Original Article

The Bayesian reconstruction and the evolutionary history of Salivirus type 1 and type 2: the worldwide spreading

Silvia Angeletti[1][#], Domenico Benvenuto[2][#], Marta Fogolari[1], Cecilia de Flora[2], Giancarlo Ceccarelli[3,4], Ivana Maida[5], Riccardo Bazzardi[6], Silvia Spoto[7], Stefano Pascarella[8], Boban Mugosa[9], Massimo Ciccozzi[2]

[1] *Unit of Clinical Laboratory Science, University Campus Bio-Medico of Rome, Italy*
[2] *Unit of Medical Statistics and Molecular Epidemiology, University Campus Bio-Medico of Rome, Italy*
[3] *Department of Public Health and Infectious Diseases, University of Rome, "Sapienza", Policlinico "Umberto I", Rome, Italy*
[4] *Migrant and Global Health Research Organization, Centro di ricerca sulla salute globale e delle popolazioni mobili (Mi-Hero) e Sanità di Frontiera Onlus, Rome, Italy*
[5] *Department of Medical, Surgical and Experimental Sciences Infectious Diseases, University of Sassari, Italy*
[6] *Controllo Microbiologico e Ispezione degli Alimenti, Istituto Zooprofilattico Sperimentale della Sardegna "G. Pegreffi", Sassari, Italy*
[7] *Diagnostic and Therapeutic Medicine Department, University Campus Bio-Medico of Rome, Rome, Italy*
[8] *Department of Biochemical Science "A. Rossi Fanelli", Sapienza University, Rome, Italy*
[9] *Institute of Public Health, Podgorica, Montenegro*

\# Authors contributed equally to this work.

Abstract

Introduction: Salivirus (SalV) represents an emerging problem in public health especially during the recent years. In this study, the Bayesian evolutionary history and the spread of the virus through the different countries have been reported.

Methodology: a database of 81 sequences of SalV structural VP1 fragment were downloaded from GenBank, aligned and manually edited by Bioedit Software. ModelTest v. 3.7 software was used to estimate the simplest evolutionary model fitting the sequence dataset. A Maximum-Likelihood tree has been generated using MEGA-X to test the "clockliness" signal using TempEst 1.5.1. The Bayesian phylogenetic tree was built by BEAST. Homology modelling was performed by SWISS-Model and protein variability evaluated by ConSurf server.

Results: the phylogenetic tree showed a clade of SalV A2 and three main clades of SalV A1, revealing several infections in humans in South Korea, India, Tunisia, China, Nigeria, Ethiopia and USA. The Bayesian maximum clade credibility tree and the time of the most common recent ancestor dated back the root of the tree to the year 1788 with the probable origin in USA. Selective pressure analysis revealed two positive selection sites, His at 100th and Leu at 116th positions that at the homology modelling resulted important to guarantee protein stability and variability. This could contribute to the development of new mutations modifying the clinical features of this evolving virus.

Conclusions: Bayesian phylogenetic and phylodynamic represented a useful tool to follow the transmission dynamic of SalV and to prevent new epidemics worldwide.

Introduction

Acute Gastroenteritis is considered one of the most significant and still prevalent diseases worldwide [1,2]. Even though sanitation and prevention strategies has determined a relevant decreasing of the mortality rate for diarrhea from 15% in 2008, to about 9% in 2015, infectious diarrheas are still a concerning problem in public health both in developing and industrialized countries [2,3]. Effective surveillance systems are critical for outbreak detection and timely implementation of public health interventions. Salivirus (SalV), previously named Klassevirus, a non-enveloped single-strand RNA virus with a genome length of 7.4 kb belonging to the Picornaviridae family, was recently described as an emerging problem in public health. Several studies, reported SalV as cause of acute gastroenteritis in children [4-6]. SalV as causal agent of acute gastroenteritis was described in 0.2-8.7 % of cases

in different countries worldwide [7,8] with highest prevalence in Asian countries [9,10].

SalV is closely related to Aichivirus, a member of the *Kobuvirus* genus that similarly was detected in fecal samples as well in sewage from different countries [4-6]. Based on capsid region variation, SalV is classified in SalV A1 and SalV A2 [11]. SalV infection has been described in three cases of unexplained diarrhea from the United States, Australia and China and in sewage from Spain and Thailand [12-16]. Gastrointestinal apparatus seems not the only source of transmission, because in China this virus has been found in respiratory samples from a child with adenovirus infection showing different symptoms and representing a new way of spreading [17]. In a recent study, toilet wastes from different international airplanes resulted positive for several pathogens including enteric and respiratory virus. Especially in case of plane from South Asia, a high prevalence of SalV and Aichivirus have been described compared to samples collected from North Asia and North America [9,18,19]. At today, few reports are available about molecular epidemiology of this virus and all of them focusing on viral genotyping evaluated the genetic diversity of the virus as causal agent of acute gastroenteritis worldwide [7,8,11,18,20]. The aim of this study, using the genotype sequences available in GeneBank, was to apply the Bayesian evolutionary methods to date the origin of SalV infection worldwide and to track the spread of the virus through the different countries where epidemics have been reported. By phylogenetic analysis, the way of transmission of SalV can be clearly defined through the ancestor strain circulation dating and geographic movement tracing. This approach is extremely useful for prompt preventive action and surveillance plan aimed to limit the further spread of the virus worldwide.

