Infection, Genetics and Evolution 59 (2018) 23–27

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Opinion

Phylodynamics of Crimean Congo hemorrhagic fever virus in South Russia

A.N. Lukashev[a,c,*], A.A. Deviatkin[b]

[a] Martsinovsky Institute of Medical Parasitology, Tropical and Vector Borne Diseases, Sechenov University, Moscow, Russia
[b] Institute of Molecular Medicine, Sechenov University, Moscow, Russia
[c] Chumakov Federal Scientific Center for Research and Development of Immune-and-Biological Preparations, Moscow, Russia

ARTICLE INFO

Keywords:
CCHFV
Phylodynamics
Reassortment

ABSTRACT

Phylodynamics of Crimean Congo Hemorrhagic fever virus (CCHFV) genotype V in South Russia was analyzed using 244 partial (452–571 nt) sequences in all three genomic segments and 38 complete genomic sequences. Despite increased number of sequences, the Russian lineage of the European genotype V (commonly termed GtVa) was distinct from GtV isolates from Turkey and the Balkan countries. No geographic pattern was observed in phylogenetic subgrouping of CCHFV within South Russia. Identical isolates could be found at distant locations spaced by hundreds of kilometers, while relatively divergent viruses circulated in the same region. Full genome analysis indicated that reassortment events within GtVa occurred every few decades (median half-life of a non-reassortant node 30–40 years) and involved M and S segments. Therefore, in South Russia CCHFV represents a highly dynamic population of frequently reassorting viruses.

1. Introduction

Crimean Congo hemorrhagic fever (CCHF) is a severe tick-borne zoonosis. Fatality rates are usually around 1–5%, but can reach 30% upon human-to-human transmission (Bente et al., 2013). The virus is a significant health concern in Africa, Asia and Southern Europe. The CCHF virus (CCHFV) is a member of the *Bunyavirales* order, the *Nairoviridae* family and the *Orthonairovirus* genus. It is an enveloped virus with three RNA segments: S (small), around 1.7 Kb, which encodes nucleoprotein; M (medium), around 4.5 Kb, which encodes two glycoproteins and a mucin-like domain; and L (large), 12.1 Kb, which encodes polymerase. The S segment is predominantly used for phylogenetic studies. Analysis of CCHFV strains collected worldwide shows a clear geographic pattern of CCHFV genotypes (Deyde et al., 2006; Hewson et al., 2004a). Reassortment of CCHFV genome segments is well documented (Hewson et al., 2004b) and was estimated to occur approximately once every 100 years (Lukashev et al., 2016). As a result, the phylogenetic relations of viruses and the genotype systems may differ between segments. Virus lineage that is predominant in Southern Europe (South Russia, Turkey, Bulgaria, Albania) was termed genotype V (GtV) in all genome segments (Deyde et al., 2006). There is evidence of reassortment within GtV, but not between GtV and other genotypes (Lukashev et al., 2016). Rarely, divergent isolates of GtVI were also found in Europe (Sherifi et al., 2014).

Phylogeographic studies of the "European" GtV indicate higher virus diversity in Turkey and South Russia than in Balkan countries, and assume westward virus spread, which is also consistent with epidemiological data. Bayesian coalescent phylogenetic analysis of partial S segment sequences estimated that the most recent common ancestor of GtV existed 82 (95% high probability density interval: 52–112) years ago, while a study of full-genome sequences yielded more ancient dates, ranging from 248 (126–400) to 402(289–538) years ago in different genome segments (Lukashev et al., 2016) or 257–517 years ago (Carroll et al., 2010). The likely location of GtV emergence was inferred to be South Russia (Lukashev et al., 2016; Zehender et al., 2013). A further subdivision of GtV into subtypes Va (South Russia) and Vb-Ve (Turkey and the Balkans) was suggested (Zehender et al., 2013).

