.

MATERIALS AND METHODS

Bacterial Strain and Virus Sample

E.coli XL1-Blue was a reserved strain from the laboratory in Hunan Center for Disease Control and Prevention; the virus positive samples were derived from the saliva samples collected from human rabies cases in the national surveillance points for rabies in Yongzhou, Hunan province in 2008.

Enzymes, Major Reagents and Equipment

pUCm-T vector, T 4 DNA ligase, UNIQ-10 mini plasmid extraction column and UNIQ-10 gel recovery column were purchased from Shanghai Sangon Biological Engineering Technology & Services Co., QIAamp viral RNA Mini kit (QIAGENE, Germany), One-step RT-PCR System with Platinum® Taq High Fidelity (Invitrogen, USA), BigDye Terminator 3.1/1.1 version sequencing kit (ABI, USA), 2 000 bp DNA Marker, SOC medium, IPTG, X-Gal, and other products were all purchased from Dalian TaKaRa Company, PCR Amplifier (PCT-200 Thermal, Germany), Applied Biosystems 3 730 sequencer (ABI, USA), electrophresis apparatus(Bio-rad, USA).

Design and Synthesis of Primers

The sequences of PCR primers were provided by the Department of Viral Encephalitis of the Institute for Viral Disease Control and Prevention, China CDC with reference to the sequence data from the Institute Pasteur (GenBank Accession No: M13215), P1: 5'-ATGTAACAC CTCTACAATGG-3', P2: 5'- GGATTGACRAAGATCTTGCTCAT-3', which were synthesized by Shanghai Sangon Biological Engineering Technology & Services Co.

Extraction of RNA, RT-PCR

Total RNA in the saliva samples was extracted with QIAamp viral RNA Mini kit, and RT-PCR was carried out using One-step RT-PCR System with Platinum® Taq High Fidelity kit. The reaction condition of RT-PCR included the pre-denaturation: 42 °C 45 min, 95 °C 3 min; then denaturation: 95 °C 25 s, annealing: 45 °C 30 s, elongation: 72 °C 1 min, for 30 cycles; and final elongation at 72 °C for 5 min.

Clone, Sequencing of N Gene

The product of RT-PCR was subjected to 0.8% agarose gel electrophoresis, the purification and recovery of PCR product were carried out according to the instruction of UNIQ-10 gel recovery kit. The recovered N gene segment was cloned to pUCm-T vector and the recombinant clone was transformed into the *E.coli* XL1-Blue competent cells[4] and the recombinant plasmid pUCm-T/N was further identified by sequencing using the BigDye Terminator 3.1/1.1 version sequencing kit with 60 ng of template per μL, 2.0 pmol of M13 and T7 primers according to the manufacturer's instructions on Applied Biosystems 3 730 sequencer.

Sequence Analysis

Seqman biological software was used for sequence calibration and splicing, and the proved sequence was named and submitted to Genbank by Sequin software, BLAST procedure was carried out to compare with the N gene of the human rabies vaccine strain CTN-1-V from GenBank (GenBank Accession No: AF367863) on both of the nucleotide and amino acid sequences. BioEdit and other biological softwares were used to analyze the characteristics of genetic constitution for the obtained sequences, and Protean and Expert Protein Analysis System (ExPASy) were used to predict and analyze the structures and functions of the encoding proteins.

Evolutionary Analysis

N gene sequences were selected from the representative rabies viruses in different areas that were gathered by GenBank, the alignment and phylogenetic evolutionary analysis of the sequences were finished using Clustal W, MEGA4.1, and other biological softwares.

RESULTS

PCR Amplification and Identification of N Gene

PCR amplification product was subjected to 0.8% agarose gel electrophoresis and a single band with similar length to the anticipated fragment was found at 1 353 bp position. The electrophoretic band

of the product was in accordance with the predicted size after the PCR identification of the T-A cloning result (Figure 1).