Methodology
Phylogenetic analysis

A dataset of all available VP1 sequences including 75 Salivirus A1 and 6 Salivirus A2 was built by downloading from GenBank (http://www.ncbi.nlm.nih.gov/genbank/). The selection criteria for sequences were known sampling date and geographical location (Table 1). The countries included in the study were: China, Ethiopia, Germany, Guatemala, Hungary, India, Nigeria, Spain, South Korea, Thailand, Tunisia, USA and Venezuela.

The 81 sequences were aligned using the Bioedit program v7.0.5 and manually edited, as already described [21], obtaining an alignment of 852bp length.

Table 1. Dataset used for the analysis. Accession number, place and year of isolation have been reported for each sequence.

Accession Number	Place of isolation	Year
JQ898343.1	Thailand	2009
MG571816.1	Venezuela	2016
MG571815.1	Venezuela	2015
KM023140.1	China	2011
NC025114.1	China	2013
KP247440.1	China	2013
GQ253936.1	USA	2008
MG026493.1	Ethiopia	2016
NC012957.1	Nigeria	2007
GQ179640.3	Nigeria	2009
MG026494.1	Ethiopia	2016
MG026496.1	Ethiopia	2016
MG026495.1	Ethiopia	2016
KT310068.1	Guatemala	2009
GQ253933.1	Spain	2008
GQ507022.1	Nigeria	2007
KX091899.1	India	2010
KX091903.1	India	2011
KX091890.1	India	2008
KX091887.1	India	2007
KX091902.1	India	2011
KX091891.1	India	2008
KX091886.1	India	2007
KX091894.1	India	2009
KX091892.1	India	2008
KX091895.1	India	2009
KP247439.1	Germany	2013
KC196335.1	South Korea	2012
HM038153.1	South Korea	2009
HM038162.1	South Korea	2009
HM038165.1	South Korea	2009
HM038166.1	South Korea	2009
HM038164.1	South Korea	2009
HM038157.1	South Korea	2009
HM038161.1	South Korea	2009
HM038159.1	South Korea	2009
HM038152.1	South Korea	2009
KU362771.1	Tunisia	2011
KU362772.1	Tunisia	2011
KU362770.1	Tunisia	2011
KU362767.1	Tunisia	2011
KU362769.1	Tunisia	2011
HM038170.1	South Korea	2009
HM038160.1	South Korea	2009
HM038163.1	South Korea	2009
HM038168.1	South Korea	2009
HM038154.1	South Korea	2009
HM038169.1	South Korea	2009
HM038150.1	South Korea	2008
HM038149.1	South Korea	2007
KT240115.1	Hungary	2013
KC196328.1	South Korea	2008
KC196329.1	South Korea	2010
KT182636.1	China	2013
KU362768.1	Tunisia	2011
HM038148.1	South Korea	2007
HM038158.1	South Korea	2009
HM038155.1	South Korea	2009
HM038151.1	South Korea	2009
HM038156.1	South Korea	2009
KX091896.1	India	2009
KX091901.1	India	2011
KX091889.1	India	2008
HM038167.1	South Korea	2009

The recombination signal was analysed by the pairwise homoplasy index (Phi) test using SplitsTree4 software [22].

Invariant sites estimation has been performed in order to perform Xia test to estimate substitution saturation and the graphical exploration of the dataset of sequences used in the analysis by mean of DAMBE-7 software [23]. The likelihood mapping analysis of 10,000 random quartets by using TreePuzzle software has been used in order to investigate the phylogenetic signal as already described [24]. The best-fitting nucleotide substitution model was chosen by JModeltest software [25]. Adaptive Evolution Server (http://www.datamonkey.org/) was used to find eventual sites of positive or negative selection. At this purpose the following tests has been used: fast unconstrained Bayesian approximation (FUBAR) [26], adaptive Branch Site REL [27], Bayesian un-restricted test (BUSTED) [28] and the mixed effects model of evolution (MEME). These tests allowed to infer the site-specific pervasive selection, the episodic diversifying selection across the region of interest and to identify episodic selection at individual sites [29]. Statistically significant positive or negative selection was based on p value < 0.05 [29].

By MEGA-X software [30] an ML tree was reconstructed using Hasegawa Kishino Yano (HKY) plus Gamma distribution (HKY+G) as evolutionary model selected by JModelTest.

The correlation between sampling dates and genetic distances from the tips to the root was explored using TempEst to evaluate the robustness in terms of molecular clock of the dataset [31].

Bayesian phylogenetic analysis

The evolutionary rate, the dated and the phylogeographic trees were estimated using a Bayesian Monte Carlo Markov Chain (MCMC) approach. Strict and relaxed clock models have been tested using HKY+G model. Two parametric (constant size, exponential) and three non-parametric (Bayesian skyline plot (BSP), Bayesian GMRF and Bayesian Skygrid) population models were compared.

By means of a Bayes factor (BF, using marginal likelihoods) the best fitting models were selected and implemented in Beast v. 1.10.4. In reference of previous studies such as Villano *et al.*, the strength of the evidence against H0 (null hypothesis) was evaluated as follows: 2lnBF < 2 = no evidence; 2–6 = weak evidence; 6–10 = strong evidence; and > 10 = very strong evidence. A negative 2lnBF indicates evidence in favor of H0. Only values ≥ 6 were considered

Table 1 (continued). Dataset used for the analysis. Accession number, place and year of isolation have been reported for each sequence.

