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/© 2014 Elsevier B.V. All rights reserved.

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.

3. Results and discussion

3.1. Phylogeny and evolutionary history

3.1.1. SOIV isolates of 1976 and 1988 are phylogenetically closely related to the 1918 pandemic reference strain and A(H1N1)seasonal virus

SOIV isolated in 1976 and 1988 are genetically divergent from Triple-Reassortment swine-origin influenza viruses (TR-SOIV) since they do not share the TRIG cassette. They have been isolated from humans previously to the emergence of TR-SOIV and present PB1 phylogenetically most closely related to the 1918 pandemic reference strain and seasonal A(H1N1) descendants (Fig. 1). Based on the phylogenetic relation and the historical molecular epidemiology discussed above, we propose SOIV and seasonal A(H1N1) to have evolved in the swine and human hosts, respectively, from a common avian ancestor.

Both SOIV strains code for a truncated non functional form of 11aa PB1-F2 (Table 2), similar to the N-terminal 11aa of the 1918 pandemic virus and the human adapted progeny PR8 (Fig. 2). Typically, all avian viruses code for a full length PB1-F2 and often this protein appears in a truncated form in swine and human adapted strains (McAuley et al., 2010). The adaptation of avian viruses to the mammalian host has in fact been proposed to include a truncation of PB1-F2, suggesting that it may not be crucial for effective transmission within these new hosts (Krumbholz et al., 2011; McAuley et al., 2010). We found that, in the human host, PR8 has evolved from its precedent 1918 pandemic virus to present a Tryptophan-Stop change, resulting in a truncated functional form of 87aa (Fig. 2). When A(H1N1) reemerged in 1977, however, PB1-F2 was coded in a shorter non-functional form of 57aa, resulting again from a Tryptophan-Stop change. Again based on phylogeny, genetics and the known historical molecular epidemiology, we presume this 1977 A(H1N1) to be the precursor for seasonal A(H1N1) viruses, that have then evolved in the human host to further accumulate genetic changes (Fig. 2).

A particular Isoleucine/Leucine, I/L, pattern at positions 10/11 is found in PB1-F2 of 1976 and 1988 isolates, 1918 pandemic strain and PR8. (Table 2, Fig. 2). Residue 10I appears to be a marker for classic North American A(H1N1) lineage, since it is also present in seasonal A(H1N1) . Residue 11L, however, is exclusively present in these four strains and could refer to an early A(H1N1) residue,

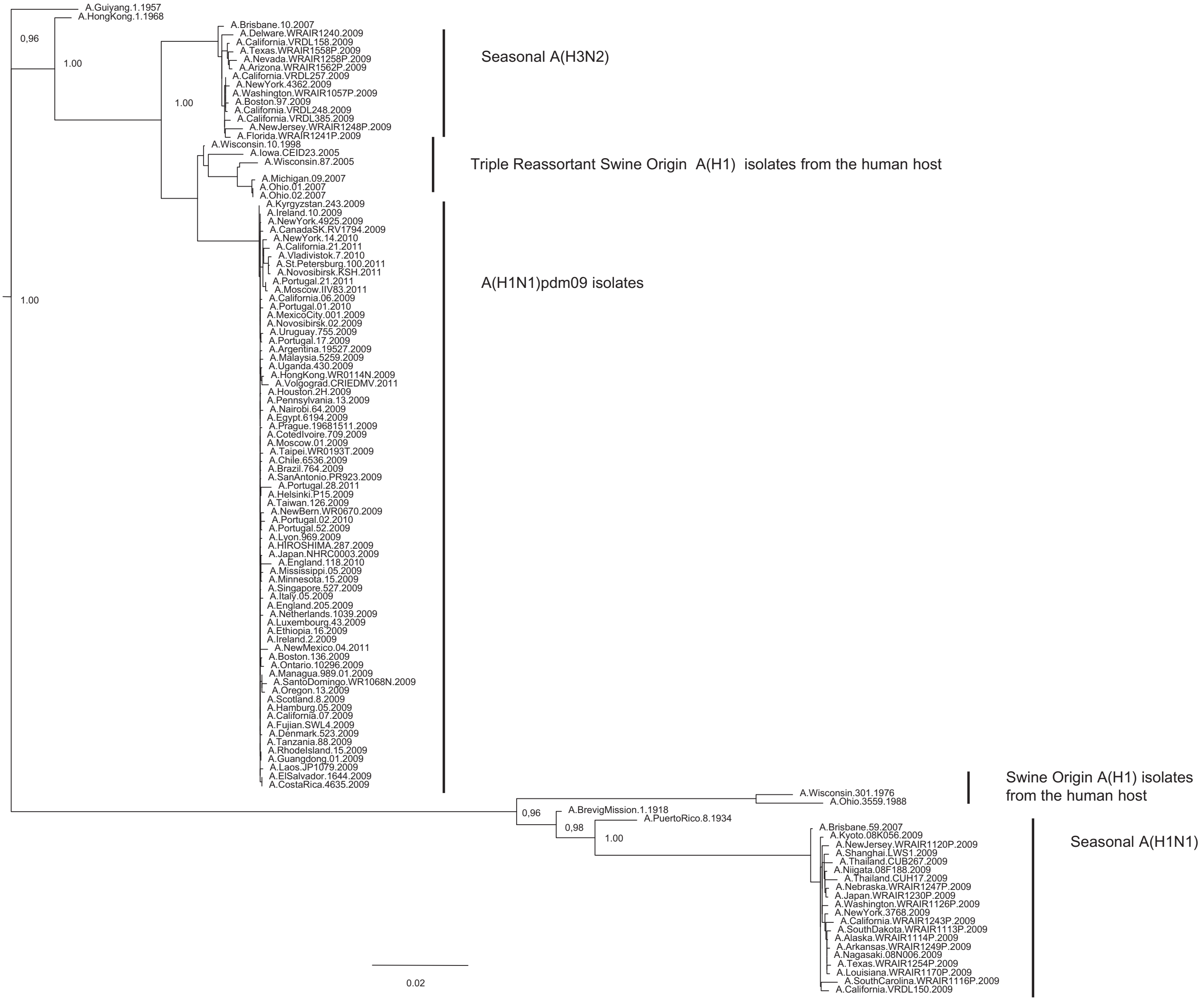

Fig. 1. Phylogenetic tree of the PB1 coding region of swine-origin influenza viruses isolates from the human host. The phylogenetic tree was constructed in *Seaview* using the model GTR+I selected by *JModelTest* software. The accession numbers of the sequences used in this analysis are provided in Table S1.

