Infection, Genetics and Evolution 22 (2014) 192–200

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Molecular phylogeny and evolutionary dynamics of influenza A nonstructural (NS) gene

Jianpeng Xu [a], Haizhen A. Zhong [b], Alex Madrahimov [a,c], Tomáš Helikar [c], Guoqing Lu [a,*]

[a] Department of Biology, University of Nebraska at Omaha, Omaha, NE 68182, United States
[b] Department of Chemistry, University of Nebraska at Omaha, Omaha, NE 68182, United States
[c] Department of Mathematics, University of Nebraska at Omaha, Omaha, NE 68182, United States

ARTICLE INFO

Article history:
Available online 22 October 2013

Keywords:
Influenza A virus
Nonstructural (NS) gene
Phylogeny
Lineage assignment
Evolutionary dynamics

ABSTRACT

While the nonstructural gene (NS) of the influenza A virus plays a crucial role in viral virulence and replication, the complete understanding of its molecular phylogeny and evolutionary dynamics remains lacking. In this study, the phylogenetic analysis of 7581 NS sequences revealed ten distinct lineages within alleles A and B: three host-specific (human, classical swine and equine), two reassortment-originated (A(H1N1)pdm09 and triple reassortment swine), one transmission-originated (Eurasian swine), and two geographically isolated avian (Eurasian/Oceanian and North American) for allele A and two geographically isolated avian (Eurasian/Oceanian and North American) for allele B. The average nucleotide substitution rates of the lineages range from 1.24×10^{-3} (equine) to 4.34×10^{-3} (A(H1N1)pdm09) substitutions per site per year. The selection pressure analysis demonstrated that the d_N/d_S ratio of the NS gene in A(H1N1)pdm09 lineage was higher than its closely related triple reassortant swine, which could be attributed to the adaptation to the new host and/or intensive surveillance after the inter-species transmission from swine to human. The positive selection sites were found in all lineages except the equine lineage and mostly in the NS1 region. The positive selection sites 22, 26, 226, 227 and 230 of the human lineage are significant because these residues participate in either forming the dimerization of the two RNA binding domain (RBD) monomers or blocking the replication of host genes. Residues at position 171 provide hydrophobic interactions with hydrophobic residues at p85β and thus induce viral cell growth. The lineages and evolutionary dynamics of influenza A NS gene obtained in this study, along with the studies of other gene segments, are expected to improve the early detection of new viruses and thus have the potential to enhance influenza surveillance.

1. Introduction

Influenza A virus is an important pathogen of respiratory infection in human and other animals (Guan et al., 2010; Holmes, 2010; Webster et al., 1992). The virus possesses a segmented genome of eight negative-sense, single-stranded RNAs. Each of the influenza viral genes plays a significant role in viral interactions with hosts and subsequent infection; therefore, understanding the evolution and phylodynamics of each viral gene can provide new insights into the epidemiological dynamics of influenza viruses. Large-scale evolutionary dynamics analyses have been conducted in several segments, including nucleoprotein (NP), neuraminidase (NA), and matrix protein (M) (Fourment et al., 2010; Furuse et al., 2009; Xu et al., 2011, 2012).

The NS gene plays a crucial role in virus replication and virulence. The NS segment encodes both the nonstructural protein NS1, found only in infected cells (Krug and Etkind, 1973), and the nuclear export protein (NS2/NEP) (Lamb et al., 1980). NS2 supports export of the viral genome from the nucleus (Neumann et al., 2000) as well as regulation of virus genome transcription and replication (Robb et al., 2009). The NS1 protein, which contains an N-terminal double-stranded RNA-binding domain (RBD) and a C-terminal effector domain, is implicated in the pathogenesis and virulence of the influenza virus (Hale et al., 2008). The major role related to NS1 is its inhibition of host immune responses, especially the limitation of both interferon (IFN) production and the antiviral effects of IFN-induced proteins, such as dsRNA-dependent protein kinase R (PKR) and 2′,5′-oligoadenylate synthetase (OAS)/RNase L (Ramos et al., 2013).

NS1 also directly modulates other important aspects of the virus replication cycle, including viral RNA replication, viral protein synthesis, and general host-cell physiology. Furthermore, NS1 (as well as other viral proteins, including PB2 and NP) interacts with

* Corresponding author. Tel.: +1 402 5543195; fax: +1 402 5543532.
E-mail addresses: jianpenxu@unomaha.edu (J. Xu), hzhong@mail.unomaha.edu (H.A. Zhong), amadrahimov@unomaha.edu (A. Madrahimov), thelikar@mail.unomaha.edu (T. Helikar), glu3@unomaha.edu (G. Lu).

1567-1348/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2013.10.011

the host molecules in order to allow for more efficient replication of the virus by both activating and inhibiting the host's various biochemical pathways (Donelan et al., 2003; Ehrhardt et al., 2007; Li et al., 2006; Ma et al., 2010; Noah et al., 2003). Having multiple roles within the virus life cycle, a live influenza intra-respiratory vaccine with truncated NS1 has recently been generated, demonstrating the possibility of novel immunoprophylactic design (Baskin et al., 2007). On the other hand, NS protein exhibits much less variations than HAs/NAs and is considered as a potential therapeutic target for antiviral therapy (Twu et al., 2006). Therefore, understanding of evolution of the NS gene is of great importance and practical relevance.

