Infection, Genetics and Evolution 27 (2014) 234–243

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Genetic evolution of PB1 in the zoonotic transmission of influenza A(H1) virus

Marta Gíria [a], Helena Rebelo de Andrade [a,b,*]

[a] Centro de Patogénese Molecular, Unidade dos Retrovírus e Infecções Associadas, Instituto de Medicina Molecular e Instituto de Investigação do Medicamento (iMed.ULisboa), Faculdade de Farmácia, Universidade de Lisboa, Lisboa, Portugal
[b] Instituto Nacional de Saúde Dr Ricardo Jorge IP, Lisboa, Portugal

ARTICLE INFO

Article history:
Received 4 June 2014
Received in revised form 16 July 2014
Accepted 21 July 2014
Available online 1 August 2014

Keywords:
Influenza A(H1)
Interspecies transmission
Polymerase basic-protein 1
Genetic evolution
Viral adaptation

ABSTRACT

The epidemiology of human infection with swine-origin influenza A(H1) viruses suggests that the virus must adapt to replicate and transmit within the human host. PB1 is essential to the replication process. The objective of this study was to identify whether PB1 retains genetic traces of interspecies transmission and adaptation. We have found that the evolutionary history of PB1 is traceable. Lineage appears to be distinguished by amino acid changes between the conserved motifs of the viral polymerase, which can have major impact in PB1 protein folding, and by changes in the expression of the *Mitochondrial Targeting Sequence* and in the predicted helical region, that putatively affect induction of cellular apoptosis by PB1-F2. Furthermore, we found genomic markers that possibly relate to viral adaptation to new hosts and to new cellular environment and, additionally, to an enhanced compatibility with HA. We found no specific trend in the amino acid substitutions. Viral fitness appears to be favored by less reactive amino acids in some positions, while in others more reactive ones are fixed. Also, more flexible conformations appear associated with higher protein stability in general, although often more restrictive conformations appear to have favored protein folding and binding. Several aspects of PB1 mapping domains and the specific roles and interaction of PB1, PB1-F2 and N40 with each other and with other viral proteins and host cellular molecules remain unclear. Tracing the genetic evolution is critical to further understand the mechanisms by which PB1 affects vital fitness and adaptation. This analysis now permits putative adaptive related polymorphisms to be experimentally evaluated for phenotypic impact.

1. Introduction

The reservoir of aquatic birds sporadically introduces avian-origin viruses into mammalian hosts and interspecies transmission occurs between the swine and human populations. When crossing the species barrier, adaptation is mostly driven by natural selection and selective sweeps (Ding et al., 2009; Bhatt et al., 2013). Within the new host, adaptation mainly occurs by purifying selection (Ding et al., 2009; Bhatt et al., 2013). PB1, as an essential player in the replication process, undergoes genetic changes through the process of viral adaptation.

Human infections with swine influenza A(H1) virus resulting in un-sustained human-to-human transmission have been documented worldwide from 1970 to the 2009 pandemics (Shinde et al., 2009; Zimmer and Burke, 2009; Dawood et al., 2009). These zoonotic viruses are designated as swine-origin influenza A(H1), SOIV A(H1). During their evolutionary history, reassortment events and interspecies transmission have placed PB1 into new viral genomic backgrounds and new host cellular environments. The extent to which PB1 retains genetic traces of interspecies transmission and adaptation is unknown. Questions arise as: are there genetic markers that outline the lineage and host origin of PB1 segment, within a particular virus? Is the genetic evolution of PB1 towards viral adaptation traceable at the amino acid (aa) level? Is it possible to identify genetic markers for (a) viral adaptation to new host cellular environments, and (b) adaptation of PB1 to new genomic backgrounds, following reassortment events? In this study, we propose to trace the genetic evolution of PB1 of swine viruses that have infected the human host and infer its putative role in fitness and host adaptation, in view of the molecular epidemiology and evolutionary history of the viruses.

1.1. Role of PB1 genomic segment in viral fitness

The role of PB1 genomic segment is believed to be diversified and determinant in replication and induction of apoptosis. The

* Corresponding author at: Instituto Nacional de Saúde Dr Ricardo Jorge IP, Lisboa, Portugal. Tel.: +351 21 7508159.
E-mail address: h.rebelo.andrade@insa.min-saude.pt (H. Rebelo de Andrade).

http://dx.doi.org/10.1016/j.meegid.2014.07.024
1567-1348/

segment encodes three proteins, PB1, PB1-F2 and N40. PB1 protein is responsible for the recognition of vRNA and initiation and elongation of cDNA and mRNA in viral transcription and replication. Interferences with the binding domains and the conserved motif of PB1 are specific targets for new antiviral research (Perez and Donis, 2001; Reuther et al., 2011; Chu et al., 2012). PB1-F2 is encoded in OFR+1 and exclusively found in infected cells. It has been associated with the induction of cellular apoptosis at a late stage of infection, which is supportive of viral replication and infectious particles release. Also, it is able to promote inflammation and it has been shown to up-regulate polymerase activity by interacting with PB1 protein (Krumbholz et al., 2011). N40 in an N-terminal truncated form of PB1. It retains the ability to bind PB2 but is unable to bind PA. It has been reported as not essential to virus survival. However, polymerase activity is significantly reduced in the absence of N40, it and even further if PB1-F2 is also absent, although it seems not to be affected by the loss of PB1-F2 alone. On the other hand, the over-expression of N40 in the absence of PB1-F2 has been associated with a shift from transcription to replication and is thought to be regulated by the accumulation of the different RNA species (Vater, 2011). Although new information regarding PB1, PB1-F2 and N40 is constantly being uncovered, several aspects of their specific roles and of their interaction with each other and with other viral proteins and host cellular molecules remain unclear. Namely, in the history of influenza virus classical reassortments, the acquisition of PB1 protein together with surface glycoproteins is a recurrent event and thereby thought to confer a biological advantage in natural selection by increasing the viral fitness (Wanitchang et al., 2010; Abt et al., 2011; Nelson et al., 2008; Khiabanian et al., 2009; Bergeron et al., 2010). Although it remains unclear as to how, the profile of gene segregation in reassortment events suggests that a functional compatibility between PB1 and HA enhances viral fitness.