Accession Number	Place of isolation	Year
KX091897.1	India	2009
KX091898.1	India	2010
KX091885.1	India	2007
KX091893.1	India	2008
KX091900.1	India	2011
JN379039.1	China	2011
KX091888.1	India	2007
KC196336.1	South Korea	2012
KC196334.1	South Korea	2011
KC196330.1	South Korea	2011
KC196333.1	South Korea	2012
KC262668.1	South Korea	2011
KC196332.1	South Korea	2011
GU245894.1	China	2009
GQ184145.1	USA	2002
NC012986.1	USA	2002

significant [32]. The MCMC chains has been run for 75 million generations and sampled every 7500 steps. Convergence was assessed by estimating the effective sampling size (ESS) after removing a 10% burn-in, by Tracer software, version 1.7 [33], ESS values ≥ 250 were accepted. The posterior probability of each mono-phyletic clade ≥ 0.89 was considered statistically significant. The obtained tree was summarized using Tree Annotator software.

A map of the SalV infections has been generated using SPREAD software v1.0.7 [34].

Homology modelling

Homology models have been built relying on the website SwissModel [35]. Structural templates have been searched and validated using the software available within the SwissModel environment and HHPred [36]. Homology models have been validated using the QMEAN tool [37]. Three-dimensional structures have been analyzed and displayed using PyMOL. Mapping of evolutionary conserved regions onto homology models have been carried out with the ConSurf website [38]. To map the structural variability of the variant VP1 of SalV fragments and their sites under selection pressure, homology modelling has been applied on the templates JQ898343 and JN379039. Prediction of the impact of single mutations on the protein stability have been predicted with the SDM server [39]. Sequence alignment and display utilized Jalview [40].

Results
Phylogenetic analysis

The Pairwise Homoplasy index (Phi) test did not detect any significant signal of recombination (p-value = 0.104).

Figure 1. Maximum clade credibility (MCC) tree with Bayesian phylogeography reconstruction of SalV sequences. Branches are scaled in time and colored according to the legend to the left where each color represents the geographic location of the sampled sequence (tip branches), as well as of the ancestral lineage (internal Branches) inferred by Bayesian phylogeography. Significant posterior probability support as indicated by a diamond colored according to location as shown in the Legend. The node age of the branches represents significant statistical support for the clade subtending that branch (posterior probability>0.89). The scale at the bottom of the tree represents time in years. The years are indicated in the internal nodes; different clusters are highlighted.
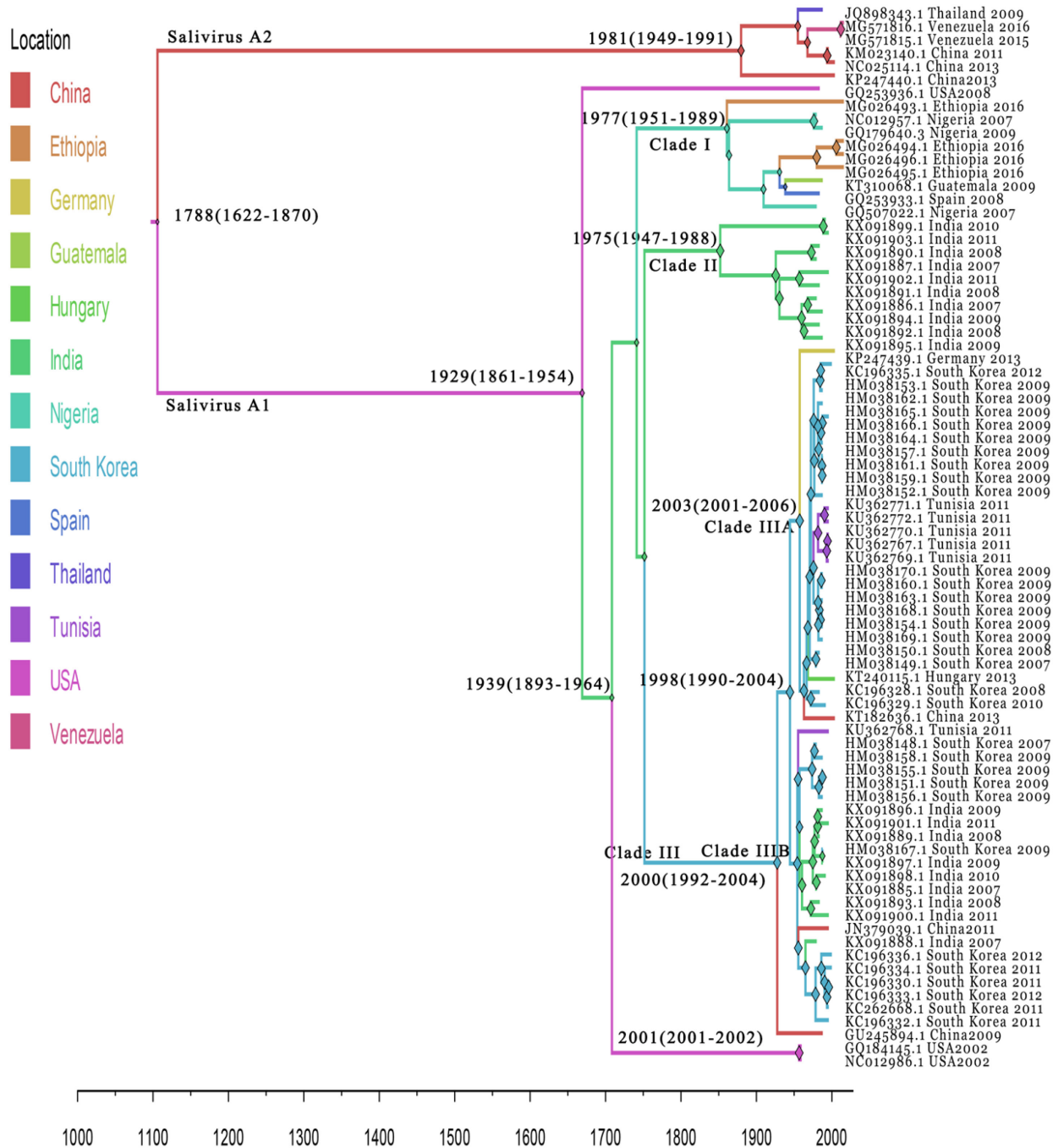
The proportion of invariants sites by iteration found a P (invariant) of 0.6154. Consequently, the Xia's test for substitution saturation has been performed suggesting little substitution saturation for the data set, since the Iss value was smaller than Issc [41].