Amino acid residues of the PB1-F2 coding region

A(H1N1) lineage	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
A/BrevigMission/01/1918	M	G	Q	E	Q	D	T	P	W	I	L	S	T	G	H	I	S	T	Q	K	R	E	D	G	Q	Q	T	P	R	L	E	H	H	N
A/PuertoRico/08/1934	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	K	.	.	.	R	.
A/USSR/90/1977	.	.	.	.	.	G	.	.	.	.	Q	.	.	.	.	.	.	.	.	.	G	.	.	.	.	K	.	.	K	.	.	.	R	.
Consensus seasonal A(H1N1) isolates from 2009	.	.	.	.	.	G	.	.	.	.	Q	.	T/I	.	.	T	.	.	.	.	E	.	.	.	.	K	I	.	K	R	.	.	R	.
A.Wisconsin.301.1976.PB1.CY026145.1.seq	.	.	.	.	.	.	.	.	.	.	.	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A.Ohio.3559.1988.PB1.CY024931.1.seq	.	E	.	.	.	.	.	L	.	.	.	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

A(H1N1) lineage	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68
A/BrevigMission/01/1918	S	T	R	L	M	D	H	C	Q	K	T	M	N	Q	V	V	M	P	K	Q	I	V	Y	W	K	Q	W	L	S	L	R	S	P	T
A/PuertoRico/08/1934	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	I
A/USSR/90/1977	.	.	.	.	.	G	.	Y	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	-	-	-	-	-	-	-	-	-	-
Consensus seasonal A(H1N1) isolates from 2009	L	.	Q	.	.	V	P	Y	R	.	.	.	.	.	.	A	.	.	.	.	.	.	.	*	-	-	-	-	-	-	-	-	-	-
A.Wisconsin.301.1976.PB1.CY026145.1.seq	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A.Ohio.3559.1988.PB1.CY024931.1.seq	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

A(H1N1) lineage	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91
A/BrevigMission/01/1918	P	V	S	L	K	T	R	V	L	K	R	W	R	L	F	S	K	H	E	W	T	S	*
A/PuertoRico/08/1934	L	.	F	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	-	-	-
A/USSR/90/1977	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Consensus seasonal A(H1N1) isolates from 2009	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A.Wisconsin.301.1976.PB1.CY026145.1.seq	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A.Ohio.3559.1988.PB1.CY024931.1.seq	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Fig. 2. Amino acid alignment of PB1-F2 from A(H1N1) lineage viruses. Alignment of the PB1-F2 coding region was performed by *ClustalW* in *Mega5.2 software*. The accession numbers are provided in Table S1.

originated from the avian origin 1918 pandemic strain and transmitted to the swine and human hosts. In the human host, 11L is present in the 1934 PR8 although already substituted for Glutamine, Q, in the reemergent 1977 A(H1N1) subtype (Fig. 2). The genomic position 11 of the PB1-F2 is not recognized, to date, as critical for protein function. However, since Glutamine is retained in 2009 seasonal A(H1N1), and is also present in seasonal A(H3N2) (Table 2), the function of the protein seems to have beneficiated from polarity over the hydrophobic and non-reactive properties of Leucine in this particular position.

The 1918 pandemic virus exclusively presents 33H, 66S, 69P and 90S and, together with PR8 further presents six residues characteristic of an early avian A(H1N1) lineage (positions 42, 43, 50, 65, 84, 86) (Table 2, Fig. 2). Residue 66S is a known signature for virulence (Krumbholz et al., 2011). It would be interesting to further evaluate the association of the remaining residues with a phenotype of enhanced virulence. However, since all are located after the stop codon in 1976 and 1988 SOIV isolates, the aa are not encoded in the swine population and therefore could not have been implicated in the adaptation to the human host.

3.1.2. *TR-SOIV isolates are phylogenetically most closely related to the 1968 pandemic reference strain and A(H3N2)seasonal virus*

As opposed to the 1976 and 1988 isolates, TR-SOIV and A(H1N1)pdm09 are phylogenetically more closely related to seasonal A(H3N2) and 1968 pandemic virus (Fig. 1). Historically, in 1968, seasonal A(H2N2) virus acquired HA and PB1 from an A(H3) avian virus (Zimmer and Burke, 2009; Nelson et al., 2008). The progeny A(H3N2) pandemic virus bearing an avian origin PB1 replaced the seasonal A(H2N2), as characteristically do all emerging virus. A(H3N2) then acquired a seasonal epidemiology and, unprecedented, remained in circulation after the reemergence of A(H1N1) in 1977. The PB1 segment of seasonal A(H3N2) was integrated into the TRIG cassette of TR-SOIV, in the swine population, and passed on to A(H1N1)pdm09. When A(H1N1)pdm09 emerged, the seasonal A(H3N2) was not eliminated. Both viruses co-circulate in the human host and share closely related PB1 proteins that originate from the same avian ancestor.

Within TR-SOIV isolates, the phylogenetic diversity of PB1 reflects the genetic evolution of the viruses in the swine host, from 1998–2007. A/Wisconsin/10/98 appears phylogenetically closest to the precursor of A(H1N1)pdm09, which is in agreement with A(H1N1)pdm09 emergence in the swine population around 1999 (Fig. 1).

We have identified residue 327R, Arginine, as being present in all analyzed SOIV and TR-SOIV (Table 1). This is a swine and avian signature, as opposed to the human signature 327K, Lysine, (Pan et al., 2010). Both are polar amino acids, positively charged and usually involved in salt bridges in active or binding sites, however, their structure is different. This position is located between the conserved motifs I and II of the viral polymerase (Chu et al., 2012). Although putatively interfering with the folding of the protein, we consider that the presence of the swine signature Arginine in SOIV is probably not detrimental to virus fitness when infecting the human host, since the A(H1N1)pdm09 retains it.