1.2. Molecular epidemiology of human infections with swine-origin influenza A(H1) viruses

Sporadic cases and clusters of human infection with SOIV A(H1) have been identified in the past years, resulting in unsustained human to human transmission. From 1970–2000, over 50 cases of human infection with SOIV have been reported worldwide, mainly by A(H1N1) from the classic swine North American lineage. This was the predominant lineage isolated from pigs until the late nineties, with very little genetic change. It originates from avian A(H1N1) viruses, thought to be introduced in the swine population by interspecies transmission at the same time as they emerged in the human population in 1918 causing a pandemic. These SOIV A(H1) then share the genetic background of the avian A(H1N1) 1918 virus and the seasonal A(H1N1) descendents (Shinde et al., 2009; Zimmer and Burke, 2009). In the swine population, multiple strains of Triple-Reassortant swine influenza A virus (TR-SIV) then emerged and became dominant in North America, as a result of a triple reassortment event between the classic swine North American, avian North American and seasonal A(H3N2) lineages (Shinde et al., 2009). The internal genes derive from swine (M, NS and NP), human (PB1) and avian viruses (PA, PB2) and this particular combination is designated as triple-reassortant internal genes (TRIG) cassette. The TRIG cassette is very tolerant to antigenic glycoproteins and has been associated with other subtypes of swine virus (H3N2, H1N2). It is extremely stable and assumed to confer a selective advantage to the virus (Ma et al., 2010). Since 2005, there have been 11 notifications of sporadic human infections with Triple-Reassortant swine-origin influenza A(H1), TR-SOIV A(H1).

A(H1N1)pdm09 then emerged in the human host, in 2009, and caused a pandemic. This emerging SOIV, although a A(H1N1)subtype, was genetically different from the previous swine A(H1N1) and TR-SIV A(H1) isolated from the human host. Its proposed origin was a reassortment event in which the backbone of TR-SIV acquired M and NA genomic segments from an Eurasian swine lineage (Dawood et al., 2009). The reassortment is presumed to have occurred in the swine population and to have suffered a long evolutionary process before the interspecies transmission to the human host. This period is phylogenetically estimated in up to 10 years and the introduction of the progeny virus occurred in single or multiple events of genetically closely related strains (Ding et al., 2009).

The epidemiology of human infections with swine influenza virus is dependent on environmental factors such as exposure, but it also reflects the genetic ability of the virus to infect the human host. Although sporadic infections have occurred, transmission among humans has been very limited until the 2009 pandemics, suggesting that the virus must adapt to replicate and transmit within the new host.

2. Methods

2.1. Study sample

This study was performed with PB1 nucleotide sequences accessed from the Influenza Virus Resource database at www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html and GISAID's EpiFlu™ database at www.gisaid.org.

The data set of SOIV included PB1 sequences from 8 isolates of SOIV A(H1) that have infected the human host and from 55 isolates of A(H1N1)pdm09 from the pandemic period, with worldwide distribution. The study sample of 8 SOIV A(H1) constitutes the entire set for which there are published PB1 sequences.

For the purpose of phylogeny and mutation trend analysis, the study sample additionally included 13 A(H1N1)pdm09 worldwide isolates from 2010/2011. For the putative adaptive mutation analysis, SOIV and A(H1N1)pdm09 sequences were also evaluated against 19 seasonal A(H1N1) and 13 seasonal A(H3N2) isolates from 2009, with worldwide distribution, and their ancestors reference strains for the previous pandemics of 1918, 1957 and 1968 and reference strain for A(H1N1) reemergence in 1977. Strains designation and accession numbers are listed in Table S1.

2.2. Phylogeny and mutation trend analysis

Nucleotide sequence alignment was performed by *ClustalW*, *Mega5.2*. Phylogeny was analyzed for the PB1 genomic segment exclusively in what concerns the PB1 protein coding region, since PB1-F2 protein is coded in different truncated forms that compromise the phylogenetic analysis. The phylogenetic tree of PB1 was constructed in PhyML, *Seaview*, using the model GTR+I selected by *JModelTest* software.

Within the branches of PB1 phylogeny, genetic analysis was performed for PB1 and PB1-F2 coding regions. For the purpose of this analysis, residues that were found exclusively in a particular lineage or host origin are indicated as putative genetic markers for that origin.

Viral RNA is used for non-coding functions such as packaging signals and promoter-related activities and, consequently, genetic mutations in the coding region of the protein may not be directly related to protein function. In our analysis of polymorphisms which have arisen and persisted in particular influenza virus lineages, however, residues considered putatively associated with viral adaptation on the basis of molecular epidemiology or amino acid substitution were identified as putative markers for adaptation.