

Molecular analysis of the virulence attenuation process in Junín virus vaccine genealogy

Sandra Elizabeth Goñi · Javier Alonso Iserte ·
Betina Inés Stephan · Cristina Silvia Borio ·
Pablo Daniel Ghiringhelli · Mario Enrique Lozano

Received: 14 October 2009 / Accepted: 18 January 2010 / Published online: 11 February 2010
© Springer Science+Business Media, LLC 2010

Abstract The Junín virus strain Candid#1 was developed as a live attenuated vaccine for Argentine hemorrhagic fever. In this article, we report sequence information of the L and S RNAs of Junín virus Candid#1 and XJ#44 strains, and show the comparisons with the XJ13 wild-type strain and with other Junín virus strains, like Romero, IV4454 and MC2 strains, and other closely and distantly related arenaviruses. Comparisons of the nucleotide and amino acid sequences of all genes of three strains from the same vaccine genealogy, revealed different point mutations that could be associated with the attenuated phenotype. A 91% of the mutations found are consistent with a hypothesis of progressive attenuation of virulence from XJ13 to XJ#44 and to Candid#1; 39% of mutations were observed in XJ#44 and conserved in Candid#1, while another 52% of the mutations appeared only in Candid#1 strain. The remaining 9% corresponded to reverse mutations in the L gene. In summary, the present work shows a set of mutations that could be related to the virulence attenuation phenomenon. This information will serve as a starting point to study this biological phenomenon, provided that a reverse genetics system for Junín virus is developed to allow the generation of infectious virions with specific mutations.

Keywords Arenavirus · Virulence attenuation · Genetic analysis · Junin virus · AHF vaccine

Introduction

Arenaviruses are important emerging pathogens and include a number of hemorrhagic fever viruses classified as NIAID category A priority pathogens and CDC potential biothreat agents. Junín virus, a member of the family Arenaviridae [1, 2] (virus code: 00.003.0.01.010.), is the etiological agent of Argentine hemorrhagic fever (AHF). The clinical symptoms of AHF include hematological, neurological, cardiovascular, renal, and immunological alterations. The mortality rate for AHF can be as high as 30%, but early treatment with immune plasma reduces fatal cases to less than 1%. The human population at risk is composed mainly of field workers, who are believed to become infected through cuts or skin abrasions or via air-borne dust contaminated with urine, saliva, or blood from infected rodents [3].

Arenaviruses are enveloped viruses with genomes consisting of two single-stranded RNA species, termed L (ca. 7 kb) and S (ca. 3.5 kb). The open reading frames of both RNA species are arranged in an ambisense manner and are separated by a non-coding intergenic region that folds into a stable secondary structure [1]. The L RNA codes for two proteins, a large polypeptide (L), presumed to be the RNA polymerase, and a small zinc finger-like protein (Z) that could be a counterpart of the matrix protein in other riboviruses [4, 5]. The complete nucleotide sequence of the S RNA from several arenaviruses has been determined, and several partial sequences are also available [6]. The S RNA species codes for the nucleocapsid protein, N, and the precursor of the envelope glycoproteins, GPC. Proteolytic cleavage of GPC in infected cells produces a stable signal peptide and the G1 and G2 polypeptides [7]. N and L proteins are translated from anti-genome-sense mRNAs, complementary to the 3' portion of the viral S or L RNA,

S. E. Goñi · J. A. Iserte · B. I. Stephan ·
C. S. Borio · P. D. Ghiringhelli · M. E. Lozano (✉)
Departamento de Ciencia y Tecnología, LIGBCM, Universidad
Nacional de Quilmes, Argentina, Roque Sáenz Peña 352,
CP 1876BXD Bernal, Buenos Aires, Argentina
e-mail: mlozano@unq.edu.ar

respectively. GPC and Z proteins are translated from genome-sense mRNAs corresponding to the 5' region of the viral S or L RNA, respectively.

A collaborative effort conducted by the US and Argentine Governments led to the production of a live attenuated Junín virus vaccine, named Candid#1 [8]. After rigorous biological testing in rhesus monkeys, the vaccine was used in human volunteers, followed by an extensive clinical trial in the AHF endemic area [9]. This vaccine was derived from the 44th mouse brain passage (XJ#44) of the prototype strain of Junín virus [10]. Molecular characterization of the vaccine strain Candid#1 (avirulent strain), and of its more virulent ancestors, XJ13 (high virulence) and XJ#44 (intermediate virulence), allows a systematic approach with the aim of studying the basis for Junín virus virulence. Here, we describe sequence information of the L and S RNAs of Junín virus Candid#1 and XJ#44 strains, and show the comparisons with the XJ13 wild-type strain and with other Junín virus strains, like Romero, IV4454 and MC2 strain, and other closely and distantly related arenaviruses.

Methods

Viral strains and cell culture

The passage history of Junín Candid#1 and XJ#44 has been described elsewhere [8]. The original virus isolate that gave rise to the Candid#1 strain was the XJ strain, isolated in Junín City (Buenos Aires, Argentina) from a human AHF patient [11]. Records of the passage history of the XJ strain come from the Yale Arbovirus Research Unit, Connecticut, USA (J. Casals Laboratory) and USAMRIID, Frederick, Maryland, USA (J. G. Barrera Oro Laboratory). A 'working stock' of Junín Candid#1 virus was produced by infection of certified fetal rhesus lung diploid (FRhL-2) cell monolayers with the master seed. The attenuated Junín virus XJ#44 was provided by J. G. Barrera Oro (USAMRIID) and was amplified in our laboratory in BHK21 cells. The more virulent XJ13 strain (prototype) was provided by C. J. Peters (CDC, Atlanta, Georgia, USA) as a lysate of infected BHK21 cells. Virions were recovered and purified from the supernatant media; viral and total RNA from infected cells was obtained according to procedures described previously [8].

cDNA synthesis and PCR amplification

During molecular cloning of Junín virus cDNA (Candid#1 and XJ#44), special attention was devoted toward avoiding spurious genetic variations that possibly could obscure changes relevant to the attenuation of virulence. Selected

regions of the S and L RNAs were amplified by RT-PCR. The primers used in the amplification are listed in Table 1A, B. The cDNA synthesis was carried out as reported previously [8] and the target sequences were amplified using high-fidelity thermostable DNA polymerases. Amplified cDNAs were analyzed on agarose gels, purified using sodium iodide and glass powder elution (QIAprep^R Miniprep Kit, Qiagen) and cloned into linearized pCR Blunt II-TOPO^R (Invitrogen) plasmid DNA. For the S RNA, five overlapping fragments comprising the entire segment were amplified and purified for direct sequencing.