All TR-SOIV analyzed code for a functional full length 90aa PB1-F2, as do seasonal A(H3N2) and 1968 pandemic virus (Table 2). It has been proposed that PB1–PB1-F2 interaction does not impact replication kinetics directly (McAuley et al., 2010). Nevertheless, A(H3N2) and TRIG cassette viruses characteristically have a high fitness and the mechanisms by which PB1-F2 may be contributing could involve mitochondrial targeting and the ability to cause cell death, the immunostimulatory properties or the interaction with the expression of N40 (McAuley et al., 2010). TR-SOIV, seasonal A(H3N2) and the reference strain for the 1968 pandemic present, however, amino acid divergences that we consider host related. Eleven residues were identified as putative swine origin markers (34, 60, 71, 74, and 89), six of which are concomitant putative avian markers (62, 82, 85, 29, 37 and 70) (Table 2). It is interesting to recognize that some residues are changed during adaptation to the human host and therefore no longer present in the descendent seasonal viruses (62, 82 85 and 70) and, in some circumstances, in the pandemic reference strain (29 and 37) (Table 2).

The A(H1N1)pdm09 codes for the exact same N-terminal amino acids as the TR-SOIV, but presents a Serine-Stop change resulting in a non functional PB1-F2 of 11aa (Fig. 3, Table 2). The reduced virulence of the 2009 pandemic virus has been attributed, in part, to the lack of a functional full length PB1-F2 (Krumbholz et al., 2011). The mitochondrial localization of PB1-F2 is responsible for changes in morphology and membrane potential, ultimately associated with the initiation of the cellular intrinsic apoptotic pathway. PB1-F2 localizes to mitochondria through a *Mitochondrial Targeting Sequence* (MTS) comprised in 69–82 or 63–75aa (Dundon, 2012). The 11aa length protein does not encode this signal, will not localize to mitochondria and because of that may be less virulent. How-

Table 1
Amino acid residues in PB1 protein of swine-origin influenza viruses isolates from the human host. Alignment of the PB1 coding region was performed by *ClustalW* in *Mega5.2 software.*

	Virus strain																	
	Isolates from human infection with swine-origin influenza A(H1) viruses									Vaccine			Seasonal influenza			Previous pandemic		
	A/Wisconsin/301/1976	A/Ohio/3559/1988	A/Wisconsin/10/1998 *	A/Wisconsin/87/2005 *	A/Iowa/CEID/23/2005 *	A/Michigan/09/2007 *	A/Ohio/01/2007 *	A/Ohio/02/2007 *	consensus A(H1N1)pdm09 isolates from 2009	A/California/07/2009 (A(H1N1)pdm09)	A/Brisbane/10/2007 (seasonal AH3N2)	A/Brisbane/59/2007 (seasonal AH1N1)	consensus seasonal A(H3N2) isolates from 2009	consensus seasonal A(H1N1) isolates from 2009	A/PuertoRico/08/1934	A/BrevigMission/01/1918	A/Guiyang/01/1957	A/HongKong/01/1968
Signature for swine origin																		
327 [b]	R	R	R	R	R	R	R	R	R	R	K	K	K	K	R	R	R	K
Putative marker for adaptation of avian origin A(H1N1) virus to the mammalian host																		
211 [a]	K	K	R	R	R	R	R	R	R	R	R	K	R	K	R	R	R	R
667 [a]	V	V	I	I	I	I	I	I	I	I	I	V	I	V	I	I	I	I
752 [a]	D	D	E	E	E	E	E	E	E	E	E	D	E	D	E	E	E	E
Putative marker for SOIV adaptation to the mammalian / human host																		
433	K	K	R	R	R	R	R	R	K	K	K	K	K	K	K	K	K	K
642	N	N	N	S	N	S	S	S	N	N	N	N	N	N/S	N	N	N	N
216 [a]	S	G	G	S	S	S	S	S	G	G	G	S	G	S	S	S	S	S
586 [a]	K	K	K	K	K	K	K	K	R	R	R	K	R	K	K	K	K	K
Putative marker for adaptation of A(H3N2) lineage viruses to the mammalian host																		
336 [b]	V	V	I	I	I	I	I	I	I	I	I	V	I	I	V	V	V	V
361 [ab]	N	N	R	R	R	R	R	R	R	R	R	S	R	S	S	S	S	S
486 [ab]	R	R	K	K	K	K	K	K	K	K	K	R	K	R	R	R	R	R
584 [a]	R	R	Q	Q	Q	Q	Q	Q	Q	Q	Q	R	Q	R	R	R	R	R
621 [a]	Q	Q	R	R	R	R	R	R	R	R	R	Q	R	Q	Q	Q	Q	Q
741 [a]	A	A	S	S	S	S	S	S	S	S	S	A	S	A	T	A	A	A
Putative marker for adaptation of the TRIG cassette to the new genetic background																		
179 [a]	M	M	I	I	I	I	I	I	I	I	M	V	M	V	M	M	M	M
339 [b]	V	V	M	M	M	M	M	M	M	M	I	I	I	I	I	I	I	I
638	E	E	D	D	D	D	D	D	D	D	E	E	E	E	E	E	E	E
Putative marker for A(H1N1)pdm09 adaptation to the new genetic background																		
12	V	V	V	V	V	V	V	V	I	I	V	V	V	V	V	V	V	V
175	D	D	D	D	D	D	D	D	N	N	D	D	D	D	K	D	D	D
364	L	L	L	V	L	L	L	L	I	I	L	L	L	L	L	L	L	L
435	A	S	T	T	T	T	T	T	I	I	T	T	T	T	T	T	T	T
587	A	A	A	V	A	A	A	A	V	V	A	A	A	A	A	A	A	A
618	K	K	E	E	E	E	E	E	D	D	E	E	E	E	E	E	E	E
728	I	I	I	I	I	I	I	I	V	V	I	I	I	I	I	I	I	I
Putative marker for enhanced compatibility between PB1 and HA																		
298 [ab]	L	L	L	L	L	L	L	L	I	I	L	I	L	I	L	L	L	L
386 [ab]	R	R	R	R	R	R	R	R	K	K	R	K	R	K	R	R	R	R
517 [a]	I	I	I	I	I	I	I	I	V	V	I	V	I	V	A	A	A	A
Putative marker for seasonal A(H1N1) and A(H3N2) viruses adaptation to the human host																		
156 [a]	T	T	T	T	T	T	T	T	T	T	T	I	T	I	T	T	T	T
176/7 [a]	KE	KE	KE	KE	KE	KE	KE	KE	KE	KE	KE	RG	KE	RG	KE	KE	KE	KE
181 [a]	I	I	I	I	I	I	I	I	I	I	I	V	I	V	I	I	I	I
195 [a]	M	M	M	M	M	M	M	M	M	M	M	V	M	V	M	M	M	M
210 [a]	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	H	Q	H	Q	Q	Q	Q
213 [a]	T	N	N	N	N	N	N	N	N	N	N	D	N	D	N	N	N	N
375 [ab]	G	S	S	S	S	S	S	S	S	S	S	N	S	N	S	S	S	S
456/7 [ab]	HE	HE	HE	HE	HE	HE	HE	HE	HE	HE	HE	YA	HE	YA	HE	HE	HE	HE
645 [a]	V	V	V	V	V	V	V	V	V	V	V	I	V	I	M	M	V	V
108 [a]	L	L	L	L	L	L	L	L	L	L	L	I	L	I	I	L	L	L
691 [a]	K	K	K	K	K	K	K	K	K	K	K	R	K	R	R	K	K	K
709 [a]	V	V	V	V	V	V	V	V	V	V	I	V	I	V	V	V	V	V
619 [a]	D	D	D	D	G	D	D	D	D	D	N	D	N	D	D	D	D	D
113 [a]	V	V	V	I	V	I	I	I	V	V	A	V	A	V	V	V	V	V
Other putative markers for PB1 lineage																		
383 [ab]	D	D	E	E	E	E	E	E	E	E	E	D	E	D	D	D	E	E
473 [ab]	L	L	V	V	V	V	V	V	V	V	V	L	V	L	L	L	V	V
576 [a]	I	I	L	L	L	L	L	L	L	L	L	I	I	I	I	I	L	L
212 [a]	L	L	V	V	V	V	V	V	L	L	V	L	V	L	L	L	L	V
54 [a]	K	K	K	K	K	K	K	K	K	K	K	R	K	R	R	R	K	K
654 [a]	T	T	S	S	S	S	S	S	S	S	S	N	S	N	N	N	S	S
430 [ab]	E	E	K	K	K	K	K	K	K	K	K	R	K	R	R	R	K	R