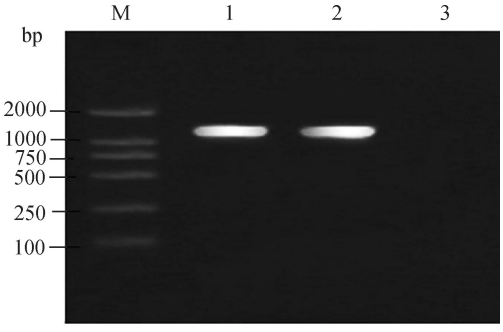

Figure 1. Agrose gel analysis of RT-PCR and T-A clone of rabies N gene. M: 2 000 bp DNA marker; 1: RT-PCR result of rabies N gene; 2: PCR identification result of plasmid pUCm-T/N; 3: negative control.

Sequencing and Homology Analysis

The sequencing result showed that the N gene segment had been successfully cloned into pUCm-T vector and the infected virus strain in the positive rabies sample was named as Hunan0806 after sequence identification. Compared with the N gene of human rabies vaccine strain CTN-1-V that was isolated from Shandong province of China in 1956, the homology in nucleotide was 90% and the homology of amino acids was 99%, and four mutation sites can be found in the amino acids of Hunan0806, i.e. Ser at position 100, Ala at position 214, Glu at position 239 and Val at position 338, while the corresponding amino acids of CTN-1-V were T, N, S, and K respectively (Figure 2), suggesting that N gene of Hunan0806 was highly homologous to CTN-1-V strain in both of the nucleotides and amino acids sequences.

```
Query  1    MDADKIVFKVNNQVVSLKPEIIVDQYEYKYPAIKDLKKPSISLGKAPDLNKAYKSVLSGM  60
Sbjct  1    MDADKIVFRVNNQVVSLKPEIIVDQYEYKYPAIKDLKKPSITLGKAPDLNKAYKSVLSGM  60
Query  61   NAAKLDPDDVCSYLAAAMQFFEGTCPEDWTSYGILIARKGDKITPDSLVEIKRTDVEGNW  120
Sbjct  61   NAAKLDPDDVCSYLAAAMQFFEGTCPEDWNSYGILIARKGDKITPDSLVDIKRTDVEGNW  120
Query  121  ALTGGMELTRDPTVSEHASLVGLLLSLYRLSKISGQNTGNYKTNIADRIEQIFETAPFVK  180
Sbjct  121  ALTGGMELTRDPTVSEHASLVGLLLSLYRLSKISGQSTGNYKTNIADRIEQIFETAPFVK  180
Query  181  IVEHHTLMTTHKMCANWSTIPNFRFLAGTYDMFFSRIEHLYSAIRVGTVVTAYEDCSGLV  240
Sbjct  181  IVEHHTLMTTHKMCANWSTIPNFRFLAGTYDMFFSRIEHLYSAIRVGTVVTAYEDCSGLV  240
Query  241  SFTGFIKQINLTAREAILYFFHKNFEEEIRRMFEPGQETAVPHSYFIHFRSLGLSGKSPY  300
Sbjct  241  SFTGFIKQINLTAKEAILYFFHKNFEEEIRRMFEPGQETAVPHSYFIHFRSLGLSGKSPY  300
Query  301  SSNAVGHVFNLIHFVGCYMGQVRSLNATVIAACAPHEMSVLGGYLGEEFFGKGTFERRFF  360
Sbjct  301  SSNAVGHVFNLIHFVGCYMGQVRSLNATVIAACAPHEMSVLGGYLGEEFFGKGTFERRFF  360
Query  361  RDEKELQEYEAAELTKTDLALADDGTVNSDDEDYFSGETRSPEAVYTRIMMNGGRLKRSH  420
Sbjct  361  RDEKELQEYEAAELTKTDLALADDGTVNSDDEDYFSGETRSPEAVYTRIMMNGGRLKRSH  420
Query  421  IRRYVSVSSNHQARPNSFAEFLNKTYSSDS  450
Sbjct  421  IRRYVSVSSNHQARPNSFAEFLNKTYSSDS  450
```

Figure 2. Comparison of amino acid sequence of Hunan0806 and the CTN-1-V vaccine strain.