The tree Puzzle showed that the percentage of dots falling in the central area of the triangle was 3.9%, since it was lower 30% of noise the dataset contained sufficient phylogenetic signal.

The percentage of the Parsimony sites was 30% while the percentage of the constant sites was 62% confirming a sufficient phylogenetic signal for further analysis.

By FUBAR analysis of 284 sites, significant (p-value < 0.05) pervasive episodic selection was found in 2 sites (100th and 116th nucleotide position using the reference sequence SalV KU362769.1). In sequences from Nigeria, Ethiopia, Usa, China, Guatemala, Spain and Thailand at the 100th position the presence of Histidine has been found, while in the other sequences the most prevalent amino acid was Arginine. This difference in the 100th position is present in the sanitation-resistant sequence isolated from sewage water in Bangkok (JQ898343.1), in the sequence isolated in the nasal swab of a child in china (KT182636.1) and in the sequences isolated in stool samples from Nigerian children suffering from non-polio acute flaccid paralysis (AFP) (GQ179640.3, GQ507022.1). At the 116th position, the presence of Leucine instead of Threonine has been found in sequences from SalV A2 sequences.

Significant (p < 0.05) pervasive negative selection in 179 sites (63%) has been evidenced and confirmed by FUBAR.

Bayesian phylogenetic analysis

The root-to-tip divergence plots showed R value of 0.56, indicating good correlation and high temporal and "clocklikeness" signal between the genetic distance to the root. The selected model (lnBF > 6) was the costant demographic model using strict clock (Supplementary Table 1). The evolutionary rate of the SalV VP1 sequences was 1.0851×10^{-3} substitutions site per year (95% HPD 6.2859×10^{4}–1.5531×10^{3}). In Figure 1, Bayesian maximum clade credibility (MCC) tree was showed.

The time of the most common recent ancestor (tMRCA) dated the root of the tree back to the year 1788 (1622 – 1870). In the MCC tree, three main clades of SalV A1 sequences and an outgroup clade of SalV A2 sequences, statistically supported (posterior probability (pp) > 0.89) were represented.

Figure 2. Map of spreading of SalV generated using Spread v1.0.7. The connected lines represent the virus transmission route.
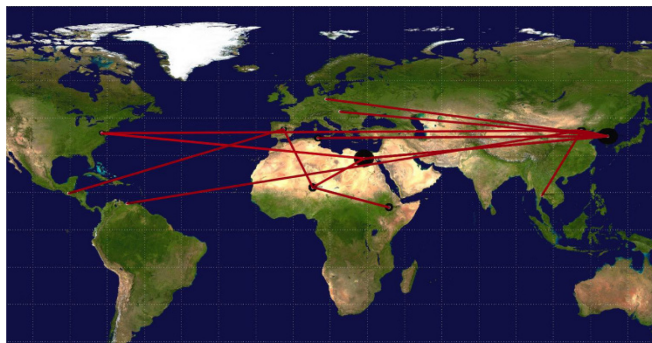
The outgroup clade of SalV A2 including sequences from Thailand, Venezuela and China dated back to 1981 (1949-1991).

The SalV A1 clade dating back to 1929 (1861-1954) included three statistically supported clades I, II and III. Clade I composed of nine sequences isolated from 2007 to 2016 in Guatemala, Nigeria, Spain and Ethiopia dating back to 1977 (1951 – 1989), clade II consisting of ten Indian sequences isolated from 2007 to 2011, dating back to 1975 (1947 - 1988) and clade III was composed by two statistically supported clusters (IIIa and IIIb), dating back to 2000 (1992 - 2004). The tMRCA of cluster IIIa, including 29 sequences, dated back to 2003 (2001 – 2006). Within cluster IIIa different monophyletic epidemics are evidenced. The first included sequences isolated in South Korea from 2007 to 2009, the second more recent included sequences isolated in South Korea from 2009 to 2012, the last involved Tunisia in 2011. In cluster IIIa, sequences isolated during 2013 from Germany, Hungary and China clustered separately. The tMRCA of the cluster IIIb included twenty-three sequences and dated back to 2000 (1992-2004). Inside this cluster, two different monophyletic epidemics involving isolates from South Korea (2007-2012) and one monophyletic epidemic of Indian sequences (2007-2011) were evidenced. Moreover, in this cluster sequences from China and Tunisia clustered separately. Interestingly, isolates from USA dating back to 2001-2002, form a sort of outlier of SalV A1 clade (Figure 1).

In Figure 1, the MCC tree with Bayesian phylogeographic reconstruction of SalV sequences is showed. The probable circulation origin of SalV A1 and A2 dating back to 1788 was USA with a state posterior probability (spp) of 0.19. SalV A2 probably migrated in China (spp of 0.80) in the year 1981 where it persisted to circulate (spp 0.99) and simultaneously probably

migrated to Thailand and Venezuela with a spp of 0.67 and 0.72, respectively.