Color legend:
Blue fill: putative marker.
Light grey fill: putative marker for seasonal A(H3N2) lineage.
Dark grey fill: putative marker for seasonal A(H1N1) lineage.
Legend:
[a]Marker for PB1 lineage.
[b]Residue in the proximity of the conserved motifs of the viral polymerase.
*Triple reassortment swine-origin influenza A(H1) virus.

Table 2
Amino acid residues in PB1-F2 protein of swine-origin influenza viruses isolates from the human host. Alignment of the PB1-F2 coding region was performed by *ClustalW* in Mega5.2 software.

	Virus strain																	
	Isolates from human infection with swine-origin influenza A(H1) viruses									Vaccine			Seasonal influenza			Previous pandemic strains		
	A/Wisconsin/301/1976	A/Ohio/3559/1988	A/Wisconsin/10/1998 *	A/Wisconsin/87/2005 *	A/Iowa/CEID/23/2005 *	A/Michigan/09/2007 *	A/Ohio/01/2007 *	A/Ohio/02/2007 *	consensus A(H1N1)pdm09 isolates from 2009	A/California/07/2009 (A(H1N1)pdm09)	A/Brisbane/10/2007 (seasonal AH3N2)	A/Brisbane/59/2007 (seasonal AH1N1)	consensus seasonal A(H3N2) isolates from 2009	consensus seasonal A(H1N1) isolates from 2009	A/PuertoRico/08/1934	A/BrevigMission/01/1918	A/Guiyang/01/1957	A/HongKong/01/1968
Protein lengh (amino acids)	11	11	90	90	90	90	90	90	11	11	90	57	90	57	87	90	90	90
Putative marker for swine origin																		
34			S	S	S	S	S	S			N	N	N	N	N	N	N	N
60			P	P	P	P	P	P			L		L		Q	Q	Q	Q
71			Y	Y	Y	Y	Y	Y			S		S		F	S	S	S
74			I	I	I	I	I	I			T		T		T	T	T	T
89			I	I	I	I	I	I			T		T			T	T	T
Putative marker for swine and avian origin																		
62			L	L	L	L	L	L			P		P		L	L	L	L
82			L	L	L	L	L	L			P		P		L	L	L	L
85			K	K	K	K	K	K			R		R		K	K	K	K
70			G	G	G	G	G	G			V		V		V	V	E	G
29			R	R	R	R	R	R			K	K	K	K	K	R	K	K
37			R	R	R	R	R	R			Q	Q	Q	Q	R	R	Q	Q
Putative signature for viral adaptation to the mammalian host																		
27			T	T	T	I	I	I			I	I	I	I	T	T	T	T
Putative marker for viral adaptation to swine host																		
23			S	N	S	N	N	N			S	D	S	D	D	D	S	S
83			F	S	F	S	S	S			F		F		F	F	F	F
Putative marker for lineage																		
2	G	E	E	E	E	E	E	E	E	E	E	G	E	G	G	G	E	E
10	I	I	T	T	T	T	T	T	T	T	T	I	T	I	I	I	T	T
14			E	E	E	E	E	E			E	G	E	G	G	G	E	E
18			I	I	I	I	I	I			I	T	I	T	T	T	I	I
21			R	K	K	K	K	K			K	E	G	E	R	R	R	K
22			G	G	G	G	G	G			G	E	G	E	E	E	G	G
25			R	R	L	R	R	R			R	Q	R	Q	Q	Q	Q	Q
26			Q	Q	Q	Q	Q	Q			Q	K	Q	K	Q	Q	Q	Q
28			Q	Q	Q	Q	Q	Q			Q	P	Q	P	P	P	R	R
31			G	G	G	G	G	G			G	E	G	E	E	E	E	E
33			P	P	P	P	P	P			P	R	P	R	R	H	P	P
35			S	S	S	S	S	S			S	L	S	L	S	S	L	L
40			D	D	D	D	D	D			D	V	D	V	G	D	D	D
45			I	I	I	I	I	I			I	T	I	T	T	T	T	I
52			H	H	H	H	H	H			H	P	H	P	P	P	H	H
55 [a]			T	T	T	T	T	T			T	I	T	I	I	I	T	T
57 [a]			F	S	F	S	S	S			S	Y	S	Y	Y	Y	S	S
59 [a]			R	R	R	R	R	R			R		R		K	K	K	K
73 [a]			R	R	R	R	R	R			R		R		K	K	K	K
75 [a]			H	H	H	H	H	H			H		H		R	R	R	R
76 [a]			A	A	A	A	A	A			A		A		V	V	V	V
79 [a]			Q	Q	Q	Q	Q	Q			Q		Q		R	R	R	R
81 [a]			K	K	K	K	K	K			K		K		R	R	K	K
87			G	G	G	G	G	G			G		G		E	E	E	G
20			K	K	K	K	K	K			R	K	R	K	K	K	K	K
44			R	K	R	K	K	K			R	K	R	K	K	K	R	R
Putative marker for avian origin A(H1N1) lineage																		
11	L	L	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	L	L	Q	Q
42			Y	Y	Y	Y	Y	Y			Y	Y	Y	Y	C	C	Y	Y
43			L	L	L	L	L	L			L	R	L	R	Q	Q	L	L
50			D	D	D	D	D	D			D	A	D	A	V	V	D	D
65			K	K	K	K	K	K			K		K		R	R	K	K
84			N	N	N	N	N	N			N		N		S	S	N	N
86			Q	Q	Q	Q	Q	Q			Q		Q		H	H	Q	Q
Putative marker for virulence																		
66			N	N	N	N	N	N			N		N		N	S	N	N
69			Q	Q	Q	Q	Q	Q			Q		Q		L	P	Q	Q
90			N	N	N	N	N	N			N		N			S	N	D

