Saudi Journal of Biological Sciences (2014) 21, 179–183

King Saud University

Saudi Journal of Biological Sciences

www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

Evidence of selection pressures of neuraminidase gene (NA) of influenza A virus subtype H5N1 on different hosts in Guangxi Province of China

Youhua Chen [a,*], You-Fang Chen [b]

[a] Department of Zoology, University of British Columbia, Vancouver V6T 1Z4, Canada
[b] School of Software, Harbin Normal University, Heilongjiang Province, China

Received 24 August 2013; revised 17 September 2013; accepted 20 September 2013
Available online 2 October 2013

KEYWORDS

Positive selection;
Influenza A virus;
Host specificity;
Adaptive evolution

Abstract In the present study, the possible evidence of positive selection was analyzed for the neuraminidase (NA) sequences of Guangxi H5N1 strains of China. Based on an overall site-specific positive selection analysis, it was found that NA gene of H5N1 Guangxi strains underwent purifying selection and no significant positively selected sites were identified. For the branch-specific positive selection analysis, there was no positive selection evidence for the branches leading to different poultry hosts (chicken, duck and goose). Conclusively, positive selection seems not possible (if not rare) for the NA gene in influenza H5N1 subtype, at least for the samples found in Guangxi Province of China.

1. Introduction

Positive selection of virus genes could help elucidate how the genes are modified and adapted to the changing environment. The occurrence of positive selection is in particular likely when the viruses are transferred from original hosts to new hosts for adaptation. There are some advances in recent years contributing to search for the evidence of positive selection of some important genes for different viruses (Tang et al., 2008; Wang et al., 2009; Shen et al., 2011; Perez et al., 2012).

Subtype H5N1 of Influenza A virus is an economically important and highly pathogenic epidemic that could cause a great amount of mortality in poultry and animals (Bataille et al., 2011), and also cause illnesses on human beings occasionally (Kawaoka, 2012; Morens et al., 2012). The evolution of IAV genomes has been widely studied from different facets, including codon usage patterns (Zhou et al., 2005; Wong et al., 2010; Fancher and Hu, 2011; Goni et al., 2012), homologous recombination (Boni et al., 2008; He et al., 2009), and phylogenetics (Liu et al., 2009). In Asian countries, there are growing reports of patients being infected by H5N1. Consequently, understanding evolutionary forces and

* Corresponding author. Tel.: +1 5877786218.
E-mail address: haydi@126.com (Y. Chen).

Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

1319-562X © 2013 Production and hosting by Elsevier B.V. on behalf of King Saud University.
http://dx.doi.org/10.1016/j.sjbs.2013.09.009

adaptation of H5N1 would be of great help to understand its transmission and infection of animals and human beings (Kawaoka, 2012; Morens et al., 2012).

Neuraminidase (NA) is an important enzyme on the surface of influenza A virus particles which is widely used as the antigenic determinant. NA would help virus to be relaxed from the host cell, thus being important in the transmission of influenza viruses. As such, adaptive evolution of NA gene over different host species might be profound. Evolutionary studies on the NA gene have been widely carried out at different perspectives (Zhou et al., 2005; Han et al., 2010; Liu et al., 2010; Li et al., 2011). In particular, there are some progresses on reporting positive selection evidence of influenza A virus genes. For example, in a recent study, the evolutionary pressure of NA gene in influenza A virus subtype H1N1 has been carried out (Li et al., 2011). In another study, the selection pressure of HA gene was compared between pandemic (2009), human and swine influenza A strains (Furuse et al., 2010). In the present study, we would focus on the adaptive evolution of NA genes in Guangxi H5N1 strains for the purpose to throw new insights into the epidemiology and transmission mechanisms of influenza subtype H5N1 on the provincial scale in the southern part of China.

2. Materials and methods

2.1. Sequence data

93 NA gene sequences for influenza A virus subtype H5N1 found in Guangxi Province, China were extracted from the GenBank database (http://www.ncbi.nlm.nil.gov/; access date: Sep 10th, 2013). After removing partial sequences, 81 full ones are retained and aligned using MUSCLE program (Edgar, 2004). Ambiguous regions of the alignment were removed and the final aligned sequences were checked manually by BioEdit software (Hall, 1999). The aligned sequences were available upon request. The associated GenBank accession information is presented in Fig. 1.