Sequence determination and alignment tool

For each of the analyzed regions at least four independent cDNA clones were sequenced by the chain termination method. Additionally, direct sequencing of PCR products was used to confirm the sequences of cDNA clones. Nucleotide sequences of the following arenaviruses S and L RNAs were obtained from the GenBank database (accession numbers are indicated between brackets): Junín XJ13 (L: NC_005080, S: NC_005081); Junín XJ#44 (L: DQ489718, S: GQ121040); Junín Candid#1 (L: AY918707, S: FJ969442); Junín Romero (L: AY619640, S: AY619641); Junín MC2 (L: AY216507, S: D10072); Junín IV4454 (S: DQ272266, Z gene: DQ538136); Machupo Carvalho (AAT40450.1, AAT40449.1); Machupo Chicava (AAT45080.1, AAT45079.1); Machupo Mallele (AAT40454.1, AAT40453.1); Tacaribe (NP_694848.1, NP_694847.1). Sequence alignments were done using the CLUSTAL X software [12].

RNA secondary structure predictions

RNA secondary structure predictions were done at the Rensselaer bioinformatics web server using the Mfold software (<http://mfold.bioinfo.rpi.edu/>, [13]). Mfold predicts nucleic acid structures by minimizing free energy of base pairing. The size of the molecule which structure is being predicted must be 800 bases long for on line works and 6,000 bases long for in batch works. The length of genomic S RNAs matches well with the maximum allowed prediction but the genomic L RNAs does not match. However, our interests are focused on the predictions of the RNA's panhandle structure. Therefore, the prediction was made on chimeric molecules containing one hundred bases of the 5'- and 3'-termini of the genomic RNA joined by a polyA sequence. The size of the chimeric molecule was up to 800 bp. The prediction was repeated varying the length of the genomic termini and polyA spacer. In all cases the predicted structure of the panhandle was identical.

Table 1 Primers used in PCR amplification

Study region	Name viral primers	Sequence (5′–3′)	Name viral complementary primers	Sequence (5′–3′)
A. L RNA amplification				
ARena	ARLV-250V	CGCACAGTGGATCCTAGGC	ARLV-250C	CAKYTTAGACAMAGRTARTGRT
	ARLV-2200V	TCACCTGYAGTGAYCAYTA	ARLV-2200C	ATCAACAARWSWGCNTTYCA
	ARLV-2800V	CAAACACCTATRAAWCCWGATG	JLV-2250C3	CAATCAACAATAATCCGCATT
	ARLV-3700V	ATTGTGTCWATTGTWTC	JLV-2300C2	GGCGTGAAGAAGATCTTGGC
	JLV-4200V	ACCNACYTYGTYTTRTA	ARLV-3100C	CCTAYACHTCWAGYGAYGA
	JLV-4900V	TGCATTGACACCTCTGC	ARLV-3700C	AAWSWTA YAARGARCARGT
	ARLV-5000V	GGCTCAATAAAACTTCTC	JLV-3700C1	ACCTGAAAGTTCGGACTAAAGT
	JLV-5000V4	TCWAWGAAAYTTCTCAAAACAT	JLV-4600C	TTATGATAAGGAGAAAAC
	ARLV-6100V	GTCCTCTTGTCCAGTGTGAG	ARLV-5000C2	GAYCAAAATCAARTGYTTYGA
	ARLV-6600V	ARMGKRYTYCKYCKNGTRTT	JLV-5600C	TTTCCCTTGCATTGGTC
		ATNAGRCGDHNRATYTC	JLV-6500C	TAGAAAACATCGAATGGCTTG
B. S RNA amplification				
FS1		GTAAAACGACGGCCAGAGGGGATCCTAGG	RS1	CAGGAAACAGCTATGACATAGCCTCCAGGGG
FS2		GTAAAACGACGGCCAGCTCCAAGACTGG	RS2	CAGGAAACAGCTATGACATCTGCAACCACCC
FS3		GTAAAACGACGGCCAGTCCGTAATGACTGG	RS3	CAGGAAACAGCTATGACATGAAAGGACCAAGC
FS4		GTAAAACGACGGCCAGA ACTGCCTGGCTTG	RS4	CAGGAAACAGCTATGACTGCATGCATGACTG
FS5		GTAAAACGACGGCCAGCCTGTTTCTCCATC	RS5	CAGGAAACAGCTATGACCGCACAGTGGATCC
C. L RNA RACE				
PoliT		CCGGATCCTCTAGAGCGGCCGC(T) ₁₇ V	150C-RACE	ATTCTGGCGGTGGCTGTGTG
6970V-RACE		ACAACAGAGAGCTGGGGTGATG	250C2-RACE	CATCTNAGRCANAGRTARTG
6850V-RACE		GAAGGGTGGACATAGTATGCC	540C-RACE	TGATATCAGATCCCCAAGGGCA
3212V-RACE		TGTCAGCAATTAGCTTGGCG	161C-RACE	ATGGCAATGAGACTGACTGC

Names of the generic primers, designed to amplify members of *Arenaviridae*, begin with “AR”; Junin virus specific primer names begin with a “J”. Genomic sense primers have a V letter in their name, and antigenomic sense primers have a C letter in their name. A The nucleotide generic positions are indicated using the universal code, B the *underlined* nucleotides correspond to universal primers sequence (M13 forward [F] and M13 reverse [R]), C and D have names with “RACE” added at the end

Pattern search

The protein sequences were scanned in order to find putative domains and motifs. The tools used were ScanProsite (ExPASy Proteomics Server, ScanProsite, www.expasy.ch, [14]) for identifying common motifs, and NETNGLYC 1.0 (<http://www.cbs.dtu.dk/services/NetNGlyc/>; [15]) to search for N-glycosylation sites.