SalV A1 from USA spread to India (spp 0.36) in 1939 and then branched into three different clades. Clade I originating in Nigeria (spp 0.60) in 1977 spread to Ethiopia and Spain. Clade II originated in India since 1939 and persisted circulating there (spp 0.93). Clade III including the clusters IIA and IIIb, originated in South Korea (spp 0.90). Cluster IIIa although persisting in South Korea also spread in Tunisia and sporadically in China, Hungary and Germany. Cluster IIIb persisted in South Korea but spread also in India, Tunisia and China (Figure 1).

Figure 2 showed SalV worldwide migration route confirming the phylogeographic analysis that involve American, African, Asiatic and European continents. In detail phylogeographic analysis showed a greater posterior probability for USA in respect to China (spp = 0.19 vs 0.08 respectively), confirming the likelihood of origin for SalV 1 and 2 in USA and thereafter the posterior probability of 0.80 in the clade dated 1981 suggest the migration in China as probable origin for SalV 2. All the connections are statistically significant (red lines). The selective pressure analysis showed two sites under positive of selection, while 63% of sites seemed to be under negative selection, suggesting that SalV virus could be highly conserved.

Homology modelling

A multiple sequence alignment of fragments of the VP1 protein of SalV protein has been obtained. According to the alignment, the SalV VP1 fragments here considered cover the region 59-278 of the full-length Aichi Virus VP3 protein (PDB code 5AOO,

chain C). This portion is part of the VP3 hetero 15-mer and contains 3 α-helices and 10 β-sheets.

Among the available structural templates, the chain C of the PDB structure with code 5AOO corresponding to the crystal structure of the VP3 domain from Aichi Virus, has been selected. The decision has been taken in consideration of the suitable quality of the structure in terms of resolution (2.10 Å), the global and local quality estimate for the high sequence identity to the targets (about 53%) (Figure 3).

The sites subject to selective pressure in SalV VP1 fragments have been located onto the predicted three-dimensional structure. The two sites are at the positions 100th and 116th of the alignment corresponding to positions 42 and 58 of 5AOO chain C structure. The mutation at position 116 is exposed on the outer surface of the protein and located on the capsid surface. The mutation at position 100 is located in the inner surface of the protein, pointing toward the inner volume of capsid (Figure 4).

The homology model of the SalV VP1 protein compared with the VP3 fragment of Aichi Virus shows a difference in length and orientation of β-sheets encompassed by sequence positions 172 – 178 and 234 – 243. These secondary structure elements are predicted to interact, as in the Aichi Virus capsid, with an equivalent symmetry-related β-hairpin of an adjacent VP1 subunit forming an extended four-strand sheet.

ConSurf analysis conducted on the SalV VP1 domain suggests that both the position 100 and 116 are within variable sites of the VP3 protein of Aichi Virus. The variability may reflect the potentiality of future mutation that can induce more stability of the protein

Figure 3. Sequence alignment among two SaliV VP1 variants and the VP3 of Aichi Virus (chain C of the PDB coordinate set 5AOO). Residue colouring follows ClustalX scheme. Triangles below alignment mark the positions under selective pressure.
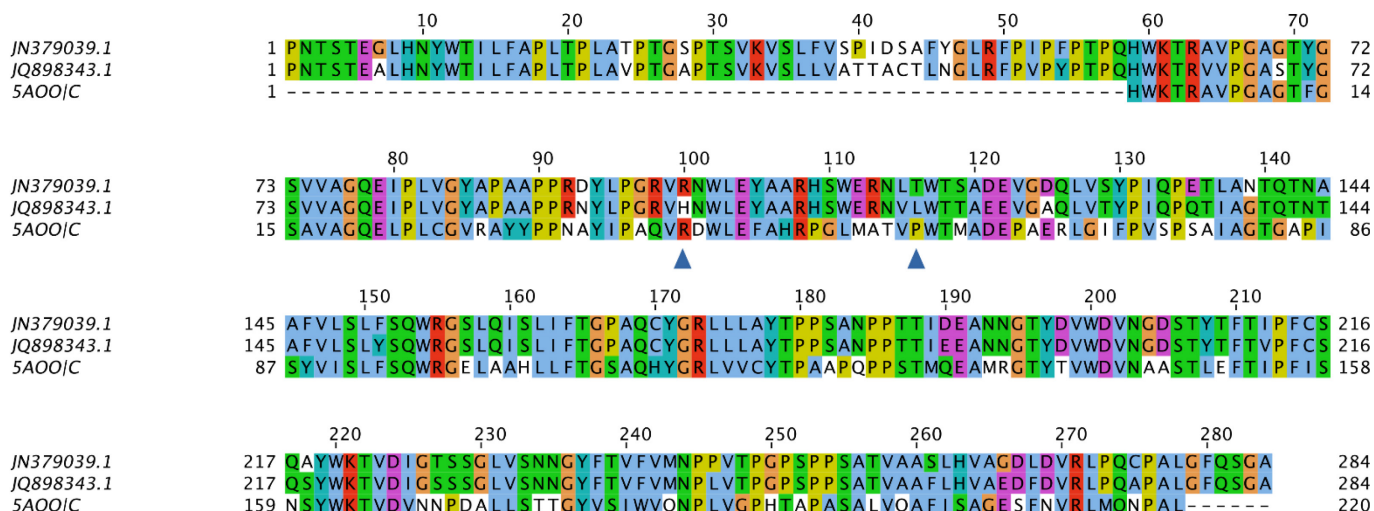
responsible for the clinical features of SalV that presented these mutations (Figure 5).

Prediction of thermodynamic impact of the two mutations on the SalV VP1 has been carried out for the homology models built on the templates JQ898343 and JN379039. In both cases, the mutations His48 to Arg and Leu58 to Thr are predicted to be moderately destabilizing while the inverse mutations obviously stabilizing (data not shown).