Color legend:
Light blue fill: putative marker for avian origin.
Dark blue fill: putative marker for adaptation to mammalian host.
Green fill: putative marker for virulence.
Light grey fill: putative marker for seasonal A(H3N2) lineage.
Dark grey fill: putative marker for seasonal A(H1N1) lineage.
Legend:
[a] Residue located within the predicted helical region of the PB1-F2.

Amino acid residues of the PB1-F2 coding region

A(H3N2) lineage	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
A.Guiyang.1.1957.PB1.153896.seq	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	.	.
A.Hong.Kong.1.68.PB1.8435.seq	M	E	Q	E	Q	D	T	P	W	T	Q	S	T	E	H	I	N	I	Q	K	K	G	S	G	Q	Q	T	R	K	L	E	R	P	N
consensus Seasonal A(H3N2) isolates from 2009	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	R	G	.	.	.	.	.	I	.	.	.	.	.	.	.
consensus A(H1N1)pdm09 isolates from 2009	.	.	.	.	.	.	.	.	.	.	.	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A.Wisconsin.87.2005.PB1.291901.seq	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	N	.	R	.	.	Q	R	.	G	H	.	S
A.Iowa.CEID23.2005.PB1.DQ889683.1.seq	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	L	.	.	.	.	.	.	.	.	.
A.Wisconsin.10.98.PB1.AF342823.seq	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	.	.
A.Michigan.09.2007.PB1.388380.seq	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	I	.	.	.	.	.	.	.
A.Ohio.01.2007.PB1.291859.seq	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	I	.	.	.	.	.	.	.
A.Ohio.02.2007.PB1.291907.seq	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	I	.	.	.	.	.	.	.

A(H3N2) lineage	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68
A.Guiyang.1.1957.PB1.153896.seq	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
A.Hong.Kong.1.68.PB1.8435.seq	L	T	Q	L	M	D	H	Y	L	R	I	M	S	Q	V	D	M	H	K	Q	T	V	S	W	K	Q	W	L	S	L	K	N	P	T
consensus Seasonal A(H3N2) isolates from 2009	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	L	.	P	.	.	.	.	.	.
consensus A(H1N1)pdm09 isolates from 2009	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A.Wisconsin.87.2005.PB1.291901.seq	S	.	R	.	.	.	.	.	.	K	.	.	N	.	.	.	.	.	.	.	.	.	.	.	R	P	.	.	.	.	.	.	.	.
A.Iowa.CEID23.2005.PB1.DQ889683.1.seq	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	F	.	.	.	.	.	.	.	.	.	.	.
A.Wisconsin.10.98.PB1.AF342823.seq	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	F	.	.	.	.	.	.	.	.	.	.	.
A.Michigan.09.2007.PB1.388380.seq	S	I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A.Ohio.01.2007.PB1.291859.seq	S	I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A.Ohio.02.2007.PB1.291907.seq	S	I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

A(H3N2) lineage	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91
A.Guiyang.1.1957.PB1.153896.seq	.	E	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	E	.	.	.	*
A.Hong.Kong.1.68.PB1.8435.seq	Q	G	S	L	K	T	R	V	L	K	R	W	K	L	F	N	K	Q	G	W	T	D	*
consensus Seasonal A(H3N2) isolates from 2009	.	V	.	.	.	.	.	.	.	.	.	.	.	P	.	.	R	.	.	.	.	.	*
consensus A(H1N1)pdm09 isolates from 2009	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A.Wisconsin.87.2005.PB1.291901.seq	.	.	Y	.	R	I	H	A	.	.	Q	.	.	.	S	.	.	.	.	.	I	N	*
A.Iowa.CEID23.2005.PB1.DQ889683.1.seq	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*
A.Wisconsin.10.98.PB1.AF342823.seq	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*
A.Michigan.09.2007.PB1.388380.seq	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	*
A.Ohio.01.2007.PB1.291859.seq	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	*
A.Ohio.02.2007.PB1.291907.seq	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	*