Phylogenetic analyses

The ORFs nucleotide sequences from the N, GPC, Z, and L genes, obtained from those arenaviruses whose complete genome is at the GenBank, were aligned independently using Clustal X. Each alignment was bootstrapped (100 replicas) and a neighbor joining and parsimony analyses were performed using the Phylip package.

Ends sequence determination

The analysis of the viral genomic ends was performed by a RACE method [16]. To determine the 5'-end sequences, cDNAs were synthesized using a specific primer and elongated adding dC nucleotides with terminal transferase (Promega Life Science). To determine the 3'-end sequences, viral genomic or infected-cells-derived RNAs were elongated by adding a poly-A tail using a polyA polymerase (United States Biotechnology). The primers used for the RACE analysis are listed in Table 1C, D (for the L and S RNAs, respectively).

Results

In order to achieve the amplification of the different regions of the Junín virus genome, we designed a large set of primers. Some of these primers were designed comprising degenerate bases because there was not enough information available about Junín and closely related arenavirus sequences at the beginning of this study (Table 1A).

The amplification products obtained for Junín virus strains Candid#1 and XJ#44 were sequenced directly or after cloning. In both instances, several sequencing reactions were performed.

Sequence data was analyzed using a series of bioinformatics tools. In a previous work, a small set of changes between S RNA from the Candid#1 and XJ13 Junín virus strains were reported [8]. However, when we compared the sequence of XJ#44 and the re-sequenced Candid#1 S RNAs to the recently sequenced XJ13 strain, we found more differences that could be associated with the attenuated phenotype (Fig. 1). As depicted in Fig. 1a1, one of these

changes is found in the signal peptide (I₃₅ > V > V), four of them are at the middle portion of G1 (T₁₆₈ > A > A; E₁₈₆ > E > G; S₂₀₆ > S > P; and P₂₀₈ > L > L), two are at the carboxyl terminus of G2 (F₄₂₇ > F > I and T₄₄₆ > T > S), and five more are at the amino half of N (V₄₇ > V > E; K₅₉ > R > R; I₁₅₈ > V > V; E₂₆₈ > D > D; and T₃₂₂ > I > I). An alignment of the coding sequences of the L gene of Junín virus strains showed only nine nucleotide changes between XJ13, XJ#44, and Candid#1 strains implicated in amino acid substitutions (Fig. 1b1). Seven of these changes may be related with the attenuation process (H₇₆ > Y > Y; V₄₁₅ > V > A; D₄₆₂ > N > N; L₉₃₆ > L > P; R₁₁₅₆ > K > K; S₁₆₉₈ > S > F; and I₁₈₈₃ > I > V) and two changes (R₈₈₁ > G > R; S₉₂₁ > G > S) could be considered reversions. All these changes are presented as XJ13 > XJ#44 > Candid#1 residues. On the contrary, no changes were found in the amino acid sequence as well as at the nucleotide level of the Z protein, among these three Junín virus strains.

Furthermore, we compared the nucleotide sequences obtained from vaccine-related strains with other reported Junín virus strains. There are only two other Junín virus strains whose genome has been fully sequenced, the Romero and MC2 strains. The Romero strain (incorrectly named “Rumero” in GenBank), classified as a high virulence strain, was isolated from an AHF patient and was passed twice in fetal rhesus lung cells and once in Vero cells [17, 18]. On the other hand, the MC2 strain was isolated from a rodent captured in the endemic area of AHF, and was classified as an intermediate virulence strain [19–21]. Moreover, the complete S RNA and Z gene sequences were obtained for the Junín virus strain IV4454. This strain, classified as an intermediate virulence strain, was isolated from an AHF patient.

Differences between deduced amino acid sequences from the six analyzed Junín virus strains are shown as vertical bars in Fig. 1 and were classified into two types:

Type 1: Positions where the nucleotide or amino acid sequence of one of the field strains of Junín virus (XJ13, Romero, IV4454 or MC2) was different from all other field strains, shown as vertical bars located *below* each RNA diagram in Fig. 1.

Type 2: Positions with mutations among the vaccine strains (XJ13, XJ#44, and Candid#1), shown as vertical bars located *over* each RNA diagram in Fig. 1.

To identify those type-2 mutations that could be more confidently related with the attenuation process, we compared the homologous positions at the genomes obtained from field strains of Junin virus (XJ13, Romero, IV4454 and MC2). At positions GPC₃₅, N₁₅₈, N₂₆₈, and N₃₂₂, the same variations present between vaccine-related strains were found among field strains. At positions L₈₈₁ and L₉₂₁,