Complete CCHFV genomes provide good resolution of phylogenetic methods (robust statistical support for most tree nodes), but only about 80 are available worldwide. Short genome fragments may lack resolution on a short timescale, because the virus accumulates mutations at a rate of approximately 1×10^{-4}–3×10^{-4} substitutions/site/year (Carroll et al., 2010; Lukashev, 2005; Sherifi et al., 2014), and thus about one mutation per 100 nt in 30–100 years.

In 2016, 21 complete genomes (KR814833-KR814893, KU161582-KU161587) and partial sequences in all three genome segments of over 440 CCHFV samples from South Russia (S segment: KR814894-KR815339, positions 115–652 according to AF428144; M segment: KU161576-KU161581, positions 4640–5074 according to KR814873 and L segment: positions 110–546 according to KX013468) were produced by the Stavropol Antiplague Scientific Research Institute of the State Sanitary Surveillance of Russia. Sequences KU161582–KU161587

* Corresponding author at: Martsinovsky Institute of Medical Parasitology, Tropical and Vector Borne Diseases, Malaya Pirogovskaya 20-1, Moscow 119435, Russia.
E-mail address: alexander_lukashev@hotmail.com (A.N. Lukashev).

https://doi.org/10.1016/j.meegid.2018.01.016
Received 1 November 2017; Received in revised form 19 January 2018; Accepted 21 January 2018