2.2. Homologous recombination

By using the program RDP (Martin and Rybicki, 2010; Martin et al., 2010), homologous recombination signals were detected among the 81 sequences based on the seven recombination detection methods including GENECOV (Padidam et al., 1999), Bootscan/Rescan (Martin et al., 2005), Chimaera (Posada and Crandall, 2001), MaxChi (Maynard Smith, 1992), SiScan (Gibbs et al., 2000), 3Seq (Boni et al., 2007) and RDP (Martin and Rybicki, 2010). If there are putative recombinant sequences, they will be removed from the dataset since homologous recombination would bias the inference accuracy of phylogeny (Schierup and Hein, 2000; Anisimova et al., 2003).

2.3. Construction of phylogenetic tree for NA gene

The phylogenetic tree was constructed using the maximum likelihood method with PhyML software (Guindon et al., 2010). 1000 Bootstrapping replicates were applied to reveal the support of reconstructed tree. The tree was used in all subsequent positive selection analyses and presented in Fig. 1.

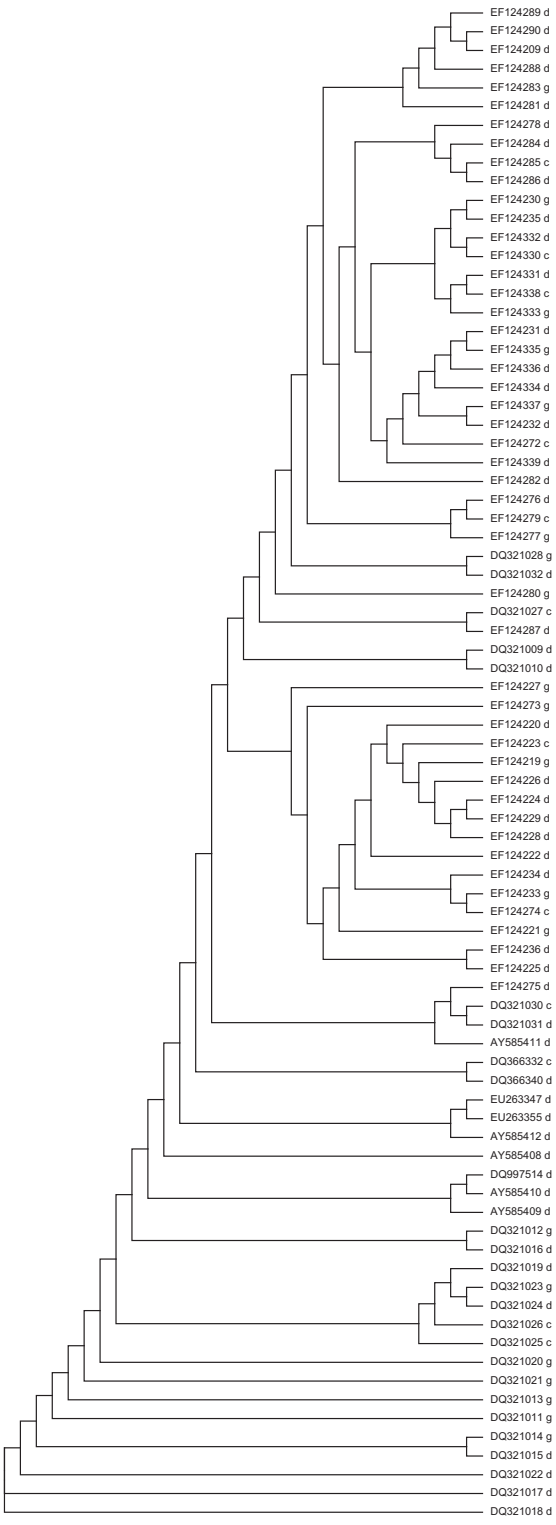

Figure 1 Maximum likelihood phylogenetic tree topology reconstructed for the 81 NA gene sequences sampled from Guangxi Province, China. This tree is used for all subsequent positive selection analyses and the labels indicated the GenBank accession numbers. At the end of these labels, the letters "d", "c" and "g" indicate the hosts of the virus are duck, chicken and goose respectively.

Table 1 Estimations of parameters for the six models of different ω among sites for the NA gene in 32 Guagnxi H5N1 samples.