Discussion

SalV, a worldwide spread virus recognized as causative of gastroenteritis in children, is becoming an emerging serious concern. Previous study has proven that Salivirus RNA can be detected in both treated and untreated sewage water. This could be due to the fact that the treatment processes of sewage water are not enough efficient in order to remove SalV particles, probably because of possible virus crystalline formation or aggregated forms, having an effect to increase viral survival in environment water samples [42].

The causative role of this virus in acute gastroenteritis is discordant at today as so as the seasonality not well established.

The treatment-resistance, the wide diffusion, the misdiagnosis and the evolving way of diffusion of SalV makes it a possible public health threat in the future years. Few studies have investigated SalV in terms of evolutionary analysis and only basic phylogenetic trees to characterize this virus have been described. In this study, we applied the Bayesian phylogenetic and evolutionary analysis determining a probable date of insurgence of the first SalV in USA, MRCA to the year 1788 (Figure 1), after that this virus divided in two different genotypes, SalV A2, the most recent, dating back to 1982 and the most ancient Salv A1 dated in 1929. The virus was the cause of different monophyletic epidemic events forming almost three different clades and two different sub-clusters dated between the year 1975 and 2003.

The phylogeographic analysis (Figure 2, Figure 3) showed the probable origin of this virus is USA. From USA, the virus had two different way of spread, SalV A2 strains travelled in China causing various epidemics and after spread in Thailand and Venezuela by different way and date of transmission. The SalV A1, the most representative strain, travelled from USA to India in 1939 and successively in three different clades, spread in Nigeria (clade I) in 1977, persisted in India (clade II), and spread in South Korea (clade III) in 2000, identified by three different monophyletic epidemics. We did not identify any sort of intermingled strains in each clade or

Figure 4. Cartoon model of the structural superposition between the homology model and the VP3 protein of Aichi virus (PDB code 5AOO). Positions under positive selection have been indicated by arrow.
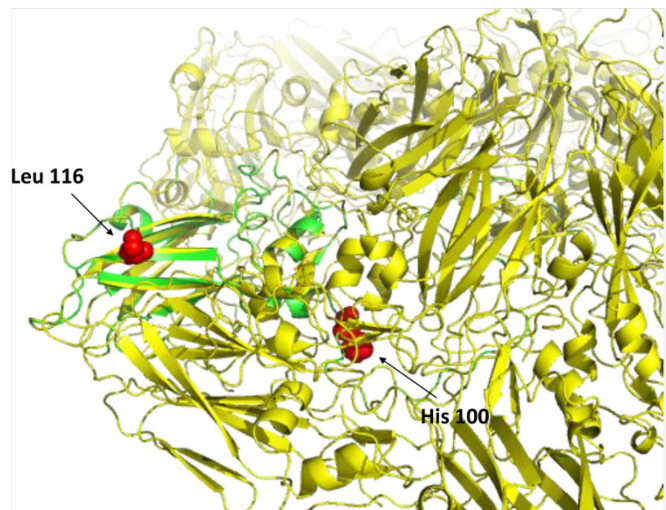

Figure 5. Cartoon depiction of the model based on the 5AOO.2.L structure colored according to the conservation score calculated by ConSurf web tool. The color scale ranges from cyan (high sequence variability) to dark red (high sequence conservation). Positions subject to selective pressure are represented as stick models and labelled accordingly. All of them occur in highly variable positions of the VP1 of SalV.
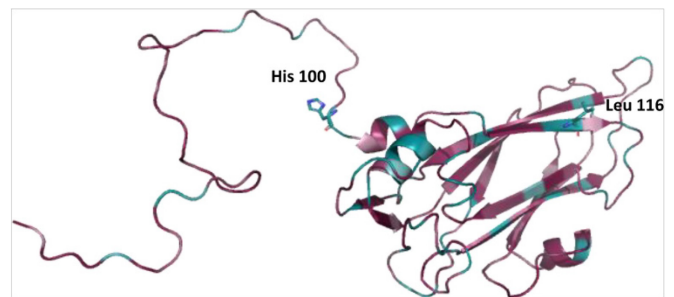
cluster underling the patent of territorial epidemics. This virus probably travelled by infected individuals that, once arrived in the living country, infected and contaminated environmental water becoming the potential source of human infection and creating a sort of fecal-oral circuit implying also the waste water as further source of virus isolation.

In this study, only two sites under positive of selection were found, suggesting that SalV virus could be highly conserved. This analysis do not have to encourage a light surveillance because previous studies, on other viruses, demonstrated that even only one site under positive selection can be important for

transmission vector change, as happened in CHKV epidemics [43,44]. Moreover, we did not exclude the amount of infectivity or therapy resistant or failed therapy so as. At the 116th position, the presence of Leucine instead of Threonine has been found from SalV A1 sequences, indicating that SalV with Histidine in the 100th position could induce over time severe and resistant infection.

We need to get a better understanding of SalV impact on public health in order to prevent and resolve future epidemics and infection [45], so as to prevent the diffusion of this evolving virus, how has been recently demonstrated [11,46,47].

The presence of the positive selection sites where the probability to fix a mutation over the time is high (> than 90%), can probably favor the resistance of the virus to the sanitation processes and the stagnation in sewage with other similar viruses. An example of the evolving nature of SalV is the recent isolation of the virus in respiratory specimen of child in China.