Molecular phylogeny and evolution analyses of the NS gene have been conducted in previous studies (Chen et al., 2009; Chen and Holmes, 2006; Christman et al., 2011; Jahangir et al., 2012; Sevilla-Reyes et al., 2013; Smith et al., 2009b). However, the evolutionary dynamics of the NS gene (e.g., the time of most common ancestor, selection, and substitution rate) on a large scale remains less well-understood. In this study, a large-scale phylogenetic analysis of NS sequences was conducted to infer evolutionary relationships of influenza A viruses. The Bayesian modeling and Markov Chain Monte Carlo (MCMC) sampling methods were employed to estimate the substitution rates and the time of most recent common ancestor (TMRCA) of different lineages. We also estimated non-synonymous to synonymous substitution rate ratios (d_N/d_S ratio) as well as investigated the positive and negative selection sites for each lineage. Protein structures were modeled to show the found selection sites and their potential significance in influenza surveillance and antiviral drug development.

2. Materials and methods

2.1. Sequence data

Influenza A NS gene sequences as well as the information of host, subtype, isolation time, and isolation location were downloaded from GenBank (Bao et al., 2008). After excluding duplicated sequences or sequences from laboratory strains, a total of 7581 sequences (longer than 810 nucleotides) were obtained.

2.2. Phylogenetic analysis and lineage assignment

Sequence alignment was created using MUSCLE program (Edgar, 2004) followed by manual adjustment using TranslatorX (Abascal et al., 2010). Phylogenetic analysis was conducted using the maximum likelihood (ML) method in RAxML (Stamatakis et al., 2005). RAxML uses rapid algorithms for bootstrap and maximum likelihood searches and is considered one of the fastest and most accurate phylogeny programs for large datasets. Analyses of 1000 bootstrap replicates were performed using GTR-GAMMA, the GTR model of nucleotide substitution with the Gamma model of rate heterogeneity. Lineages were defined based upon the groupings with strong bootstrap support in the ML tree. Additional information such as hosts and geographic regions were also considered in the classification. Two alleles (A and B) were revealed for the NS gene in early studies and such nomenclature has been followed broadly by influenza virologists (Kawaoka et al., 1998; Treanor et al., 1989). We followed this convention to assign lineages respectively in each of the two alleles. The phylogenetic trees were visualized and color-coded with FigTree (http://www.tree.bio.ed.ac.uk/software/figtree/).

2.3. Substitution rate and the time of most recent common ancestor (TMRCA) analyses

The substitution rate and TMRCA were estimated using a Bayesian Markov Chain Monte Carlo (MCMC) method as implemented in BEAST (Drummond and Rambaut, 2007). In all cases, the GTR + Γ_4 nucleotide substitution model was employed as this was the best-fit model supported by Modeltest (Posada and Crandall, 1998). For each analysis the Bayesian skyline coalescent model was used, as it describes the fluctuating population dynamics characteristics of influenza virus (Rambaut et al., 2008). Three clock models were compared statistically for each dataset using a Bayes factor test in the Tracer program (Suchard et al., 2001): a strict clock, an uncorrelated lognormal relaxed clock (UCLD) and an uncorrelated exponential relaxed clock (UCED) (Drummond et al., 2006). The UCED model was found to provide the best fit for all lineages (Bayes factor >30). We thus used this model for the estimation of evolutionary dynamic parameters. In each case, MCMC chains were run for sufficient time to achieve convergence, with uncertainty in parameter estimates reflected in the 95% highest probability density (HPD). The maximum clade credibility (MCC) tree across all plausible trees was then computed from the BEAST trees using TreeAnnotator, with the first 10% of trees removed as burn-in.

2.4. Measurement of selection pressures

To examine the selection experienced by each lineage, we estimated the ratio of non-synonymous (d_N) to synonymous (d_S) substitutions per site (ratio d_N/d_S) for each lineage, using the HYPHY package (Pond et al., 2005). Positively selected codons were detected using the single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL) and internal fixed effects likelihood (IFEL) methods with a significance level of 0.05. In the SLAC method, the nucleotide and codon model parameter estimates were used to reconstruct the ancestral codon sequences at the internal nodes of the tree. The single most likely ancestral sequences were then fixed as known variables, and applied to infer the expected number of non-synonymous or synonymous substitutions that have occurred along each branch, for each codon position. The FEL method is based on maximum-likelihood estimates. The FEL method estimates the ratio of non-synonymous to synonymous substitutions on a site-by-site basis for the entire tree or only the interior branches (IFEL).

2.5. Protein structural analyses

To map positive selection sites onto the protein structures and to better understand how the positive selection sites may interact with other proteins, we carried out structural analyses of NS1 proteins using molecular Operating environment (MOE) (http://www.chemcomp.com/). The NS1 protein contains an N-terminal double-stranded RNA-binding domain (RBD) and a C-terminal effector domain. NS1 protein with double-strand RNA (PDB ID: 2ZKO) (Cheng et al., 2009) was downloaded from the Protein Databank (www.pdb.org). Proteins for studying C-terminal domain interactions were downloaded from PDB website with PDB IDs: 3L4Q (NS1/p85β) (Hale et al., 2010), 3RVC (Kerry et al., 2011), 3KWI (Xia and Robertus, 2010), and 3EE8 (Xia et al., 2009). These four C-terminal domain proteins were structurally superposed in MOE. The α-carbons (Cα) of positive selection residues were highlighted in space-filling model (balls).