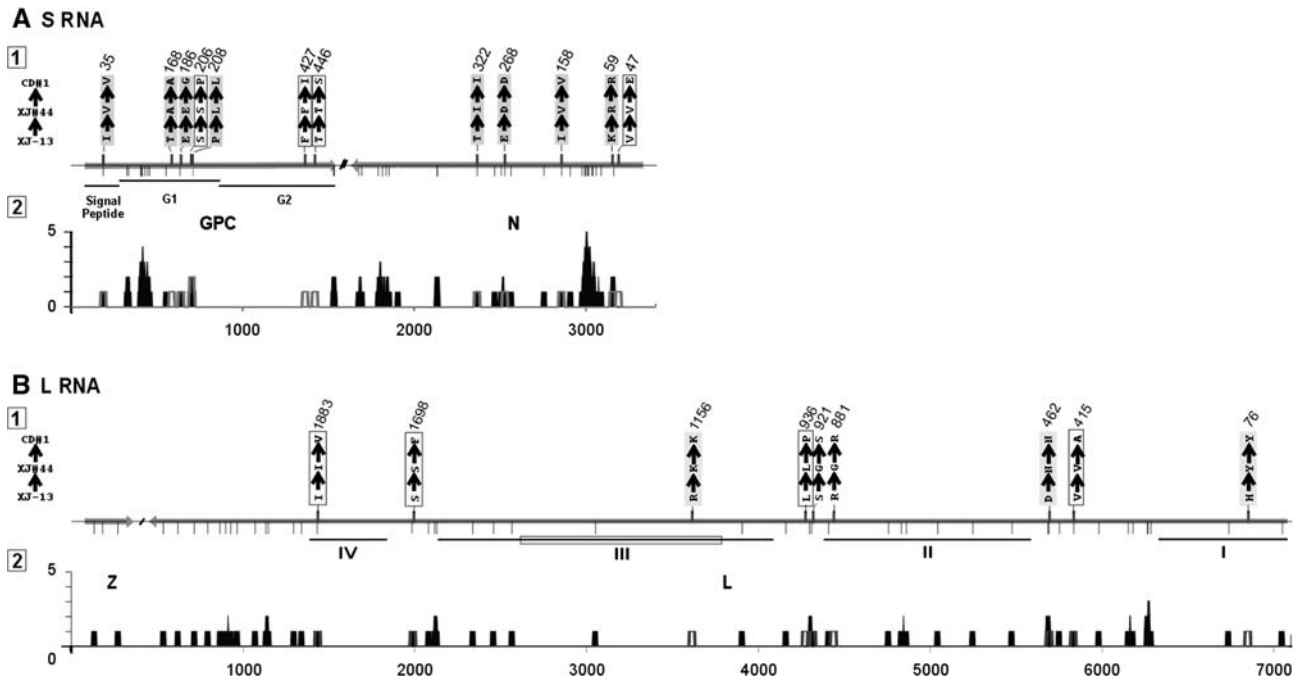


Fig. 1 Schematic of the changes detected in Junín virus proteins. Comparisons were done among all fully sequenced genomes from different Junín virus strains. A nucleotide rule is depicted below the diagram to facilitate the location of each position. **a** S RNA, **1** the open reading frames corresponding to the N and GPC genes are shown as *open rectangles* with *arrowheads* indicating the direction of translation. Non-coding sequences are shown as *horizontal thin lines*. The three cleavage products of GPC protein are shown by *horizontal lines* below the diagram. Amino acid changes detected between vaccine genealogy strains (XJ13, XJ44, and Candid#1; type-2 mutation) are represented as *vertical lines* over the genes. Above them, the detected changes and the position are indicated. Amino acid

changes between field strains of Junín virus (XJ13, Romero, IV4454 and MC2; type-1 mutation) are indicated as *vertical lines* below the genes; **2** plot of relative mutation frequency for type-1 mutations (*black areas*) and for type-2 mutations (*bold line*). **b** L RNA, **1** the open reading frames corresponding to the Z and L genes and the position of the changes in the amino acid sequence are shown as in (**a**). The four conserved regions of the RNA polymerase of arenaviruses as described by Vieth et al. [24] are shown by *horizontal lines* below the diagram of the L RNA. Inside region III, the polymerase domain is signaled with an *open box*; **2** plot of relative mutation frequency for type-1 mutations (*black areas*) and for type-2 mutations (*bold line*)

there is a variation from XJ#44 to Candid#1. However, Candid#1-derived sequence is identical to the field strains (including XJ13)-derived sequences. Thus, it is probable that these positions represent naturally occurring mutations, not related with the attenuation process. Furthermore, sequence variations among field strains are subject to natural selection pressure, whereas sequence variations among vaccine genealogy strains were subjected to an arbitrary selection pressure. We calculated a mutation frequency index defined as number of mutations per amino acid, and the graph was constructed with a program designed by J. A. Iserte (unpublished results), using an overlapped windows-based strategy (11 residues) adding up the number of mutations within each window and plotting the value at the middle of the window. This analysis was made along the entire Junín virus genome for both types of mutations (Fig. 1a2, b2). The regions of the S or L RNA with a high mutation frequency index for type-1 mutations are shown as closed black areas, while regions with a high mutation frequency index for type-2 mutations are shown as open,

bold outlined, areas. Mutations of type 2 found outside black areas, identified at S RNA positions GPC₁₆₈, GPC₄₂₇, GPC₄₄₆, and N₄₇, and L RNA positions L₇₆, L₉₃₆, L₁₁₅₆, could be more confidently involved in the virulence attenuation process.

The non-coding regions at the genomic ends are highly conserved among analyzed strains, varying between 93 and 97% of nucleotide sequence homology. The 3'-non-coding region, which in the virions shows a high degree of complementarity with the 5' non-coding region, exhibits very few differences in independent clones of each strain and varies only slightly from one strain to another.

On the other hand, when infected cell-derived RNAs were sequenced, a high degree of sequence variability was observed at the 5'- and 3'-non-coding regions among RNAs derived from the same viral strain. We performed a RACE analysis to examine specifically the genomic or the antigenomic forms of Junín virus RNAs. As a result, a series of non-template bases were found at the 5'-end of Candid#1 L and S RNAs (Fig. 2b, shadowed). In the

comparison between the 5'-end of genomic RNAs and the 3'-end of antigenomic RNAs (which are used as a template for the former), at least one additional guanine was present in all 5'-end genomic clones comprising extra bases, similarly to what has been detected for other arenaviruses [22]. The 3'-RACE analysis of genomic S and L RNAs obtained from Candid#1-infected cells rendered several clones harboring short deletions (Fig. 2a, c). RNA secondary structure analysis from Candid#1 S and L RNAs predicted a

panhandle structure between 5'- and 3'-ends of both genomic RNAs. Deletions at the 3'-end were localized inside this panhandle (Fig. 2e). These results are consistent with a model involving the use of cellular RNAs to prime the viral RNA synthesis and the use of 5'-end sequences from viral RNA into a non-completed panhandle structure, as template for the 3'-end sequence completion.