The presence of mutation in sequences of SalV that are sanitation resistant and more virulent and the homology between VP1 of SalV and VP3 protein of Aichivirus, suggests that probably this mutation facilitates the viral entry into the host cell. The presence of His at the 100th position suggests that this loop region can be characterized by a narrower angle compared with the sequences with Arg in the same position. This effect could probably lead to a more compact structure as predicted by SDM server.

The sequence conservation and mapping analysis has shown the presence of multiple regions with high variability, this could lead to the development of new mutations that could modify the clinical features of this evolving virus. Since SalV is already wide spreading, more attention should be paid to the diagnosis, prevention and treatment of infections.

This emphasized the need to improve the molecular epidemiological surveillance of SalV.

Conclusions

In light of these considerations, molecular and evolutionary studies are important to reveal the potential abilities of SalV to trigger a significant epidemic in countries also far each other's. For newly discovered and continuously evolving viruses like SalV so as for the most recent increasing number of epidemics and the presence of a positive selection sites, the dynamic nature of SalV could represent a serious threat for public health in the future.

Bayesian phylogenetic and phylodynamic represent useful tools to follow the dynamic transmission of this virus and to prevent the possibility of new epidemics through the world.

References

1. Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, Lawn JE, Cousens S, Mathers C, Black RE (2016) Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the sustainable development goals. Lancet 388: 3027-3035.
2. GBD Diarrhoeal Diseases Collaborators (2017) Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study, 2015. Lancet Infect Dis 17: 909-948.
3. Child Health Epidemiology Reference Group of WHO and UNICEF (2010) Global, regional, and national causes of child mortality in 2008: a systematic analysis. Lancet 375: 1969-1987.
4. Greninger AL, Runckel C, Chiu CY, Haggerty T, Parsonnet J, Ganem D, De Risi JL (2009) The complete genome of klassevirus - a novel picornavirus in pediatric stool. Virol J 6: 82.
5. Li L, Victoria J, Kapoor A, Blinkova O, Wang C, Babrzadeh F, Mason CJ, Pandey P, Triki H, Bahri O, Oderinde BS, Baba MM, Bukbuk DN, Besser JM, Bartkus JM, Delwart EL (2009) A novel picornavirus associated with gastroenteritis. Virol J 83: 12002-12006.
6. Holtz LR, Finkbeiner SR, Zhao G, Kirkwood CD, Girones R, Pipas JM, Wang D (2009) Klassevirus 1, a previously undescribed member of the family Picornaviridae, is globally widespread. Virol J 6: 86.
7. Lasure N, Gopalkrishna V (2016) Clinico-epidemiology and genetic diversity of Salivirus in acute gastroenteritis cases from Pune, Western India: 2007-2011. Infect Genet Evol 44: 425-430.
8. Lasure N, Gopalkrishna V (2017) Epidemiological profile and genetic diversity of sapoviruses (SaVs) identified in children suffering from acute gastroenteritis in Pune, Maharashtra, Western India, 2007-2011. Epidemiol Infect 145: 106-114.
9. Greninger AL, Holtz L, Kang G, Ganem D, Wang D, De Risi JL (2010) Serological evidence of human klassevirus infection. Clin Vaccine Immunol 17: 1584-1588.
10. Han TH, Kim CH, Chung JY, Park SH, Hwang ES (2010) Klassevirus infection in children, South Korea. Emerg Infect Dis 16:1623-5.
11. Aldabbagh S, Eckerle I, Muller A, Delwart EL, Eis-Hubinger AM (2015) Salivirus type 1 and type 2 in patients with acute gastroenteritis, Germany. J Clin Virol 72: 16-19.
12. Yang J, Yang F, Ren L, Xiong Z, Wu Z, Dong J, Sun L, Zhang T, Hu Y, Du J, Wang J, Jin Q (2011) Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. J Clin Microbiol 49: 3463-3469.
13. Lysholm F, Wetterbom A, Lindau C, Darban H, Bjerkner A, Fahlander K, Lindberg AM, Persson B, Allander T, Andersson B (2012) Characterization of the viral microbiome in patients with severe lower respiratory tract infections, using metagenomic sequencing. PLoS One 7: e30875.
14. Bibby K (2013) Metagenomic identification of viral pathogens. Trends Biotechnol 31: 275-279.
15. Ng TF, Marine R, Wang C, Simmonds P, Kapusinszky B, Bodhidatta L, Oderinde BS, Wommack KE, Delwart E (2012) High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. Virol J 86: 12161-1275.

16. Shan T, Wang C, Cui L, Yu Y, Delwart E, Zhao W, Zhu C, Lan D, Dai X, Hua X (2010) Picornavirus salivirus/klassevirus in children with diarrhea. China. Emerg Infect Dis 16: 1303-1305.

17. Pei N, Ma J, Li L, Ji J, Li J, Deng J, Liu X (2016) Complete Genome Sequence of a Salivirus in Respiratory Specimens from a Child with Adenovirus Infection. Genome Announc 4: e00159-16.

18. Hjelmsø MH Mollerup S, Jensen RH, Pietroni C, Lukjancenko O, Schultz AC, Aarestrup FM, Hansen AJ (2019) Metagenomic analysis of viruses in toilet waste from long distance flights-A new procedure for global infectious disease surveillance. PLoS One 14: e0210368.

19. Han TH, Kim SC, Kim ST, Chung CH Chung JY (2014) Detection of norovirus genogroup IV, klassevirus, and pepper mild mottle virus in sewage samples in South Korea. Arch Virol 159: 457-463.

20. Itta KC, Patil T, Kalal S, Ghargi KV, Roy S (2016) Salivirus in children with diarrhoea, western India. Int J Infect Dis 52: 14-15.