Characteristics Analysis of N Gene Sequence

The coding sequence of N gene started with the initiation codon ATG and terminated with the stop codon TAA, and the complete length of the open reading frame (ORF) was composed of 1 353 nucleotide residues, including 383 base A (28.31%), 273 base C (20.18%), 329 base G (24.32%), and 368 base T (27.20%), the GC content was 44.49% and the AT content was 55.51%, and the molecular weight of a single strand was about 409.7 KD. The sequence of N gene totally coded 450 amino acids (20 kinds), the amino acid with the high content was Ser (serine), which accounted for 8.22% of the total content of proteins (37/450), while the amino acid with the lowest content was Trp (tryptophan), which only accounted for 0.67% (3/450) (Table 1).

Physical and Chemical Characteristic Analysis of N Protein

It can be found from the analysis on the titration curve of this protein that the theoretical isoelectric point of the nucleoprotein was 6.04 (Figure 3). Results from ProtParam analysis showed that the

total number of negatively charged amino acid residues in nucleoprotein (Asp+Glu) was 56, while the total number of positively charged amino acid residues (Arg+Lys) was 49, and the total number of atoms was 7 066; among them the atom numbers of C, H, O, N and S were 2 269, 3 496, 679, 604, and 18 respectively, thus its molecular formula was deduced as $C_{2269}H_{3496}N_{604}O_{679}S_{18}$.

Table 1. Amino Acid Composition Analysis of Hunan0806 N Gene

Amino Acids	Amount (%)	Amino Acids	Amount (%)	Amino Acids	Amount (%)	Amino Acids	Amount (%)	Amino Acids	Amount (%)
A	33 (7.33)	C	6 (1.33)	D	24 (5.33)	E	32 (7.11)	F	27 (6.0)
G	30 (6.67)	H	13 (2.89)	I	26 (5.78)	K	26 (5.78)	L	35 (7.78)
M	12 (2.67)	N	20 (4.44)	P	16 (3.56)	Q	10 (2.22)	R	23 (5.11)
S	37 (8.22)	T	28 (6.22)	V	28 (6.22)	W	3 (0.67)	Y	21 (4.67)

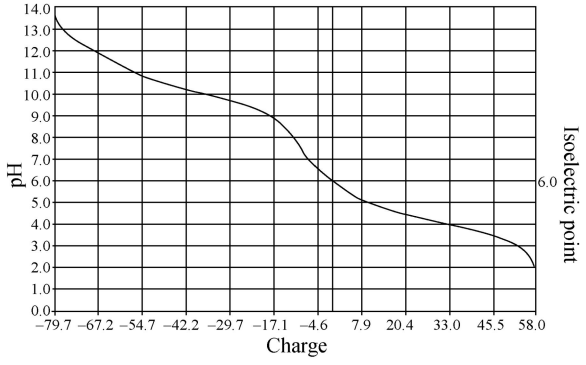

Figure 3. Isoelectric point and titration curve of the nucleoprotein of Hunan0806.

Hydrophobicity and Hydrophilicity Analysis of N Protein

ProtScale displayed the results of hydrophobicity analysis by diagrams. It was found that this protein contained about 10 highly hydrophobic domains, which were located at the amino acids positions 23, 60, 75, 140, 175, 215, 225, 240, 320, and 330, while the five major domains of nucleoprotein with the lowest scores were found at the amino acids positions 25, 105, 160, 270, and 370, which indicated that these domains were highly hydrophilic (Figure 4).

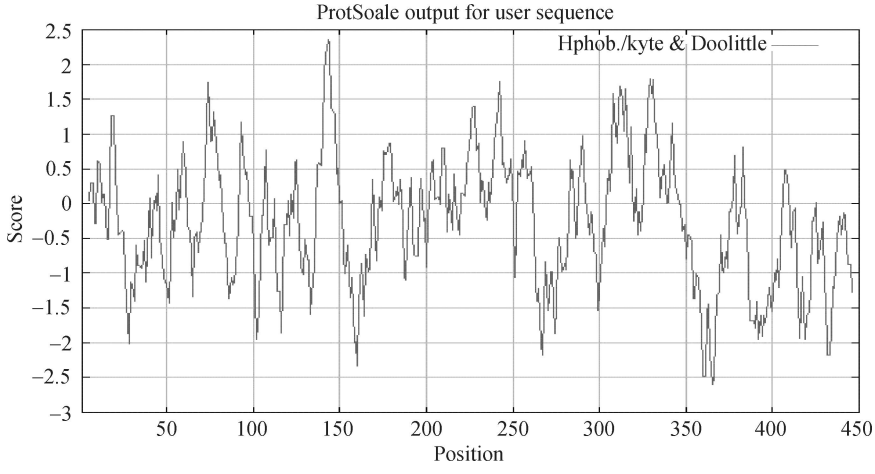

Figure 4. Hydrophilicity and hydrophobicity profile of the nucleoprotein of Hunan0806. The abscissa is the position of the amino acid sequence and the ordinates is the scaling value of the amino, the Hphob.Kyte & Doolittle scale is the higher value of the hydrophobic amino acid (>0 is indicated hydrophobicity while <0 is indicated hydrophilicity)