Interestingly, comparison between 5'- and 3'-end sequences from both genomic RNAs (S and L segments) showed highly conserved positions (Fig. 3). These positions could be related to the minimal viral promoter sequence. The 5'- and 3'-non-coding sequences from Candid#1 S RNA have approximately 80 nucleotides in length, similarly to the 5'-end from L RNA non-coding sequences. However, 3'-non-coding sequences from L RNA have only 30 nucleotides in length. When comparing the 80 nucleotides from non-coding sequences, the homology between 5'-end sequences of the L and S RNAs was of 60%, while that from 3'-end sequences was only 50%. However, if we compare only 30 nucleotides from the 3'-end of both genomic RNAs the homology of this region ascends to 71%. Because both, genomic or antigenomic, non-coding regions must be recognized by the viral RNA polymerase to complete the viral replicative cycle, the promoter region should be present in the first 30 nucleotides of the antigenomic L RNA. Genomic ends of arenaviruses comprise a highly conserved region of 19 nucleotides, called the arena region. Outside this region the 3'-end of Junin virus L RNA comprises only 11 nucleotides (GCTCAAGTGCC). These nucleotides show a high degree of homology with two regions of S RNA 3'-end sequences (Fig. 3a, shadow boxes). Thus, two boxes at the S RNA appear to match with a unique box at the L RNA. The S RNA boxes (positions 1–12 and 35–45 in the alignment) could be related to translation or transcription process. Other arenaviruses have similar characteristics at their genomic ends. For example, an extended analysis using the genomic sequences from other New World arenaviruses belonging to B1 subclade (Machupo, Junín and Tacaribe viruses; [23]), show that the 38–46 box (GCUCAAGUG for the Junin virus L RNA and GCUCAGUG for the Junin virus S RNA) was conserved among members of the group (Fig. 3b, shadow box). Consequently, it is possible that a sequence motif could be present at the 3'-end of both genomic RNAs of arenaviruses. Furthermore, for Junín, Machupo and Tacaribe viruses, this motif, described by the sequence GSYC(A)₁₋₂GUR, shows a relative degree of conservation in position in the RNA secondary structure calculated by bioinformatics tools (Fig. 3c).

An independent parsimony analysis was done for each arenavirus gene (GPC, N, Z, and L). A binary tree-file was constructed adding the obtained parsimony tree-files from each gene in order to obtain a consensus tree for the four genes.

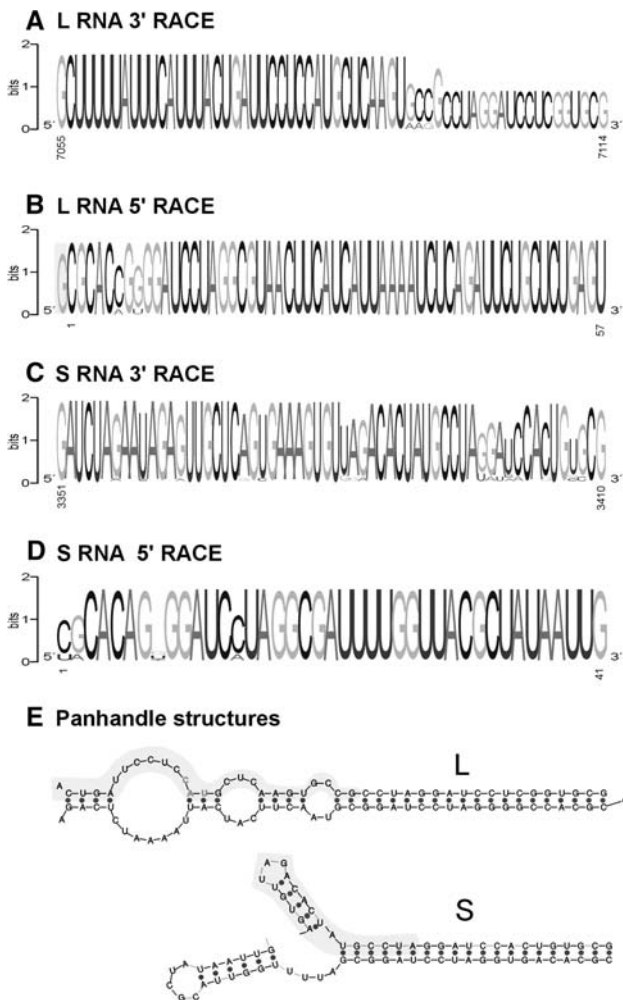
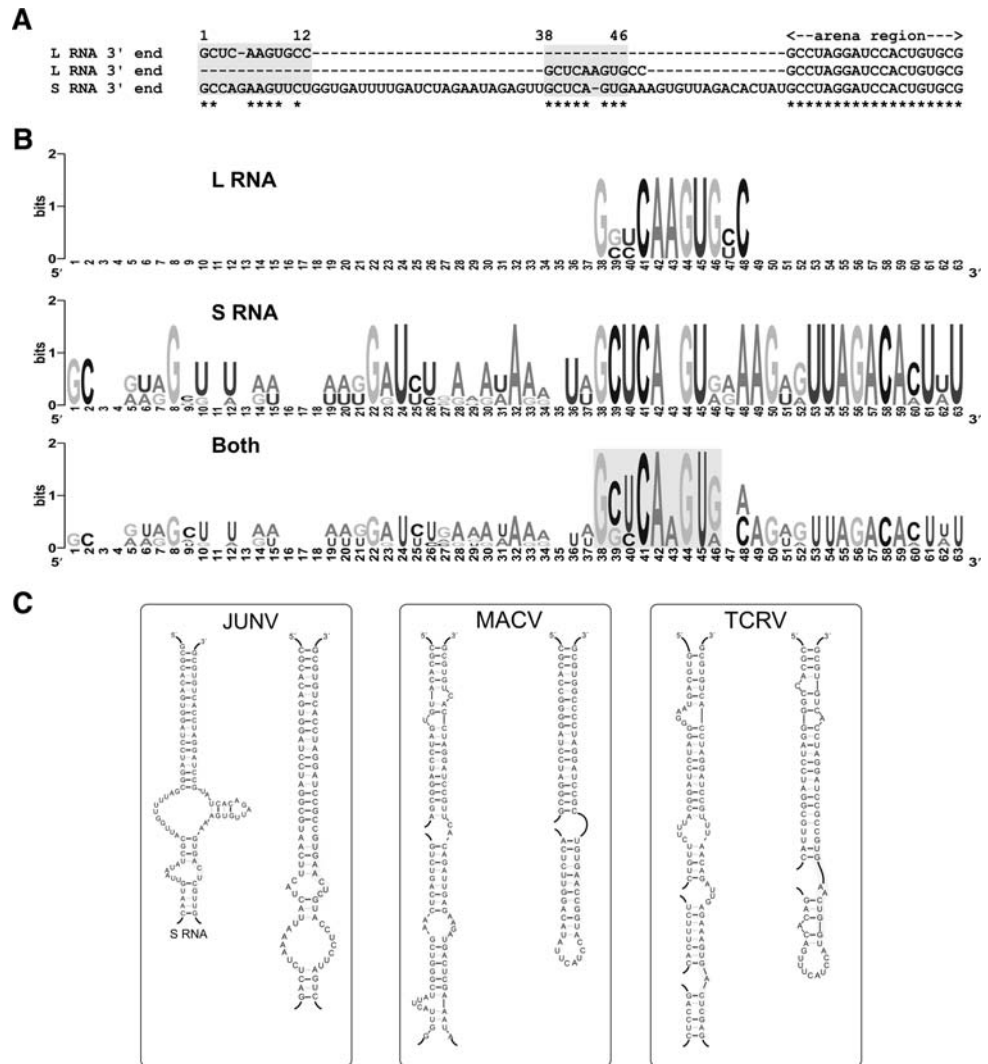


