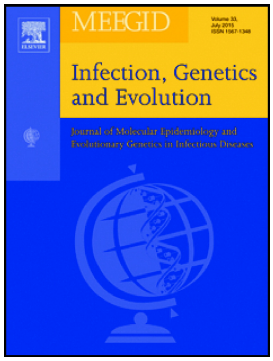

Codon usage bias in the N gene of rabies virus

Wanting He, Hongyu Zhang, Yuchen Zhang, Ruyi Wang, Sijia Lu, Yanjie Ji, Chang Liu, Pengkun Yuan, Shuo Su

PII: S1567-1348(17)30269-1
DOI: doi: 10.1016/j.meegid.2017.08.012
Reference: MEEGID 3233

To appear in: *Infection, Genetics and Evolution*

Received date: 8 May 2017
Revised date: 11 August 2017
Accepted date: 12 August 2017

Please cite this article as: Wanting He, Hongyu Zhang, Yuchen Zhang, Ruyi Wang, Sijia Lu, Yanjie Ji, Chang Liu, Pengkun Yuan, Shuo Su , Codon usage bias in the N gene of rabies virus, *Infection, Genetics and Evolution* (2017), doi: 10.1016/j.meegid.2017.08.012

Codon usage bias in the N gene of Rabies Virus

Wanting He[a], Hongyu Zhang[a], Yuchen Zhang[a], Ruyi Wang, Sijia Lu, Yanjie Ji, Chang Liu, Pengkun Yuan, Shuo Su*

Jiangsu Engineering Laboratory of Animal Immunology, Institute of Immunology and College of Veterinary Medicine, Nanjing Agricultural University, Nanjing, China

[a] Wanting He, Hongyu Zhang and Yuchen Zhang contributed equally to this work.

*Corresponding author:

shuosu@njau.edu.cn

Abstract

Since its emergence, Rabies virus (RABV) has been a major worldwide concern especially in developing countries. The nucleoprotein (N) of RABV is highly conserved and key for genetic typing, thus a better understanding of the N gene evolutionary trajectory can assist the development of control measures. We found that the N gene of RABV has a low codon usage bias with a mean effective number of codons (ENC) value of 56.33 influenced by both mutation pressure and natural selection. However, neutrality analysis indicated that natural selection dominates over mutation pressure. Additionally, we found that dinucleotide bias partly contributed to RABV codon usage bias. On the other hand, based on the clades of phylogenetic tree, we found that the evolutionary rate of the Africa 2 clade was the highest with a mean value of $3.75x10^{-3}$ substitutions per site per year. Above all, our results regarding N gene of RABV codon usage will serve future RABV evolution research.

Key words: Rabies virus; Nucleoprotein; Evolution analysis; Phylogenetic analysis; Codon usage bias.

Introduction

Rabies, as a major viral fatal disease that has caused epidemics in 150 countries on every continent except for Antarctica, is an unheeded public health problem[1]. More than 55,000 human rabies-related deaths occur every year worldwide with most fatalities (almost 99%) found in developing countries[2, 3]. Notably, in Asia human rabies cases account for over 80% of the world total cases[4]. Rabies virus (RABV) is a member of the Rhabdoviridae, genus Lyssavirus[5]. The natural evolution of RABV provides an example of multiple host switches, which allows comparative studies of the evolutionary patterns, processes, and dynamics associated with host adaptation[6]. As previous studies showed, RABV isolates mainly cluster in two phylogenetic groups: the dog- and the bat-related RABV groups[7]. Compared to the bat-related isolates that mainly circulate in bats and rural skunks and raccoons[7, 8], the dog-related isolates can be found not only in domestic dogs worldwide but also in wildlife such as foxes in Europe, Middle East and Americas, raccoons in Asia, mongooses in Africa, and so on[9-11]. Importantly, dogs are the main virus reservoir and the major source for the dissemination of the disease[4].

The RABV genome is composed of a single RNA molecule of negative polarity that encodes for the nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G) and the large RNA dependent RNA polymerase (L)[12]. Previous research described that N, P, L are responsible for synthesis of the viral RNA[13] while M and G are essential for release and virus infectivity[14, 15]. During viral replication, the ribonucleoprotein (RNP) complex is formed and the N protein encapsidates the viral RNA. The viral polymerase complex that includes the P and L proteins, serves as a template for the transcription and replication of the viral RNA[16, 17]. The G protein is part of the lipid bilayer[15] and mediates infection of the host cell[18]. Importantly, the M protein is an important determinant of pathogenicity[19]. The L protein is a multifunctional protein[20] while