Prediction of N Protein Secondary Structure

The amino acid sequence of N gene was submitted to PredictProtein for predicting the secondary structure of the protein, and it was found that the secondary structure of nucleoprotein was composed of the N terminus domain (NTD) and the C terminus domain (CTD), and among them CTD was the most variable region in the integrase amino acid sequence, including nine α-helixes, while NTD

contained HX3-7HX23-32CX2C zinc-finger structure in HHCC region, which included seven α-helixes and two β-foldings (Figure 5).

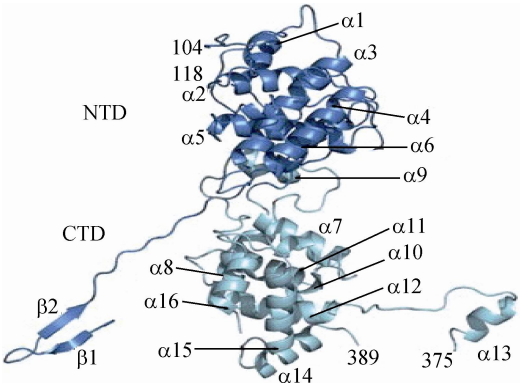

Figure 5. Helix secondary structure of the nucleoprotein of Hunan0806. The deep blue area indicates the amino terminal structural domain and the light blue area indicates the carboxyl terminus structural domain.

Phylogeny Evolution Analysis of N Gene

Phylogenic analysis was carried out on N gene from the representative rabies viruses all over the world in GenBank (Table 2) and the isolated strain of Hunan0806 (Figure 6), the N gene sequence of Hunan0806 was clustered into the same branch with all other genotype I rabies viruses, indicating that Hunan0806 belonged to the genotype I rabies virus, and most of the isolated rabies viruses strains in Hunan province were all clustered into the branch Ia, and the intervals from the different evolutionary groups to the original point of the division was not significant. The phylogenic analysis of N gene indicated that the distribution of rabies virus showed certain regional similarity, in other word, the samples from the same geographic setting tended to be clustered into the same evolutionary cluster or sub-cluster, which was in accordance with the conclusion that rabies virus revealed significant regional characteristics. Analysis was carried with the two human rabies vaccine strains used in China(CTN strain and 3aG strain), and the results showed that Hunan0806 had closer genetic relationship with the CTN strain as well as the prevailing rabies viruses distributed in most of the areas of China, but was relatively farther distant from the 3aG strain, which might be related to the species difference and the hosts difference for the isolation of these two viruses strains.

Table 2. Rabies Virus Used in the Study of Molecular Epidemiology of Rabies Virus

Isolate or Strain	GenBank Accession No	Region of Origin	Host	Isolated Year
CTN	AF367863	Shangdong	Human	1956
SADB19	M31046	USA	Vaccine strain	1935
SRV9	AF499686	China	Vaccine strain	--
PV	M13215	France	Vaccine strain	1965
3aG	AF155039	Beijing	Vaccine strain	1931
RC-HL	D16331	Japan	Vaccine strain	2000
CVS	AF406696	France	Dog	2001
CHN0635H	EF990621	Hunan	Human	2006
CHN0642D	EF990622	Hunan	Dog	2006
CHN0701D	EF990623	Hunan	Dog	2007
D774-45	DQ267925	Thailand	Dog	--
Hunan-Dk13	DQ666307	Hunan	Dog	2004
Hunan-Wg432	DQ666316	Hunan	Dog	2004
Guizhou-A148	DQ666291	Guizhou	Dog	2004
HuNPN16	DQ515993	Hunan	Dog	2006
GX074	DQ866107	Guangxi	Dog	2007
GX304	DQ866117	Guangxi	Dog	2006
HNDB33	EU008922	Hunan	Dog	2007
Hunan-Xx35	DQ666319	Hunan	Dog	2004
Hunan-Xx33	DQ666317	Hunan	Dog	2004
HNDB11	EU008919	Hunan	Dog	2007
HNDB18	EU008921	Hunan	Dog	2007
GXN119	DQ866111	Guangxi	Dog	2006
CTN181	EF564174	Shandong	Human	2007
Guizhou-A103	DQ666290	Guizhou	Dog	2004
HuNPN01	DQ496219	Hunan	Pig	2006
Guizhou-A158	DQ666292	Guizhou	Dog	2004