Fig. 2 End sequence determination of Junín virus, Candid#1 strain RNAs. **a** Sequence logos representing the distribution in the last 60 nucleotides of 3'-end sequence of L RNA clones as determined by a genome specific RACE technique, **b** sequence logos representing the distribution in the first 60 nucleotides of 5'-end sequence of L RNA clones, determined by a genome-specific RACE technique. *Shadowed box* corresponds to an additional extended G of the genomic sequence, **c** sequence logos representing the distribution in the last 60 nucleotides of 3'-end sequence of S RNA clones, determined by RACE technique, **d** sequence logos representing the distribution in the last 60 nucleotides of 5'-end sequence of S RNA clones, determined by RACE technique, **e** panhandle structures predicted for Candid#1, L and S RNAs. Shadowed with *gray* is the region present in all 3'-end clones

Fig. 3 Junín virus Candid#1 strain, S and L RNA 3'-ends. **a** Alignment between Junín virus S and L RNAs 3'-end sequences. *Shadowed boxes* indicate the homology of one L RNA sequence block with two different regions at the S RNA 3'-end. The "arena region" indicates the 19-nt long sequence shared by the genomic RNAs of all arenaviruses, **b** sequence logos obtained after comparison of 3'-end sequences from L and S RNAs (excluding arena region). Alignments were done for L RNA, S RNA, or L plus S RNA (both) sequences from New World arenaviruses belonging to B1 subclade. The region shadowed corresponds to the sequence consensus motif described in the text. **c** panhandle structures predicted for L and S RNAs. To show location, the above-mentioned motif was outlined



Clades and subclades of Old World and New World arenaviruses are represented according to Charrel et al. [6]. Phylogenetic analysis shows that all Junín virus strains (Candid#1, XJ13, XJ#44, MC2, and Romero) group together with other hemorrhagic New World arenaviruses. Only the subclade B1 branch structure of the arenaviral tree is shown (Fig. 4).

Discussion

Candid#1, the most attenuated Junin virus strain, has a set of putative attenuation markers in its GPC-, N-, and L-protein ORFs. We propose that changes found in the genomic regions harboring a low wild-type mutation frequency index (GPC₁₆₈, GPC₄₂₇, GPC₄₄₆, N₄₇, L₇₆, L₉₃₆, L₁₁₅₆), could be more confidently associated with the virulence attenuation process. For L proteins, Vieth et al. [24] describe four conserved regions among all arenaviruses. Within region III, Lan et al. [20] found the polymerase

domain, and proposed the presence of four motifs: A, B, C, and D. In this region, we only detected one change (R > K > K) at the 1156 position. Although this change is classified as conservative, it could be related to the attenuation process because this region does not seem to support naturally occurring mutations. Other changes within L are located between regions II and III (L₉₃₆, L > L > P) or inside region I (L₇₆, H > Y > Y). Change L₉₃₆, in spite of being present in a region of high index for type-2 mutations, could only be associated with a structural change of the Candid#1-derived L protein based on the structural characteristics of the Proline residue. The change in L₇₆ falls near the putative ATP/GTP-binding site (P-loop) predicted using the ExPasy web site. Although our results suggesting the involvement of the RNA polymerase in the attenuation of virulence are preliminary, they are consistent with reports on other viruses [25–27]. Furthermore, some changes in the structural proteins, nucleoprotein, and both mature glycoproteins, could also be related with the