Fig. 3. Amino acid alignment of PB1-F2 from A(H3N2) lineage viruses. Alignment of the PB1-F2 coding region was performed by *ClustalW* in *Mega5.2 software.* The accession numbers are provided in Table S1.

ever, an A(H1N1)pdm09 virus genetic engineered to express PB1-F2, has been reported to maintain low level virulence (Hai et al., 2010). It has even been suggested that PB1-F2 does not have an evolutionary importance to the virus and that it does not have major impact in virus fitness (Trifonov et al., 2009). We can conclude that the impact of PB1-F2 in viral pathogenesis, virulence and fitness has to be strain dependent to some extent and that further research is crucial at this point.

3.1.3. Marker residues distinguish PB1 genomic segment from A(H1N1) and A(H3N2) lineages

We have identified marker residues that distinguish PB1 genomic segment from A(H1N1) and A(H3N2) lineages on the basis of protein structure and a phosphorylation site, and that we propose may contribute to the known differences in viral fitness. In PB1 coding region, 9 of the 35 markers for lineage (298, 361, 375, 383, 386, 430, 456/7 and 473) are located between the conserved motifs of the viral Polymerase (motif I 303–306, motif II 403–412, motif III 438–450, motif IV 474–484) (Reuther et al., 2011; Chu et al., 2012; Biswas and Nayak, 1994) (Table 1). Both the composition and length of these inter-motifs sequences are critical to protein folding and could seriously alter the sites for binding and recognition. Regarding PB1-F2, a Serine-Leucine substitution occurred in position 35 of 2009 seasonal A(H1N1). Based on our genetic analysis, and since the reemergent 1977 A(H1N1) presented 35S, we propose that the substitution has occurred between 1977 and 2009, during viral evolution in the human host. Residue 35S is a phosphorylation site associated the regulation of PB1-F2 interaction with PB1, which directly relates to virus titers (Krumbholz et al., 2011). The lack of this site has been reported to have a detrimental effect in fitness and we consider that it could have contributed to the phenotype of more reduced fitness in A(H1N1) seasonal viruses, when compared to A(H3N2). Particularly since within A(H3N2) seasonal lineage, the exact opposite event appears to have occurred (Table 2, Fig. 3). The 1968 A(H3N2) avian origin pandemic virus lacking this phosphorylation site seems to have acquired it during its evolution in the human host and introduced it into the TRIG cassette, since it is now present in the 2009 seasonal isolates and TR-SOIV. Still within the markers for lineage in PB1-F2, 8 of the 32 (positions 55, 57, 59, 73, 75, 76, 79 and 81) are located within the predicted helical region (aa 55–85) which includes the MTS. The formation of an helical structure in this C-Terminus is essential because the positively charged amino acids have to be presented to the negatively charged mitochondrial membrane and other cellular compartments. From this interaction results the formation of pores that initiate the apoptotic process. Any changes in the folding of the protein or in the amino acids that are presented and interact with the membrane can affect the virus ability to control cellular apoptosis.

3.2. Genetic markers for viral adaptation

3.2.1. Residues found in PB1 and PB1-F2 proteins putatively associate with viral adaptation to the mammalian host, given the molecular epidemiology and the specific aa substitutions

In SOIV isolates from 1976 and 1988, Lysine, K, Valine, V, and Aspartate, D, replaced Arginine, R, Isoleucine, I, and Glutamate, E, in positions 211, 667 and 752 of PB1 (Table 1). Because the aa have similar properties, the consequences are not evident. However, seasonal A(H1N1) also present these substitutions and, although the swine and seasonal lineages appear to share a common avian ancestor, their genetic evolution paths occurred independently in different hosts. Also, 752 is located in the binding domain to PB2 (678–757aa) (Reuther et al., 2011) and 667 is part of a broader ranged binding domain (600–757aa) (Ohtsu et al., 2002), established prior to the more contemporary fine mapping. We then propose that, within A(H1N1) subtype, avian origin viruses must have beneficiated from these substitutions during their adaptation to mammalian hosts.

In TR-SOIV, residues 433R, Arginine, and 642S, Serine, of the PB1 protein appear as a putative genetic marker and a dominant aa, respectively (Table 1). Arginine is frequently associated with active or binding sites because it is able to form multiple hydrogen bonds. The substitution of Arginine for Lysine is possible since both are positively charged. It is, however, very uncommon because Lysine is more limited in the number of hydrogen bonds it can establish. Although 433 has not been recognized as part of a binding domain, in the particular case of the TRIG cassette, a more flexible or less strong bonding capacity in this position appears to have favored A(H1N1)pdm09 fitness in the human host and putatively reflects an adaptation. Opposing, in position 642, a more reactive aa appears to have benefitted the fitness of A(H1N1)pdm09. Asparagine, N, is a polar aa, usually located at proteins surface in contact with the aqueous environment and predominantly involved in binding sites. This position is part of the broader ranged binding domain to PB2 (Ohtsu et al., 2002), referred above and although it may not be essential for binding, aa changes could alter the affinity. Given that NA and M were newly acquired by A(H1N1)pdm09, these could be compensatory mutations for adaptation to the new genetic background that promotes the structural interaction of PB1 and M proteins.

A different circumstance presents in positions 216 and 586. A(H1N1)pdm09 and seasonal A(H3N2) present Glycine, G, and Arginine, R, as opposed to Serine, S, and Lysine, K, respectively (Table 1). Molecular epidemiology suggests a phenotypic purpose to the substitutions. A Lysine–Arginine change is apparently neutral in terms of aa structure and function. Glycine, however, is not as reactive as Serine and is very particular because its structure allows the most flexible conformations. Under this circumstance, any substitution of Glycine, even for another small aa like Serine, most probably alters the structure of the protein.