Model	Parameter values	Likelihood	Positively selected sites (EBE support)
M0	$\omega = 0.215$	−4072.372	Not allowed
M1	$p_0 = 0.850$, $\omega_0 = 0.074$, $p_1 = 0.150$, $\omega_1 = 1.000$	−4042.562	Not allowed
M2	$p_0 = 0.850$, $\omega_0 = 0.074$, $p_1 = 0.150$, $\omega_1 = 1.000$, $p_2 = 0.000$, $\omega_2 = 38.186$	−4042.561	Not found
M3	$p_0 = 0.036$, $= 0.074$, $p_1 = 0.815$, $\omega_1 = 0.074$, $p_2 = 0.149$, $\omega_2 = 1.002$	−4042.561	Not found
M7	$p = 0.176$, $q = 0.665$	−4043.924	Not allowed
M8	$p_0 = 0.853$, $p = 8.18$, $q = 99$, $p_1 = 0.147$, $\omega = 1.008$	−4042.587	Not found

Table 2 Positive selection analysis for the lineages specific to different host species (chicken, duck, and goose) of NA genes in 81 Guangxi samples of China $\chi^2_{0.05,1} = 3.84$.

Hosts	Models	Likelihood	Parameter estimates	$2\Delta l$ (P value)	Positively selected sites (BEB support $\geqslant 0.95$)
Chicken	Null	−4042.562	$\omega_0 = 0.074$, $\omega_1 = 1$, $\omega_2 = 1$, $p_0 = 0.850$, $p_1 = 0.150$, $p_{2a} = 0.0$, $p_{2b} = 0.0$		Not allowed
	Alternative	−4042.562	$\omega_0 = 0.074$, $\omega_1 = 1$, $\omega_2 = 3.949$, $p_0 = 0.850$, $p_1 = 0.150$, $p_{2a} = 0.0$, $p_{2b} = 0.0$	0 ($P > 0.05$)	Not found
Duck	Null	−4042.001	$\omega_0 = 0.065$, $\omega_1 = 1$, $\omega_2 = 1$, $p_0 = 0.807$, $p_1 = 0.136$, $p_{2a} = 0.048$, $p_{2b} = 0.008$		Not allowed
	Alternative	−4042.001	$\omega_0 = 0.065$, $\omega_1 = 1$, $\omega_2 = 1$, $p_0 = 0.807$, $p_1 = 0.136$, $p_{2a} = 0.048$, $p_{2b} = 0.008$	0 ($P > 0.05$)	Not found
Goose	Null	−4041.569	$\omega_0 = 0.067$, $\omega_1 = 1$, $\omega_2 = 1$, $p_0 = 0.781$, $p_1 = 0.136$, $p_{2a} = 0.070$, $p_{2b} = 0.012$		Not allowed
	Alternative	−4041.489	$\omega_0 = 0.067$, $\omega_1 = 1$, $\omega_2 = 1.804$, $p_0 = 0.813$, $p_1 = 0.141$, $p_{2a} = 0.039$, $p_{2b} = 0.007$	0.08 ($P > 0.05$)	Not found

2.4. Site-specific positive selection analysis

Identification of positive selected sites required the construction of phylogenetic trees. As such, we constructed maximum likelihood tree using the program PhyML (Guindon et al., 2010) with 1000 bootstrapping replicates. The constructed tree and aligned sequences are then subjected to the positive selection analysis using the codeml program from the PAML package (Yang, 1997, 2007). Several site-specific models (M0, M1, M2, M3, M7 and M8) which tested different models were used for the detection of positive selected sites (Tang et al., 2008; Wang et al., 2009).

Among the models, M0 and M3, M1 and M2, M7 and M8 are pairs of models, in which the former ones are null models, while the latter ones allow the existence of positive selection. The likelihood ratio test (LRT) could be implemented to test these alternative nested models to identify the possibility of positive selected sites. If the sites are subjected to positive selection, then the nonsynonymous/synonymous substitution ratio $\omega = dN/dS$ would be larger than 1, otherwise $0 < \omega < 1$ for purifying selection and $\omega = 1$ for neutral selection. A χ^2 distribution could be used for testing significance. As expected, if positive selection is in function, then we would observe $\omega > 1$ and the likelihood for the positive selection models (M3, M2 and M8) would be significantly higher than that of null models. The Bayes Empirical Bayes (BEB) calculation of posterior probabilities for site classes was implemented to calculate the probabilities of sites under positive selection (Yang et al., 2005; Tang et al., 2008). Only those sites with BEB $\geqslant 0.95$ are considered to be positively selected.