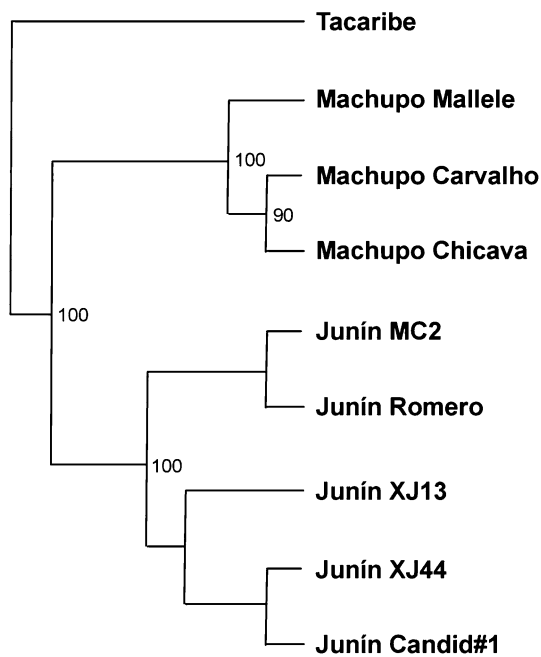


Fig. 4 Phylogenetic analysis. Neighbor joining phylogeny was done for the four arenaviral RNA-derived ORF sequences of New World arenaviruses belonging to Subclade B1. Clades supported by bootstrap values of over 70 are indicated

attenuation of virulence. The carboxyl-terminus of N protein, which contains a zinc-binding domain [28] is highly conserved, and the N_{47} change ($V > V > E$) falls outside this region and would be associated with another characteristic of the protein. As is known, the N protein has a dual function during the virus life cycle. First, it is involved in essential steps of genome replication, promoting the synthesis of the full-length antigenomic copy of S RNA, and second, it associates with the genomic RNA to form the nucleocapsid. For the glycoprotein precursor, we found a mutation in the carboxy-terminus of G1 (GPC₁₆₈, $T > A > A$). This change directly affects the conserved sequence (N₁₆₆R₁₆₇T₁₆₈K₁₆₉) for the principal N-glycosylation site predicted by NETNGLYC 1.0. The glycoprotein G2 has three domains, the outer (carboxyl-terminus) domain located outside the virion, the transmembrane domain, and the inner (amino-terminus) domain, located inside the virion. The outer domain interacts with G1, the transmembrane domain interacts with the signal peptide, and the inner domain could interact with Z or the nucleocapsid [29, 30]. G2 changes fall inside the transmembrane domain (GPC₄₂₇, $F > F > I$) or at the inner domain (GPC₄₄₆, $T > T > S$) and could affect such important interactions.

It has been previously suggested that changes in the intergenic region could play a role in the attenuation processes of arenaviruses [31]. However, our sequence analysis of the intergenic regions from XJ13, XJ#44, and

Candid#1 revealed 100% conservation in both genomic RNAs. If nucleotide changes were not tolerated in this region, it could suggest that a major evolutionary constraint is operating, perhaps related to the calculated secondary structure conformation and its proposed function in the transcription regulation process [32, 33]. The S RNA nucleotide sequence was less conserved than that of the L RNA, indicating a faster rate of evolution in the S-encoded polypeptides. Lan et al. [25] reported the genome comparison of virulent and avirulent strains of the Pichinde arenavirus, and found a lower number of attenuation process-related mutations, but at comparable genomic regions. If we only compare the field strains of Junín virus (MC2, Romero, IV4454, and XJ13), the mutations distribute very differently among large and small segments. We considered specific changes for MC2, Romero, or IV4454 when their sequences differed from XJ13 sequence, and specific changes for XJ13 when its sequence differed from all three of the previous Junín virus strains. Analyzing the protein sequences of the field strains, there were 45 divergence sites for the S RNA ORFs and 48 for the L RNA ORFs.

In the analysis of the non-coding regions a high degree of sequence variability has been observed at the 5'- and 3'-genomic ends among RNAs derived from the same strain. This heterogeneity could have arisen from different transcription-related editing of the subgenomic RNAs, reported previously for other arenaviruses [34]. Our results are consistent with a model involving the use of cellular RNAs to prime the viral RNA synthesis and the use of 5'-end sequences from viral RNA into a non-completed panhandle structure, as template for the 3'-end sequence completion. However, the contribution of these regions in the attenuation process remains to be evaluated. On the other hand, we also described a strongly conserved motif at the 3'-end of arenaviral genomic RNAs that could function as a viral promoter (Fig. 3). Although we have not yet experimentally associated any function with the conserved motif, it is a primary target to examine using other methodologies, such as a reverse genetics system. This is the first description of the conserved motif for arenaviruses.

As shown in Fig. 4, the phylogeny correlates with the genealogy of the vaccine strain, Candid#1, and a small set of nucleotide changes seems to be central to define the phenotypic variation from virulence to attenuation. If this is confirmed in a more extensive study, any surveillance program designed to monitor the natural vaccine variations should search for possible point mutations at those positions related with attenuation.

In summary, the present work shows a set of mutations that could be related to the virulence attenuation phenomenon. Furthermore, most of both described types of mutations (type 1 and type 2) could be grouped into a few regions of the genome (Fig. 1). Based on these results, we

propose to design a set of primers in order to generate PCR fragments, comprising most detected mutations that could be directly sequenced. These primers could be used in a rapid screening method, based on RT-PCR and nucleotide sequencing, in order to analyze genomic variability on field samples and to search for the presence of type-1 or -2 mutations between field strains, allowing to observe the biodiversity in nature or to develop an epidemiological surveillance program of vaccinated people. The information accumulated by sequence analysis of viral genomes with different degrees of virulence will certainly serve as a starting point to study this biological phenomenon, provided that a reverse genetics system for Junín virus is developed to allow the generation of infectious virions with specific mutations.

Acknowledgments We are very grateful to Dr. J. G. Barrera Oro from USAMRIID (Frederick, Maryland, USA) who provided us with a XJ#44 seed, to Dr. Ana María Ambrosio and Dr. María del Carmen Saavedra who prepare Candid#1 virions and Dr. Diego Golombek who carefully read this manuscript. This work was supported by PICT 38138 from ANPCYT and PIP 5813 from CONICET (Argentina). S. E. Goñi, B. I. Stephan, and C. S. Borio have a research fellowship from CONICET, J. A. Iserte holds a research fellowship from ANPCYT, P. D. Ghiringhelli and M. E. Lozano are members of the research career of CONICET.