In PB1-F2 coding region, residue 27T, Threonine, has been reported as phosphorylation site (Krumbholz et al., 2011). The phosphorylation status of PB1-F2 contributes to regulate the functionality of the protein in its direct interaction with PB1. TR-SIOV and seasonal A(H1N1) and A(H3N2), present 27I, Isoleucine (Table 2, Fig. 3). Based on this genetics analysis, we propose Threonine to be a putative avian origin genetic marker that was introduced into the seasonal lineages A(H1N1) and A(H3N2), in 1918 and 1968 respectively. Since the reemergent 1977 A(H1N1) retained the avian marker, we propose the substitution for Isoleucine to have occurred between 1977 and 2009. According to the known historical molecular epidemiology, within A(H3N2) lineage, the avian marker was probably introduced in the swine host in the late nineties, when the seasonal A(H3N2) PB1 genomic segment was acquired by the TRIG cassette, and, therefore, it was present in TR-SOIV 1998 and 2005 isolates. The T–I mutation must have, then, occurred independently in both hosts, strongly suggesting an adaptation of an avian origin virus to mammalian host cellular environments. We propose that this residue has probably been beneficial for avian viruses to infect mammalian hosts, but got lost in the subsequent circulation in swine and human populations. Another dramatic substitution occurs in position 83, where Phenylalanine, F, an hydrophobic aromatic aa usually involved with non-protein ligands is substituted by Serine, a small, polar, reactive aa (Table 2). In position 23, an apparently more neutral substitution occurred, in which Serine, S, was replaced by Asparagine, N (Table 2). Both are polar and reactive, usually exposed in the surface of the protein. Serine is, however, smaller in size and it could have altered the structure of the protein. According to our genetic analysis and to the known molecular epidemiology of the strains, both substitutions probably occurred in the swine population around 2005 and were fixed for being beneficial to the virus in the given environment.

3.2.2. Genomic positions putatively associate with the adaptation of PB1 from A(H3N2) lineage viruses to the mammalian host

Position 584 has been previously reported as undergoing changes in selective pressure during host shifts from avian to human. Particularly, an Arginine, R, to Glutamine, Q, change is described as an adaptive mutation (Tamuri et al., 2009). Here we found the avian marker Arginine present in pandemic strains of 1918, 1957 and 1968 (Table 1). Based on our genetic analysis and on the known molecular epidemiology data, we propose that seasonal A(H3N2) viruses have evolved to present R–Q change from the 1968 pandemic virus, subsequently incorporated in TR-SOIV and transferred to A(H1N1)pdm09. Since both 1976 and 1988 SOIV and seasonal A(H1N1) isolates retained the avian marker, we propose that this aa change was probably not essential for adaptation to the human or mammalian hosts, but instead was a genetic sweep that enhanced fitness. Both are polar aa, although Arginine is positively charged. It can be substituted by other positively charged aa but it has also been reported to tolerate changes to non charged ones. In fact, a more neutral side chain appears to have favored viral protein interactions of A(H3N2) viruses in the human cellular environment.

We propose that similar situations occur in 336, 361, 486, 621 and 741 (Table 1). None of the substitutions were neutral. Particularly, 336 and 361 are located between conserved motifs I and II of the viral Polymerase and 486 distances only 2 amino acids from the end of motif IV. All may have had an impact in protein folding. In position 336, the 2009 seasonal A(H1N1) isolates present the putative marker for adaptation Isoleucine, I. Isoleucine is similar to Valine, although more restrictive in the conformations it can adopt and which appears to be favored in the mammalian cell environment. In 486, Arginine is replaced by Lysine. Both are polar positively charged but this substitution is putatively prejudicial when interfering with structural sites, as discussed above. In position 361, a polar aa was replaced by an equally polar but positively charged one, and again in positions 621 and 741, less reactive or neutral amino acids were substituted for more reactive positively charged ones. In these cases, more reactive aa seem to benefit adaptation of avian TRIG cassette viruses to mammalian hosts, possibly by increasing the interaction of PB1 with host cellular proteins.

3.2.3. Residues putatively relate to the adaptation of PB1 to new genomic backgrounds on the basis of their molecular epidemiology

Residues 179I, 339M and 638D are exclusive in PB1 of TRIG cassette viruses (Table 1). Position 638 is located in the broader range binding domain to PB2, referred above (Ohtsu et al., 2002). As opposed to Aspartate, D, in TRIG cassette viruses, both 1976 and 1988 SOIV, seasonal A(H1N1) and A(H3N2) and previous pandemic strains present Glutamate, E. These are polar amino acids, frequently exposed in the surface of the proteins and associated with active or binding sites. Aspartate confers a more rigid structure to the site, which is usually more advantageous for binding. In the TRIG cassette, PB1 and PB2 originate from seasonal A(H3N2) and avian lineages, respectively, and their interaction appears to have been favored by a more rigid structure.

In 179 and 339, Methionine, M, Isoleucine, I, and Valine, V, are present in specific patterns (Table 1). All are hydrophobic amino acids, not very reactive and not usually involved in protein function. Their roles are predominately associated with recognition or binding sites. Although Methionine is even more limited in the roles it can play in protein function because of its atomic composition, Valine and Isoleucine are more restrictive in the conformations they can assume because of their structure. It is intriguing why these positions, apparently not corresponding to particular functions and not recognized as active or binding sites, present such defined patterns of aa. It is clear, however, that Methionine and Isoleucine must have been beneficial, to some extent, to PB1 activity, or were fixed as compensatory mutations for genetic changes in other viral proteins.

In the A(H1N1)pdm09, 12I, 175N, 364I, 435I, 587V, 618D and 728V are exclusive residues in PB1 coding region (Table 1). Position 12 marks the terminus of PB1 binding domain to PA (Perez and Donis, 2001). In a previous study, this position has been assign a particular role in PB1-PA binding, and a Valine, V, to Aspartate, D, change was reported to decrease it by 40% (Perez and Donis, 2001). Any change in the sequence of the domain could have a serious impact in the polymerase heterodimer formation. The V–D change previously reported replaces a hydrophobic amino acid for a polar negatively charged one. In our analysis, however, Valine is replaced by Isoleucine, I, which is similar in structure and function. Both are hydrophobic and usually located in protein cores, non-reactive and mostly involved in recognition sites. Because of their atomic composition, both are very restrictive regarding the conformations they can adopt. We propose that either Isoleucine appeared as a spontaneous mutation in the quasi-species of A(H1N1)pdm09 and prevailed since it was not detrimental to PB1-PA binding, or the substitution of has in fact a specific positive phenotypic translation, despite being so similar.