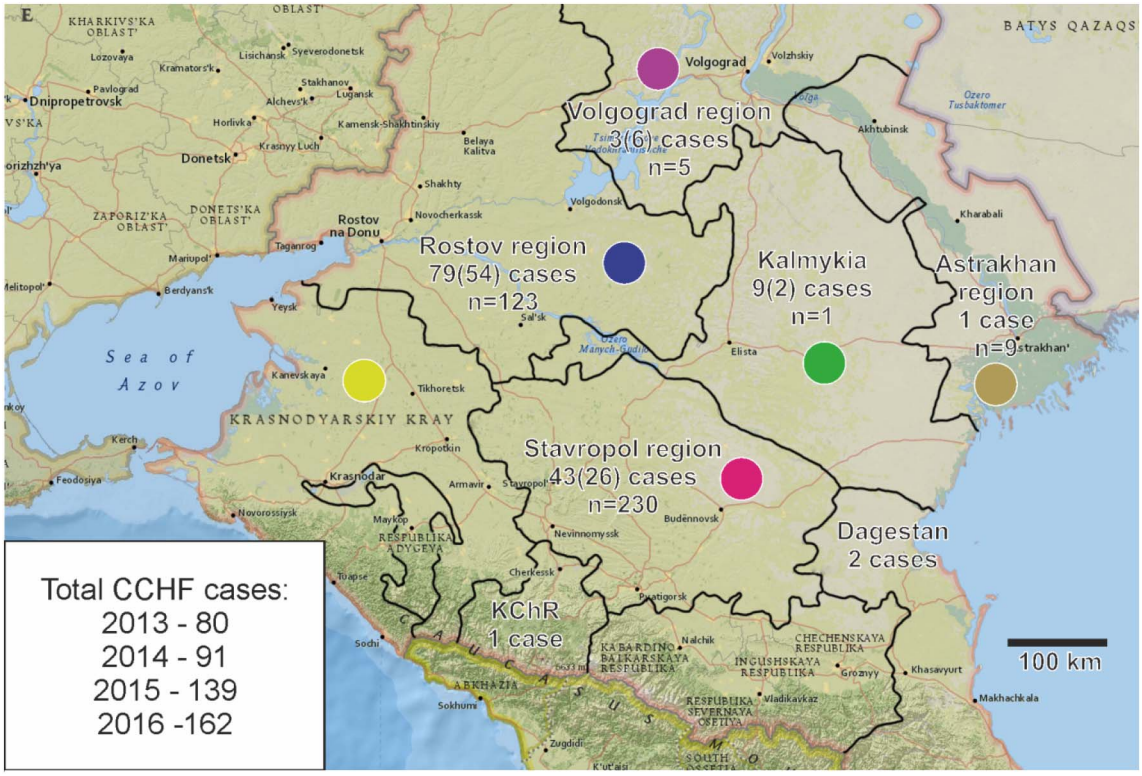

Fig. 1. Incidence of CCHF in 2015 (2014) and the number of unique sequences available in Genbank from the endemic regions of South Russia. Total incidence was taken from annual reports on infectious disease incidence by the Russian Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing - Rospotrebnadzor - (Rospotrebnadzor, 2016). Dots correspond to branch colors on the phylogenetic tree (Fig. 2a). The map was obtained using ArcGIS software [Esri Inc., version 10.4.1, Redlands, California, USA].

were analysed previously (Kulichenko et al., 2016). These new sequences allowed for the investigation of CCHFV phylodynamics with an unprecedented resolution.

CCHF is endemic in a large part of South Russia between the Black sea on the West and the Caspian sea on the East (about 700 km), and between Caucasian mountains on the South and Volgograd region on the North (about 1000 km). The annual number of reported cases varied between 80 and 162 in 2013-2016 (Fig. 1). Most cases are sporadic; however outbreaks involving up to 40 epidemiologically related cases have also occurred (Onishchenko et al., 2001).

2. Materials and methods

As of December 2016, there were 487 partial S segment sequences from South Russia available in Genbank. To simplify analysis and data presentation, redundant entries (sequences of the same isolates or 100% identical sequences from the same region) were omitted. Importantly, sequences of several isolates were published independently by different laboratories and were identical, indicating high data quality. The sequences (genome positions 150–652 according to AF428144) were aligned with homologous Genbank reference sequences of GtV from other countries, also omitting identical sequences. The final data set contained 244 sequences (Online resource 1).

For reassortment analysis, all available complete genome sequences of viruses isolated in Russia were used (as of December 2016). Samples without known collection dates were omitted. Strains Kelkit06 (Turkey) and Hoti (Kosovo) were added to the data set to root the trees.

The multiple sequence alignment was performed using MAFFT server (Kuraku et al., 2013). Only protein-coding regions were used for the analysis. The final data set consisted of 38 sequences.

A reversible jump-based substitution model (Bouckaert et al., 2014) was used to choose the substitution model and estimate the appropriate number of parameters, while sampling the tree was used for Bayesian coalescent analysis. Next, different clock assumptions and population models were compared by a Bayes factors test. The highest Bayes factor was observed in the combination of the uncorrelated lognormal relaxed clock and the exponential population model in all segments, although there were no significant differences between the estimates for other combinations of two clock (strict and relaxed clock) and two population models (constant and exponential). Trees were sampled every 10,000 (5000; 3000) generations for 100 (50; 30) million generations in total and annotated with a burn-in of 10 (5; 3) million generations in L, M and S segments, respectively. The convergence of parameter estimates was checked using Tracer (v1.5) and indicated by an effective sample size > 200.

3. Results and discussion

As expected, phylogenetic analysis of 244 relatively short and very similar sequences could not produce a robust bootstrap support for most tree nodes. Nevertheless, several important conclusions could be drawn.

South Russian isolates were distinct from viruses found in Turkey and the Balkans. (Fig. 2).

Despite a hugely increased dataset, there was no evidence of virus transfers between South Russia and Turkey-Balkan region; however, viruses from both these areas were occasionally introduced to Iran (Fig. 2). Isolates from distinct regions of South Russia (sometimes hundreds of kilometers apart, Fig. 1) were intermixed. There were a number of occasions when identical isolates could be obtained in distinct (and often non-adjacent) regions. The most vivid examples of such virus distribution, circled in Fig. 2, included five or six identical strains from the Astrakhan, Rostov, Stavropol, Krasnodar, Kalmykia and Volgograd regions. Similar conclusions could be drawn regarding M and L segment analysis; however, as fewer sequences of these regions were available from Turkey and Southern Europe, this data is not shown.