Infection, Genetics and Evolution 13 (2013) 198–210

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

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

Review

Evolutionary and ecological factors underlying the tempo and distribution of yellow fever virus activity

Christine V.F. Carrington [a,*], Albert J. Auguste [b]

[a] Biochemistry Unit, Department of Pre-Clinical Sciences, The University of the West Indies, St. Augustine, Trinidad and Tobago
[b] Institute for Human Infections and Immunity and Department of Pathology, University of Texas Medical Branch, Galveston, TX 77555, United States

ARTICLE INFO

Article history:
Received 29 June 2012
Received in revised form 15 August 2012
Accepted 16 August 2012
Available online 3 September 2012

Keywords:
Yellow fever virus
Phylogeny
Evolution
Phylogeography
Geographic distribution

ABSTRACT

Yellow fever virus (YFV) is historically one of the most important viruses to affect human populations. Despite the existence of highly effective vaccines for over 70 years, yellow fever remains a significant and re-emerging cause of morbidity and mortality in endemic and high-risk regions of South America and Africa. The virus may be maintained in sylvatic enzootic/epizootic, transitional and urban epidemic transmission cycles with geographic variation in terms of levels of genetic diversity, the nature of transmission cycles and patterns of outbreak activity. In this review we consider evolutionary and ecological factors underlying YFV emergence, maintenance and spread, geographic distribution and patterns of epizootic/epidemic activity.

1. Introduction

Yellow fever, an acute viral haemorrhagic disease, was the first human disease to be attributed to a virus, and the first demonstrated to be transmitted by an arthropod (Reed et al., 1901; Strode, 1951; Finlay, 2001). For centuries the disease has been a recognised public health problem in both Africa and the Americas. It is now considered to be historically one of the most important diseases of humans, most notable for its roles in the decimation of European colonist and their armies during campaigns in the Americas, and for thwarting early attempts to construct the Panama Canal (Tomori, 1999; Oldstone, 2010).

Although the earliest reports of yellow fever-like disease and the first recorded epidemics were both from the Americas (in the late 15th and mid-17th centuries respectively) (Scott, 1939), phylogenetic analyses indicate that the virus originated in Africa (Chang et al., 1995; Wang et al., 1996; Mutebi et al., 2001; Bryant et al., 2007), where outbreaks were associated with higher mortality rates among colonist encountering it for the first time than amongst the African population (Boyce, 1911; Carter et al., 1931; Oldstone, 2010). Likewise, on being introduced to the Americas, it took a significant toll on the indigenous populations (Oldstone, 2010).

At the beginning of the 20th century the viral aetiology of yellow fever was confirmed and the urban dwelling, anthropophilic *Aedes* (*Stegomyia*) *aegypti* mosquito was identified as the vector (Reed et al., 1901). The virus is an enveloped single-stranded positive-sense RNA virus belonging to the family *Flaviviridae* and is the prototype for the genus *Flavivirus*. A description of the yellow fever virus (YFV) genome and encoded proteins is given in Box 1.

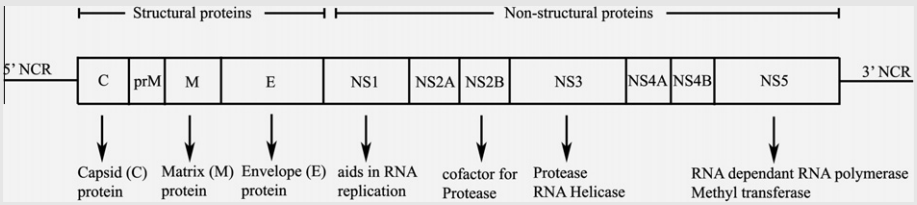
2. YFV ecology and disease epidemiology

In addition to the urban "man-mosquito-man" cycle mediated by *Ae. aegypti* (Reed et al., 1901; Strode, 1951) that results in large explosive epidemics, YFV is maintained in enzootic sylvatic or "jungle" transmission cycles involving non-human primates and arboreal mosquitoes (Shannon et al., 1938; Monath, 1989; Barrett and Monath, 2003) such as *Aedes* spp. mosquitoes in Africa (Beaty et al., 1980; Reiter et al., 1998), and *Haemogogus* and *Sabethes* spp. mosquitoes in South America (De Rodaniche and Galindo, 1957; De Rodaniche et al., 1957; Dutary and Leduc, 1981; Downs, 1982b; Rawlins et al., 1990; Vasconcelos et al., 1997, 2003; Mondet et al., 2002; Barrett and Monath, 2003; Cardoso Jda et al., 2010) (see Box 2). Additionally, an "intermediate" or "savannah" cycle is recognised in humid and semi-humid African savannahs, where human populations at the forest fringe interact with the sylvatic YFV cycle, resulting in small-scale rural epidemics (Smithburn et al., 1949; Monath, 1989; Barrett and Monath, 2003).

Despite the existence of a highly effective vaccine since 1937 (Theiler and Smith, 1937; Monath, 1989) and largely successful efforts to eradicate *Ae. aegypti* in the early to mid 20th century, failure to sustain vaccination and mosquito control programs has

* Corresponding author. Tel.: +1 868 662 1873; fax: +1 868 662 1873.
E-mail address: christine.carrington@sta.uwi.edu (C.V.F. Carrington).

