Infection, Genetics and Evolution 32 (2015) 255–264

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Time-dependent selection pressure on two arthropod-borne RNA viruses in the same serogroup

Abinash Padhi *, Li Ma *

Department of Animal and Avian Sciences, University of Maryland, College Park, MD 20742, USA

ARTICLE INFO

Article history:
Received 16 January 2015
Received in revised form 11 March 2015
Accepted 15 March 2015
Available online 20 March 2015

Keywords:
Orthobunyavirus
Ruminants
Positive selection
Purifying selection
Mutation rate
Sequence divergence

ABSTRACT

Understanding the genetic basis of viral adaptation to taxonomically diverse groups of host species inhabiting different eco-climatic zones is crucial for the discovery of factors underpinning the successful establishment of these infectious pathogens in new hosts/environments. To gain insights into the dynamics of nonsynonymous (*dN*) and synonymous substitutions (*dS*) and the ratio between the two ($\omega = dN/dS$), we analyzed the complete nucleotide coding sequence data of the M segment, which encodes glycoproteins of two negative-sense RNA viruses, Akabane virus (AKV) and Schmallenberg virus (SBV) that belong to the same serogroup. While AKV is relatively older and has been circulating in ruminant populations since 1970s, SBV was first reported in 2011. The ω was estimated to be 1.67 and 0.09 for SBV and AKV, respectively, and the estimated mutation rate of SBV is at least 25 times higher than that of AKV. Given the different evolutionary stages of the two viruses, most of the slightly deleterious mutations were likely purged out or kept in low frequency in the AKV genome, whereas positive selection together with the accumulation of slightly deleterious mutations might contribute to such an inflated mutation rate of SBV. The evolutionary distance (*d*) is nonlinearly and negatively correlated with ω, but is positively correlated with *dN* and *dS*. Collectively, the different patterns in ω, *dN*, *dS*, and *d* between AKV and SBV identified in this study provide empirical evidence for a time-dependent selection pressure.

1. Introduction

Understanding the genetic basis of viral adaptation to the phylogenetically diverse range of host species inhabiting different eco-climatic zones is crucial to uncover the factors underpinning the cross-species transmission and successful establishment of these infectious pathogens in new hosts/environments. Subsequently, this knowledge would help predict possible future outbreaks. One of the most popular approaches to inferring the genetic basis of adaptation of all living forms, including infectious viral pathogens, is based on the estimation of the ratio of *dN* (the number of nonsynonymous changes per nonsynonymous site) to *dS* (the number of synonymous changes per synonymous site) in protein-coding genes (Nielsen, 2005; Yang and Bielawski, 2000). When the *dN*/*dS* ratio (ω) is less than one, the protein coding gene is said to be under purifying selection and therefore functionally constrained. Alternatively, when ω is greater than one, selection is said to favor adaptive changes that alter amino acids at the specific sites/domains in the protein coding gene, which is referred to as positive selection (Nielsen, 2005; Yang and Bielawski, 2000). According to the neutral theory of molecular evolution, a vast majority of the nucleotide substitutions in protein-coding genes in all biological entities are likely to have been subjected to purifying selection (Kimura, 1977, 1983). Although positive selection is less frequent than purifying selection, one of the fundamental goals in evolutionary genetic studies is to detect the genomic regions/genes/specific codons within a gene that have evolved under positive selection and to understand the factors that have caused such adaptive evolutionary changes. While the ω approach has been widely used to quantify the mode and strength of selection pressures at the population level, among the closely related lineages/species as well as among the distantly related lineages/species, recent studies have reported limitations of this approach, notably, in inferring the selection pressures at the population level as well as among closely related lineages/species (Kryazhimskiy and Plotkin, 2008; Mugal et al., 2014; Peterson and Masel, 2009; Rocha et al., 2006; Wolf et al., 2009). The main contention was that since the *dN*/*dS* method, which was originally developed in the phylogenetic context to infer selection pressures based on coding nucleotide sequences of distantly related lineages, presumably estimates the ratio of fixed nonsynonymous to synonymous differences between lineages (Kryazhimskiy and Plotkin, 2008; Mugal et al., 2014; Peterson and Masel, 2009), application of this

* Tel.: +1 301 405 1389; fax: +1 301 405 7980 (L. Ma).
E-mail addresses: abinash74@gmail.com (A. Padhi), lima@umd.edu (L. Ma).

http://dx.doi.org/10.1016/j.meegid.2015.03.019
1567-1348/

approach to recently emerged populations (Kryazhimskiy and Plotkin, 2008) or relatively younger lineages/recently diverged species (Mugal et al., 2014; Peterson and Masel, 2009; Rocha et al., 2006; Wolf et al., 2009) may yield biased results.