21. Hall TA (1999) BioEditA user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41: 95-98.

22. Bruen TC, Philippe H, Bryant D (2006) A simple and robust statistical test for detecting the presence of recombination. Genetics 172: 2665-2681.

23. Xia X (2018) DAMBE7: New and improved tools for data analysis in molecular biology and evolution. Mol Biol Evol 35: 1550-1552.

24. Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18: 502-504.

25. Darriba D, Taboada GL, Doallo R, Posada D (2012) j Model Test 2: more models, new heuristics and parallel computing, Nat Methods 9: 772.

26. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K (2013) FUBAR: a fast, unconstrained bayesian approximation for inferring selection. Mol Biol Evol 30: 1196-1205.

27. Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL (2015) Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol Biol Evol 32: 1342-1353.

28. Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A (2015) Gene wide identification of episodic selection. Mol Biol Evol 32: 1365–1371.

29. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL (2012) Detecting individual sites subject to episodic diversifying selection. PLoS Genet 8: e1002764.

30. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol 35: 1547-1549.

31. Rambaut A, Lam TT, Max Carvalho L, Pybus OG (2016) Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evolut 2: vew007.

32. Villano U, Lo Presti A, Equestre M, Cella E, Pisani G, Giovanetti M, Bruni R (2015) Molecular epidemiology and phylogenetic analysis of Hepatitis B virus in a group of migrants in Italy. BMC Infect Dis 15: 531.

33. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol 67: 901-904.

34. Bielejec F, Rambaut A, Suchard MA, Lemey P (2011) SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics 27: 2910-2912.

35. Waterhouse A, Bertoni M, Bienert S (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46: W296-W303.

36. Zimmermann L, Stephens A, Nam SZ (2018) A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J Mol Biol 430: 2237-2243.

37. Benkert P, Künzli M, Schwede T (2009) QMEAN server for protein model quality estimation. Nucleic Acids Res. Available: https://swissmodel.expasy.org/qmean/. Accessed 22 October 2019.

38. Ashkenazy H, Abadi S, Martz E (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res 44: W344-350.

39. Pandurangan AP, Ochoa-Montaño B, Ascher DB, Blundell TL (2017) SDM: a server for predicting effects of mutations on protein stability. Nucleic Acids Res 5: W229-W235.

40. Waterhouse AM, Procter JB, Martin DMA (2009) Jalview Version 2-A multiple sequence alignment editor and analysis workbench. Bioinformatics 25: 1189-1191,

41. Xia X, Xie Z, Salemi M, Chen L, Wang Y (2003) An index of substitution saturation and its application. Mol Phylogenet Evol 26: 1-7.

42. Adineh M, Ghaderi M, Mousavi-Nasab SD (2019) Occurrence of Salivirus in Sewage and River Water Samples in Karaj, Iran. Food Environ Virol 11: 193-197.

43. Lo Presti A, Cella E, Angeletti S, Ciccozzi M (2016) Molecular epidemiology, evolution and phylogeny of Chikungunya virus: an updating review. Infect Genet Evol 41: 270-278.

44. Benvenuto D, Cella E, Fogolari M, De Florio L, Borsetti A, Donati D, Garilli F, Spoto S, Ceccarelli G, Angeletti S, Ciccozzi M (2019) The transmission dynamic of Madariaga Virus by bayesian phylogenetic analysis: Molecular surveillance of an emergent pathogen. Microb Pathogen 132: 80-86.

45. Kumthip K, Khamrin P, Yodmeeklin A, Maneekarn N (2017) Salivirus infection in children with diarrhea, Thailand Arch Virol 162: 2839-2841.

46. Bergallo M, Daprà V, Rassu M, Bonamin S, Cuccu R, Calvi C, Montanari P, Galliano I (2018) Prevalence and clinical profile of human Salivirus in children with acute gastroenteritis in Northern Italy, 2014-2015. Intervirology 61: 49-52.

47. Nielsen AC, Gyhrs ML, Nielsen LP, Pedersen C. Bottiger B (2013) Gastroenteritis and the novel picornaviruses aichi virus, cosavirus, saffold virus, and salivirus in young children. J Clin Virol 57: 239-242.

Corresponding author

Prof. Silvia Angeletti
Unit of Clinical Laboratory Science
University Campus Bio-Medico of Rome
Via Álvaro del Portillo, 21
00128 Rome, Italy
Tel: +3906225411461
Fax: +390622541456
E-mail: s.angeletti@unicampus.it

Conflict of interests: No conflict of interests is declared.

Annex – Supplementary Items

Supplementary Table 1. The Bayes factor (BF) results upon model comparison.

Model Comparison	BF	Evidence against H0
Const Strict (H0) vs. Relaxed (H1) clock	29.89	very strong
Const Strict (H0) vs. Expo Relaxed (H1) clock	6.42	strong
Const Strict (H0) vs. Expo Strict (H1) clock	5.15	weak
Const Strict (H0) vs. GMRF strict (H1) clock	22.24	very strong
Const Strict (H0) vs. SkyGrid strict (H1) clock	180.23	very strong

Cost: constant population size; Expo: exponential population growth; GMRF: Gaussian Markov random field; Strict:strict molecular clock; Relaxed: relaxed molecular clock. Bayes Factor (BF) is the difference (in loge space) of the marginal likelihood of null (Ho) and alternative (H) model. BFs were estimated by comparing the approximate marginal likelihoods of different models given in Supplementary Table 1. Evidence against Ho is assessed in the following way: 2 > BF > 6 indicates positive evidence against the null model; 6 > BF > 10 indicates strong evidence against the null model; BF 10 indicates very strong evidence against the null model.