References

- J.C.S. Clegg, M.D. Bowen, M.J. Buchmeier, J-P. Gonzalez, I.S. Lukashevich, C.J. Peters, R. Rico-Hesse, V. Romanowski. *Arenaviridae*, in *Virus Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses*, ed. by M.H.V. van Regenmortel, C.M. Fauquet, D.H.L. Bishop, E.B. Carstens, M.K. Estes, S.M. Lemon, J. Maniloff, M.A. Mayo, D.J. McGeoch, C.R. Pringle, R. B. Wickner (Academic Press, New York, 2000), pp. 633–640
- C.A. Tidona, G. Darai, *The Springer Index of Viruses* (Springer, New York, 2001)
- D.A. Enria, A.M. Briggiler, Z. Sanchez, *Antiviral Res.* **78**(1), 132–139 (2008)
- T. Strecker, R. Eichler, J. van der Meulen, W. Weissenhorn, H.D. Klenk, W. Garten, O. Lenz, *J. Virol.* **77**(23), 12927 (2003)
- M. Perez, D.L. Greenwald, J.C. deLa Torre, *J. Virol.* **78**(20), 11443–11448 (2004)
- R.N. Charrel, X. de Lamballerie, S. Emonet, *Curr. Opin. Microbiol.* **11**, 362–368 (2008)
- J.C. de la Torre, *Antiviral Res.* **80**(3), 239–250 (2008)
- S.E. Goñi, J.A. Iserte, A.M. Ambrosio, V. Romanowski, P.D. Ghiringhelli, M.E. Lozano, *Virus Genes* **32**, 37–41 (2006)
- A.M. Ambrosio, M.C. Saavedra, L.M. Riera, R.M. Fassio, *Acta Bioquím. Clín. Lat.* **40**(1), 5–17 (2006). (ISSN 0325-2957)
- D. Enría, J.G. Barrera Oro, *Curr. Top. Microbiol. Immunol.* **263**, 239–261 (2002)
- A.S. Parodi, D.J. Greenway, H.R. Rugiero, E. Rivero, M.J. Frigerio, N.E. Mettler, F. Garzon, M. Boxaca, L.B. Guerrero, N.R. Nota, *Día Médico (Buenos Aires)* **30**, 2300–2302 (1958)
- J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, *Nucleic Acids Res.* **25**, 4876–4882 (1997)
- M. Zuker, *Nucleic Acids Res.* **31**(13), 3406–3415 (2003)
- E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel, A. Bairoch, *Nucleic Acids Res.* **31**, 3784–3788 (2003)
- R. Gupta, E. Jung, S. Brunak, Prediction of N-glycosylation sites in human proteins, in preparation (2004) (available at: <http://www.cbs.dtu.dk/services/NetNGlyc/>)
- B.C. Schaefer, *Anal. Biochem.* **227**, 255–273 (1995)
- K.T. McKee Jr., B.G. Mahlandt, J.I. Maiztegui, G.A. Eddy, C.J. Peters, *J. Infect. Dis.* **152**, 218–221 (1985)
- N.E. Yun, N.S. Linde, N. Dziuba, M.A. Zacks, J.N. Smith, J.K. Smith, J.F. Aronson, O.V. Chumakova, H.M. Lander, C.J. Peters, S. Paessler, *Am. J. Trop. Med. Hyg.* **79**(2), 275–282 (2008)
- M.I. Berría, L.F. Gutman Frugone, R. Girda, J.G. Barrera Oro, *Medicina (Buenos Aires)* **27**, 93–98 (1967)
- N.A. Candurra, E.B. Damonte, C.E. Coto, *J. Med. Virol.* **27**, 145 (1989)
- M.C. Weissenbacher, R.P. Laguens, C.E. Coto, *Curr. Top. Microbiol. Immunol.* **134**, 79–116 (1987)
- S.J. Polyak, S. Zheng, D.G. Harnish, *J. Virol.* **69**(5), 3211–3215 (1995)
- M.L. Flanagan, J. Oldenburg, T. Reigner, N. Holt, G.A. Hamilton, V.K. Martin, P.M. Cannon, *J. Virol.* **82**(2), 938–948 (2008)
- S. Vieth, A.E. Torda, M. Asper, H. Schmitz, S. Günter, *Virology* **318**, 153–168 (2004)
- S. Lan, L. McLay, J. Aronson, H. Ly, Y. Liang, *Arch. Virol.* **153**(7), 1241–1250 (2008)
- M.J. Endres, C. Griot, F. Gonzalez-Scarano, N. Nathanson, *J. Virol.* **65**, 5465–5470 (1991)
- Y. Riviere, R. Ahmed, P.J. Southern, M.J. Buchmeier, M.B.A. Oldstone, *J. Virol.* **55**, 704–709 (1985)
- M.A. Tortorici, P.D. Ghiringhelli, M.E. Lozano, C.G. Albariño, V. Romanowski, *J. Gen. Virol.* **82**, 121–128 (2001)
- A.A. Capul, M. Perez, E. Burke, S. Kunz, M.J. Buchmeier, J.C. de la Torre, *J. Virol.* **81**(17), 9451–9460 (2007)
- J. York, V. Romanowski, M. Lu, J.H. Nunberg, *J. Virol.* **78**(19), 10783–10792 (2004)
- S.M. Wilson, J.C.S. Clegg, *Virology* **180**, 543–552 (1991)
- M.T. Franze-Fernandez, S. Iapalucci, N. Lopez, C. Rossi, in *The Arenaviridae*, ed. by M.S. Salvato (Plenum Press, New York, 1993), pp. 113–132
- M.A. Tortorici, C.G. Albariño, D.M. Posik, P.D. Ghiringhelli, M.E. Lozano, R.V. Rivera Pomar, V. Romanowski, *Virus Res.* **73**, 41–55 (2001)
- D. Garcin, D. Kolakofsky, *J. Virol.* **66**(3), 1370–1376 (1992)