To gain insights into the dynamics of *dN* and *dS* and to infer the mode and strength of selection pressure, in this study, we analyzed coding sequence data of two negative-sense ssRNA viruses, which belong to the same serogroup and genus but were reported to have emerged at different time points. Akabane virus (AKV) and Schmallenberg virus (SBV) are the arthropod-borne orthobunyavirus (family Bunyaviridae) of the Simbu serogroup (Hoffmann et al., 2012), which primarily affect ruminants and are associated with abortion, stillbirth, premature births and congenital abnormalities (Doceul et al., 2013; Fischer et al., 2013; Hoffmann et al., 2012; Kono et al., 2008; Yamakawa et al., 2006). Although both viruses have similar etiological features and host range, they exhibit distinct geographical distributions and emerge at different time points. While AKV is relatively older and has been causing sporadic outbreaks in the cattle industries across Asia, Australia, and Africa since 1970s (Geoghegan et al., 2014; Jun et al., 2012; Kessell et al., 2011; Kono et al., 2008; Lee et al., 2007; Oem et al., 2014, 2012; Yamakawa et al., 2006), SBV was first reported in autumn 2011 in Germany (Hoffmann et al., 2012) and subsequently, spread across most of the European countries (Beer et al., 2013; Chaintoutis et al., 2014; Doceul et al., 2013; Dominguez et al., 2014; Herder et al., 2012; van den Brom et al., 2012; Yilmaz et al., 2014). The two viruses have similar genomic features with three genomic segments, small (S), medium (M) and large (L) (Fischer et al., 2013). While the S segment encodes a small non-structural protein and the nucleocapsid protein, the L segment encodes the RNA-dependent RNA polymerase (Fischer et al., 2013). The M segment, which exhibits the greatest sequence variability (Fischer et al., 2013; Kobayashi et al., 2007), encodes the glycoproteins that play a crucial role in viral attachment and cell fusion (Fischer et al., 2013). The M segment is characterized by the presence of two surface glycoproteins (Gn and Gc) and a non-structural protein (NSm) in the order "Gn-NSm-Gc" (Fischer et al., 2013). Given the essential role of the M segment, the glycoprotein regions of both viruses are expected to evolve under positive selection. Taking the serially sampled viral M segment sequence data into account and employing the maximum likelihood (ML)-based codon substitution models, in this study, we seek to assess the mode and strength of selection pressures on these two viruses that emerged at different time points, specifically to determine whether the selection pressures are time-dependent, and ultimately to know how the variable genetic distances affect the *dN*, *dS* and ω in these two arthropod-borne viruses.

2. Materials and methods

2.1. Sequence data

To determine the phylogenetic placements of the SBV and AKV, a total of 43 complete nucleotide coding sequences of the nucleocapsid (N) gene, which is located in the S segment and relatively highly conserved (Fischer et al., 2013; Hoffmann et al., 2012), representing 23 orthobunyaviruses and three serogroups were retrieved from the GenBank (Benson et al., 2014). The M segment, which exhibits greater sequence variability (Fischer et al., 2013; Kobayashi et al., 2007) and encodes the glycoproteins that play a crucial role in viral attachment (Fischer et al., 2013), was analyzed to infer the phylogenetic relationships among the strains of each viral species, the rates of nucleotide substitutions, as well as to know whether this important protein in these two viruses (SBV and AKV) has evolved under differential selection pressures. We retrieved, respectively for AKV and SBV, 43 and 28 complete M segment, dated nucleotide sequence data from the GenBank (Benson et al., 2014). To know how selection pressures vary across the phylogeny of the simbu serogroup, complete M segment nucleotide coding sequences representing 12 orthobunyaviruses belonging to the simbu serogroup were also retrieved from GenBank. GenBank accession numbers of the nucleotide sequences used in the present study are listed in Appendix A. Codon-based sequence alignments were performed using the MUSCLE algorithm implemented in MEGA 5 (Tamura et al., 2011). Appropriate nucleotide substitution model for each dataset was selected using the Akaike Information Criterion (AIC) (Posada and Buckley, 2004) implemented in jModelTest ver.2 (Darriba et al., 2012). The best-fit model for each dataset is listed in Appendix B. To ensure that no recombinant sequences were included in the dataset, prior to further genetic analyses, a recombination detection program implemented in the RDP ver. 4 (Martin et al., 2010) was used for the detection of recombinant sequences.

2.2. Inter-species phylogeny

The ML-based phylogenetic tree from the complete N gene sequence data, which comprised 23 orthobunyaviruses viruses belonging to three different serogroups, was reconstructed using the PhyML ver 3 (Guindon and Gascuel, 2003). Using the same program bootstrap values were estimated with 1000 replicates. Phylogenetic trees were visualized using the FigTree v1.4.2 program (available at http://tree.bio.ed.ac.uk/software/figtree/).

2.3. Estimating the evolutionary rates of SBV and AKV

To assess the temporal structure in the M segment of SBV and AKV, we performed the root-to-tip genetic distance regression analyses implemented in Path-O-Gen ver. 1.4 (available at http://tree.bio.ed.ac.uk/software/pathogen/). The coefficient of correlation was determined by fitting a regression of the year-of-sampling against the root-to-tip genetic distance of each sample, measured from an ML tree. The ML tree for each virus was reconstructed using the phyML program (Guindon and Gascuel, 2003). To estimate the time to the most recent common ancestors (tMRCA) and nucleotide substitution rate, we used the dated M segment nucleotide sequences and employed a Bayesian Markov Chain Monte Carlos (MCMC) approach implemented in BEAST ver 1.80 (Drummond and Rambaut, 2007). The evolutionary rates and tMRCAs were estimated under the strict clock model, which assumes a consistent substitution rate, and a relaxed clock model (with uncorrelated log-normal distribution) under Bayesian coalescent prior (Drummond et al., 2006). The best-fit-clock model was evaluated by calculating a Bayes factor (BF), the difference between the marginal log likelihoods of two models (Suchard et al., 2001). When $2\ln(BF) > 3$, it is considered to be the indication of positive evidence for the alternative model (Gray et al., 2011). Under the appropriate nucleotide substitution models, phylogenies were evaluated using a MCMC chain length of 30 and 60 million states for SBV and AKV, respectively. Uncertainty in the data was described by the 95% highest posterior density (HPD) intervals. Multiple chains were run to assess the convergence of trees and were checked using Tracer ver. 1.5 (available at: http://beast.bio.ed.ac.uk/Tracer).

2.4. Tests for selection

2.4.1. Pairwise comparison

Pairwise *dN*, *dS* and ω among the viral strains within SBV and AKV were estimated from the complete M segment nucleotide coding sequences of the respective viruses using the method described by (Goldman and Yang, 1994) implemented in the CODEML