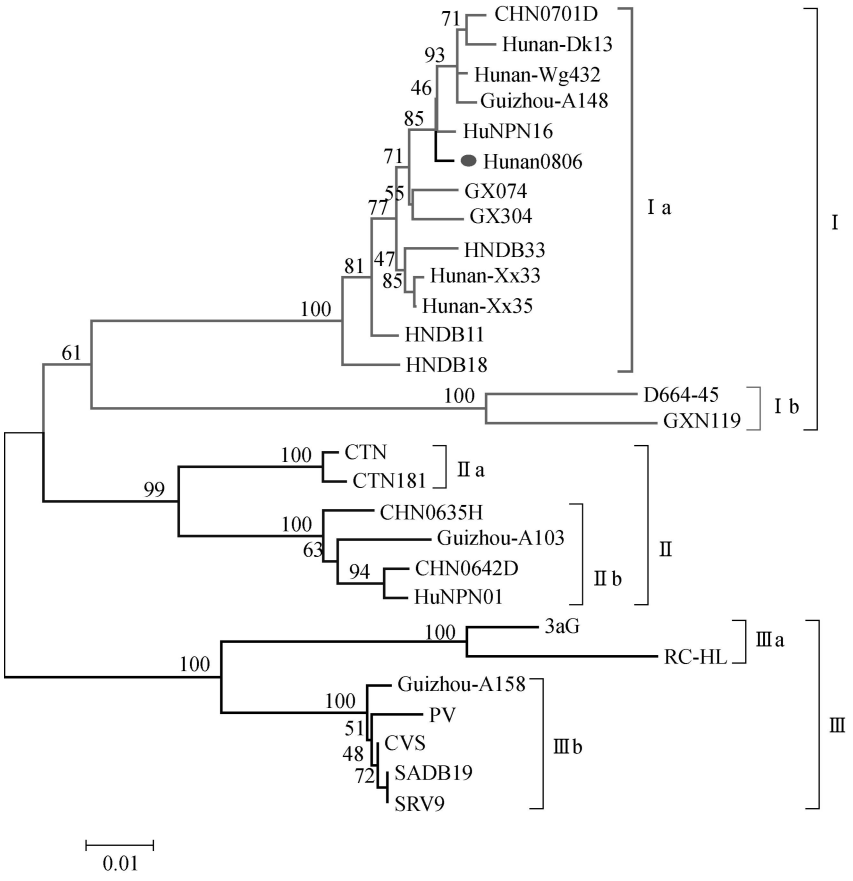

Figure 6. Phylogenetic analysis of rabies N gene (N-J method). The figure was draw by MEGE 4.1 software with maximum composite likelihood model. Bootstrap values are calculated from 1 000 repetitions and 64 238 seeds. The numbers of Sites are 1 353 and the Codon Positions are 1st+2nd+3rd+Noncoding. Scale bares represent phylogenetic distance between isolates. Numbers at each node indicate degree of bootstrap support and the Pattern among Lineages is Homogeneous.

DISCUSSION

Sequencing data of biological genomes have increased exponentially now and the sequenced data have become hotspots for genomic and proteomic investigations[5]. The present study has completed the sequencing of the whole length of the N gene of Hunan0806 rabies virus sample and deduced the composition of amino acids of the N gene, which provided the molecular basis for further investigations on genomic structures, biological characteristics, pathogenic mechanisms and other data for Hunan0806.