Two other unique residues in A(H1N1)pdm09, 618D, Aspartate, and 728V, Valine, are located , respectively, in the broader ranged and in the fine mapped PB1 binding regions to PB2 (Reuther et al., 2011; Ohtsu et al., 2002). In 618, Aspartate replaces Glutamate that is present in the TR-SOIV isolates. The exact same substitution occurred in position 638 of TRIG cassette viruses, described above. A clear preference for a more rigid conformation seems to be occurring in the binding of PB1–PB2. In the A(H1N1)pdm09, the presence of 618D has probably further increased stability and fitness advantage over TR-SOIV when infecting the human host. Residue 728V distinguishes A(H1N1)pdm09 from all other SOIV analyzed, seasonal and previous pandemic strains, who present 728I, Isoleucine. Again, given the similarities between Valine and Isoleucine, it is unclear why the viruses present them in such a strict pattern. The remaining set of four exclusive signature residues in A(H1N1)pdm09, 175N, 364I, 435I and 587V, is distributed along PB1 coding region. Their positions are close but not part of any mapped structural or functional sites and again aa changes do not seem to be consequential to a great extent. In 175N, Asparagine, and 435I, Isoleucine, a more neutral and less reactive amino acid appears to be favored in A(H1N1)pdm09, since all other SOIV, previous pandemic and seasonal strains present the polar more reactive amino acids Aspartate, D, and Threonine, T, respectively. In 364 and 587, Isoleucine, I, replaced Leucine, L, and Valine, V, replaced Alanine, A. All are hydrophobic aa, although Valine is slightly more than Alanine, and all are very similar in structure and function. We consider that all could reflect compensatory mutations in the adaptation of PB1 to the new genetic background or spontaneous mutations not detrimental to virus fitness.

3.2.4. Residues putatively relate to an enhanced compatibility between PB1 and HA, on the basis of their molecular epidemiology

Residues in the A(H1N1)pdm09 were analyzed as putative markers for enhanced compatibility between PB1 and HA, in the human host. In position 386, an Arginine, R, is replaced by Lysine, K, in A(H1N1)pdm09 and 2009 seasonal A(H1N1) isolates (Table 1). The exact same substitution was discussed above as possibly neutral but also potentially prejudicial when interfering with structural sites. This particular position, 386, lays within the conserved polymerase motifs I and II where a less rigid aa composition appears to be favoring the folding of PB1 protein. The origin of PB1 in A(H1N1)pdm09 and seasonal A(H1N1) is distinct and, consequently, this mutation is not a product of a common path of evolution. These strains additionally present the same exclusive substitutions in positions 298 and 517. Although in 517 an apparently neutral substitution occurred, Isoleucine, I, to Valine, V, in position 298, which distances only 5 aa from the beginning of motif I, the substitution of Leucine, L, to Isoleucine, I, results in a more restricted conformation. The reassortment events in the history of PB1 genomic segment suggest that a genomic compatibility between HA and PB1 proteins enhances viral fitness, although the extent of the interaction remains unclear. We consider that a pattern where similar substitutions occur in PB1 proteins of different origins, near the conserved motifs where protein folding is critical and occurring in viruses that share the A(H1) subtype, suggests an adaptation of PB1 towards an enhanced compatibility with H1 in the human host.

4. Conclusions

We have found that PB1 does retain traces of viral interspecies transmission and adaptation. In the SOIV A(H1) that have infected the human host, the evolutionary history of PB1 and PB1-F2 proteins is traceable in term of lineage and host origin. PB1 from A(H1N1) and A(H3N2) lineages appear to be distinguished by aa changes in intermotifs sequences of PB1 protein, that can have major impact in protein folding and heterotrimer formation. Lineage distinction also appears to be related to the expression of the *Mitochondrial Targeting Sequence* in PB1-F2 and to genetic changes in the predicted helical region which can affect the virus ability to control apoptosis.

Moreover, specific genomic markers appear to be putatively related to viral adaptation. Substitutions between aa with different properties occur, in genomic positions that are critical to protein function or structure, and with molecular epidemiology data further supporting the assumption that they occur as part of the process of viral adaptation. We propose residues 27I in PB1-F2 and 336I, 361R, 486K and 584Q in PB1 as putative genetic markers for viral adaptation to the mammalian host. Also, residues 638D and 618D are proposed as putative genetic markers for viral adaptation to the new genomic background, in TRIG cassette viruses and in A(H1N1)pdm09, respectively. We additionally highlight residues 298I and 386 as putatively associated with the enhancement of PB1–HA compatibility. There is no apparent trend in the evolutionary process regarding aa reactivity or structure. The fitness of the viruses appears to have been favored by more neutral and less reactive aa in some positions, while in others more reactive ones were fixed. As regards to protein structure, more flexible conformations were also putatively associated with higher protein stability in general but, in some circumstances, more rigid or restrictive conformations appear to have favored the folding of the proteins and the binding to other polymerase subunits.

The evolutionary rates determined for internal proteins are typically lower than those for HA and NA, due to differences in the selective pressure exerted by the host immune system. The present analysis does, however, highlight that PB1 genomic segment of influenza A viruses evolves to divergent lineages and adapts to host and genetic backgrounds specificities. Tracing the genetic evolution is the basis to further understand the mechanisms by which PB1 affects viral fitness and, specifically, the identification of new residues and regions with putative roles in adaptation can drive target research on antivirals. This analysis now permits putative adaptive related polymorphisms to be experimentally evaluated for phenotypic impact.

Statement of author's contribution

The phylogenetic and mutation trend analysis of the PB1 sequences, the research on evolutionary history of interspecies