2.5. Clade-specific positive selection analysis

Clade-specific positive selection analysis could allow one to test the different selection pressures on different clades in the phylogenetic tree. Clade-specific models have been widely applied in recent studies (Yu et al., 2011). We used clade-site Model A for identifying sites under positive selection along the lineages that we are interested in (Yu et al., 2011). Analogous to site-specific positive selection analysis, the likelihood ratio test (LRT) could be applied to detect the significance of difference between alternative and null models using χ^2 distribution with the number of degrees of freedom (equal to the difference between the free parameter numbers of the pairwise models).

3. Results and discussion

3.1. Detection of recombination signals

No potential recombination events were found in NA sequences of Guangxi H5N1 strains, as consistently identified by all the seven methods. As such, our result further evidenced that recombination events in influenza viruses are very rare, consistent with previous studies (Boni et al., 2008; Han et al., 2008, 2010; Han and Worobey, 2011).

3.2. Site-specific positive selection analysis

As indicated by the nested model comparison (M0 versus M3, M1 versus M2, M7 versus M8), basically no increments in terms of likelihood values are found for the models allowing positive selection in comparison to alternative ones (Table 1). There are no positively selected codon sites for NA gene of Guangxi H5N1 strains globally.

3.3. Branch-specific positive selection analysis

Among the NA lineages that are specific to three poultry hosts, the log-likelihoods for alternative models A1 which allows $\omega > 1$ did not have a significant difference to those from the null models (Table 2). Thus, positive selection is unlikely to occur among the poultry hosts of NA gene in Guangxi Province of China.

3.4. Adaptive evolution of NA gene in Guangxi Province of China

As indicated by the above analyses, NA gene seems to undergo purifying selection as no positively selected sites have been identified for either overall sequences or specific branches. Several mechanisms may lead to such an observation. First, the NA sequences sampled from the region are still largely lacked. In our study, only 81 full sequences are available, which thus may limit the divergence of the NA gene across different poultry strains. The relatively rare divergence among NA sequences would lead to short evolutionary distances of these sequences, reducing the occurrence probability of positive selection. Second, our study is only working on one subtype (H5N1) influenza A viruses, which might not fully be legitimate because the primary purpose of the study is to detect the adaptive evolution of NA gene. Thus, NA sequences from different subtypes should be utilized in the analysis so as to better reflect the evolution of NA gene. Moreover, the classification of different influenza A viruse subtypes is based on the different groupings of NA and HA (hemagglutinin) genes. Therefore, adaptive evolution is highly desired for NA (or HA) gene when the sequences sampled from different subtypes are considered given that the sequences for different subtypes should show distinct clades.

4. Conclusions

No positive selection evidence was identified for the NA gene in Guangxi H5N1 strains. Site-specific likelihood codon models did not identify any positively selected sites globally. Branch-specific analyses which focused on the branches for different poultry lineages (duck, chicken and goose) further confirmed such a statement.

References

Anisimova, M., Nielsen, R., Yang, Z., 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164, 12291–12361.

Bataille, A., van der Meer, F., Stegeman, A., Koch, G., 2011. Evolutionary analysis of inter-farm transmission dynamics in a highly pathogenic avian influenza epidemic. PLoS Pathogens 7, e1002094.

Boni, M., Posada, D., Feldman, M., 2007. An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics 176, 1035–1047.

Boni, M., Zhou, Y., Taubenberger, J., Holmes, E., 2008. Homologous recombination is very rare or absent in human influenza A virus. Journal of Virology 82, 4807–4811.

Edgar, R., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32, 1792–1797.

Fancher, K., Hu, W., 2011. Codon bias of influenza a viruses and their hosts. American Journal of Molecular Biology 1, 174–182.

Furuse, Y., Shimabukuro, K., Odagiri, T., Sawayama, R., Okada, T., Khandaker, I., Suzuki, A., Oshitani, H., 2010. Comparison of selection pressures on the HA gene of pandemic (2009) and seasonal human and swine influenza A H1 subtype viruses. Virology 405, 314–321.

Gibbs, M., Armstrong, J., Gibbs, A., 2000. Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16, 573–582.

Goni, N., Iriarte, A., Comas, V., Sonora, M., Moreno, P., Moratorio, G., Musto, H., Cristina, J., 2012. Pandemic influenza A virus codon usage revisited: biases, adaptation and implications for vaccine strain development. Virology Journal 9, 263.

Guindon, S., Dufayard, J., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology 59, 307–321.

Hall, T., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41, 95–98.

Han, G., Worobey, M., 2011. Homologous recombination in negative sense RNA viruses. Viruses 3, 1358–1373.