The primary structure of the nucleoprotein of rabies virus, the sequence of amino acids in the polypeptide chain, is the key point to determine the spatial structure of the nucleoprotein, while the composition of hydrophilic and hydrophobic amino acids is the major motive for nucleoprotein folding, and the corresponding analysis can determine transmembrane helix and other secondary structures as well as the distribution of amino acids on the surface of the nucleoprotein[6]. ExPASy is the technical alliance that is composed of Swiss Center of Bioinformatics, European Bioinformatics Center and Protein Informatics Resource, and all of the submitted sequences to them have been verified manually, the quality of the annotation is high and only little sequence redundancy can be found[7]. We carried out prediction and analyses of biological characteristics of N gene of Hunan0806, deduced amino acid composition, physical and chemical characteristics, titration curve of the protein, hydrophobicity and hydrophilicity as well as the secondary structures by ExPASy, and it was found that the N gene of Hunan0806 contained 1 353 nucleotides and encoded 450 amino acids, and its secondary structure included two domains, namely

CTD and NTD. CTD showed metal ion independent non-specific DNA binding activity which was related to the formation of integrase multimer, while NTD was able to promote the tetramerization of integrase and enhance the catalytic activity which was playing a key role in the formation of stable compound between enzymes and viruses. The sequence of Hunan0806 was compared with the currently used human vaccine strain CTN-1-V in China, and it was found that the homology in nucleotides was as high as 90% and the homology in amino acids was as high as 99%, and the differences were only found in four non-specific sites in the amino acids and the key sites related to the antigenicity of the virus were unchanged, indicating that the CTN-1-V could provide effective immunological barrier for the infectivity of this kind of viruses.

The construction of phylogenic cladogram of N gene from rabies viruses is the process of evaluation and simulation for the topology of evolutionary relationship among the sequences, which can display the possible source and transmission route of rabies as well as the cycling transmission and historic migration among the hosts in a certain degree. It can be used for laboratory diagnosis, genotyping and molecular epidemiological investigation for rabies virus infection[8]. The present study has constructed the phylogenic cladogram of N gene by N-J method, and has found that Hunan0806 belongs to genotype I of rabies viruses, which corresponds well with the previous research results[9]. Most of the isolated rabies virus strains from Hunan province have been clustered into the same branch Ia, while Ia includes seven dog rabies virus strains and one human rabies virus strain Hunan0806, indicating that mutual infection and transmission of rabies virus might exist between animal hosts and human.

This research is aimed to explore the influence and relationship of rabies virus N gene structure on the pandemic and distribution of rabies as well as the cross protection between rabies vaccine strain with the epidemic street rabies virus, which will be very important for providing the substructure data for rabies prevention and control.

ACKNOWLEDGEMENTS

The authors sincerely express their gratitude to those who were directly or indirectly involved in the samples collection and data providing.

REFERENCES

1. Knobel DL,Cleaveland S, Coleman PG. Re-evaluating the burden of rabies in Africa and Asia. Bull World Organ. 2005; 83, 360-8.
2. Pringle CR: The order monongeavirales. Arch Virol, 1991; 117(1-2), 137-40.
3. Wunner WH, Larson JK, Dietzschold B, et al. The molecular biology of rabies viruses. Rev Infect Dis, 1998; 10, 771-84.
4. Sambrook J, Fritsch EF, Maniatis T, et al. Molecular Cloning: A. Laboratory Manual. 2nd ed. Scientific Press.
5. Bansal AK. Bioinformatics in microbial biotechnology-a mini review[J]. Microb Cell Fact, 2005; 4(1), 19.
6. Cuff JA, Barton GJ. Evaluation and improvement of multiple eqnence methods for protein secondary structure prediction[J]. Proteins, 1999; 34(4), 508-19.
7. Gasteiger E, Gattiker A, Hoogland C, et al. ExPasy: the proteomics server for in-depth protein knowledge and analysis: Nucleic Acids Res, 2003; 31, 3784-88.
8. Tao XY, Tang Q, Li H, et al. Molecular epidemiology of rabies in southern people's Republic of China. Emery Infect Dis, 2009; 15, 1192-8.
9. Hu R, Tang Q, Tang J, et al. Rabies in China: an update. Vector Borne Zoonotic Dis, 2009; 9, 1-12.