Han, G., Liu, X., Li, S., 2008. Homologous recombination is unlikely to play a major role in influenza B virus evolution. Virology Journal 5, 65.

Han, G., Boni, M., Li, S., 2010. No observed effect of homologous recombination on influenza C virus evolution. Virology Journal 7, 227.

He, C., Xie, Z., Han, G., Dong, J., Wang, D., Liu, J., Ma, L., Tang, X., Liu, X., Pang, Y., Li, G., 2009. Homologous recombination as an evolutionary force in the avian influenza A virus. Molecular Biology and Evolution 26, 177–187.

Kawaoka, Y., 2012. H5N1: flu transmission work is urgent. Nature 482, 155.

Li, W., Shi, W., Qiao, H., Ho, S., Luo, A., Zhang, Y., Zhu, C., 2011. Positive selection on hemagglutinin and neuraminidase genes of H1N1 influenza viruses. Virology Journal 8, 183.

Liu, S., Ji, K., Chen, J., Tai, D., Jiang, W., Hou, G., Li, J., Huang, B., 2009. Panorama phylogenetic diversity and distribution of type A influenza virus. PLoS ONE 4, e5022.

Liu, X., Wu, C., Chen, A., 2010. Codon usage bias and recombination events for neuraminidase and hemagglutinin genes in Chinese isolates of influenza A virus subtype H9N2. Archives of Virology 155, 685–693.

Martin, D., Rybicki, 2010. RDP: detection of recombination amongst aligned sequences. Bioinformatics 26, 2462–2463.

Martin, D., Posada, D., Crandall, K., Williamson, C., 2005. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Research and Human Retroviruses 21, 98–102.

Martin, D., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuvre, P., 2010. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26, 2462–2463.

Maynard Smith, J., 1992. Analyzing the mosaic structure of genes. Journal of Molecular Evolution 34, 126–129.

Morens, D., Subbarao, K., Taubenberger, J., 2012. Engineering H5N1 avian influenza viruses to study human adaptation. Nature 486, 335–340.

Padidam, M., Sawyer, S., Fauquet, C., 1999. Possible emergence of new geminiviruses by frequent recombination. Virology 265, 218–225.

Perez, L., Arce, H., Perera, C., Rosell, R., Frias, M., Percedo, M., Tarradas, J., Doinguez, P., Nuenez, J., Ganges, L., 2012. Positive selection pressure on the B/C domains of the E2-gene of classical swine fever virus in endemic areas under C-strain vaccination. Infection, Genetics and Evolution 12, 1405–1412.

Posada, D., Crandall, K., 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proceedings of the National Academy of Sciences 98, 13757–13762.

Schierup, M., Hein, J., 2000. Consequences of recombination on traditional phylogenetic analysis. Genetics 156, 879–891.

Shen, H., Pei, J., Bai, J., Zhao, M., Ju, C., Yi, L., Kang, Y., 2011. Genetic diversity and positive selection analysis of classical swine fever virus isolates in south China. Virus Genes 43, 234–242.

Tang, F., Pan, Z., Zhang, C., 2008. The selection pressure analysis of classical swine fever virus envelope protein genes Erns and E2. Virus Research 131, 132–135.

Wang, D., Fan, W., Han, G., He, C., 2009. The selection pressure analysis of chicken anemia virus structural protein gene VP1. Virus Genes 38, 259–262.

Wong, E., Smith, D., Rabadan, R., Peiris, M., Poon, L., 2010. Codon usage bias and the evolution of influenza A viruses. BMC Evolutionary Biology 10, 253.

Yang, Z., 1997. PAML: a program package for phylogenetic analysis by amximum likelihood. Computer Applications in the Biosciences 13, 555–556.

Yang, Z., 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution 24, 1586–1591.

Yang, Z., Wong, W., Nielsen, R., 2005. Bayes empirical Byaes inference of amino acids sites under positive seleciton. Molecular Biology and Evolution 22, 1107–1118.

Yu, L., Jin, W., Zhang, X., Wang, D., Zheng, J., Yang, G., Xu, X., Cho, S., Zhang, Y., 2011. Evidence for positive selection on the Leptin gene in Cetacea and Pinnipedia. PLoS ONE 6, E26579.

Zhou, T., Gu, W., Ma, J., Sun, X., Lu, Z., 2005. Analysis of synonymous codon usage in H5N1 virus and other influenza A viruses. BioSystems 81, 77–86.