1567-1348/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2012.08.015



Box 1. Yellow fever virus genome organization. The YFV genome is a single-strand positive-sense RNA of approximately 11,000 nucleotides. It consists of a single open reading frame (ORF) flanked by non-coding regions (NCR) with a type 1 cap (m^7 GpppAmp) at the 5′ terminus. The 5′ NCR is conserved (Barrett and Monath, 2003) but there is geographic variation in terms of the number of 41–44 nucleotide long repeats (RYF repeats (Hahn et al., 2000) in the 3′ NCR, with West African strains having three repeats (RFY1, RFY2 and RFY3), two (RFY1 and RFY3) in strains from East and Central Africa, and one (RFY 3) in South American strains (Chambers et al., 1990; Wang et al., 1996; Mutebi et al., 2004). The evidence suggests that the South American variant arose from a West African variant through loss of RFY 1 and 2 in a single deletion event (Wang et al., 1996; Mutebi et al., 2004). RFY 2 in West Africans strain may have arisen from a duplication event in an East/Central African progenitor or alternatively the East/Central African strains may be the result of an RFY2 deletion in a common ancestor shared with the West African strains (Mutebi et al., 2004). The structural proteins encoded towards the 5′ end of the ORF include the capsid (C) protein, the matrix (M) protein which is post-translationally cleaved from premembrane (prM) protein and the envelope (E) protein. The latter haemagglutinates red blood cells, mediates receptor binding and acid pH-dependent fusion activity, as well as inducing a neutralizing immune response. Although the exact functions of each of the seven non-structural proteins encoded at the 3′ end of ORF remain unclear, the NS1 protein is thought to aid in virus replication (Gritsun et al., 1988, 1989; Gritsun et al., 1990), signal transduction (Jacobs et al., 2000), and virion assembly or release (Lee et al., 1989), NS5 contains the RNA-dependant RNA polymerase (Tan et al., 1996) and the NS3 encodes a serine protease and a RNA Helicase (Lindenbach and Rice, 2001).

resulted in YFV re-emerging as an important global public health concern (Bryan et al., 2004), with an estimated 200,000 cases and 30,000 deaths globally each year in the tropics of Africa and South America (Vainio and Cutts, 1998). The majority of reported cases (~80% during the period 1980–2011) are from Africa (WHO, 2012b), where outbreaks in human populations occur annually, including some very large epidemics (Kirk, 1941; Sanders et al., 1998; Monath, 2006). In contrast, the last major urban epidemic in the Americas occurred in Brazil in 1928 (Soper, 1977). Since then there have been only a small number of sporadic urban cases (i.e. cases involving *Ae. aegypti* transmission following spillover of sylvatic YFV), reported in Brazil (three cases in 1942) (Figueiredo, 2000; Tauil, 2010), Trinidad (15 cases in 1954) (Downs et al., 1955), and Bolivia (6 cases in 1998) (Van der Stuyft et al., 1999), and in 2008 there were 24 confirmed cases in Paraguay including some in urban areas (Horwood, 2008; PAHO, 2008; WHO, 2008) There is however regular epizootic activity in South America, frequently resulting in large die offs among non-human primate populations and spillover into humans (in particular unvaccinated agricultural and forest workers such as loggers and hunters, who regularly come into contact with the sylvatic mosquitoes).

The global re-emergence of YFV is neither surprising nor unexpected given that the sylvatic transmission cycle is unaffected by control measures aimed at humans and *Ae. aegypti*. Also, most countries in endemic/high-risk areas currently have estimated vaccination coverage well below the 80% threshold required for effective prevention of an epidemic (WHO, 2011). For example, in 2010, only 6 out of 23 high-risk countries in Africa for which data were available and 2 out of 10 in South America were estimated to have achieved the minimum threshold, and global vaccination coverage for 2011 was only 43% (WHO, 2012a).

3. Evolution and phylogeny

The Old World origin of YFV is evident in phylogenies including other flaviviruses, where it groups together with other *Aedes* spp. associated viruses of Old World origin (Kuno et al., 1998; Gaunt et al., 2001; Jenkins et al., 2001; Cook and Holmes, 2006) (Fig. 1). Early genetic studies of YFV based on RNAase T1 oligonucleotide fingerprinting of sixteen YFV isolates from four African and three South American countries defined four geographically defined YFV topotypes (i.e. Senegal – Gambia, Central African Republic, Ivory Coast – Burkina Faso (Upper Volta) and the Americas) that were genetically stable over long periods (Deubel et al., 1985, 1986). This pattern of genetic stability and clustering according to region was supported by subsequent studies based on analysis of nucleotide sequences from the E gene (Lepiniec et al., 1994; Chang et al., 1995), and further confirmed by phylogenies based on a 1320 nucleotide fragment from the 5′ terminus, the region encoding NS4A, 2K and NS4B, the 3′NCR (Wang et al., 1996; Mutebi et al., 2004), the prM/M and E junction (Mutebi et al., 2001; Auguste et al., 2010) (see Fig. 2) and the complete open reading frame (Auguste et al., 2010).

Despite its historical importance and current status as a re-emerging disease, there are still relatively few YFV sequences available. At the time of writing there were only 482 YFV genomic sequences in Genbank of which only 29 were complete genomes. Consequently, the full extent of YFV genetic diversity and details of its evolution remain unclear, particularly since some geographic regions and individual countries are under-represented (e.g. East and Central Africa, and most countries in the Americas). Notwithstanding this, studies have shown that YFV phylogenies generated using sequence data from a variety of gene regions are very similar (Lepiniec et al., 1994; Chang et al., 1995; Wang et al., 1996; Mutebi et al., 2001, 2004; Bryant et al., 2007; Auguste et al., 2010), suggesting a constant evolutionary rate across the genome. Studies have also demonstrated that YFV evolves significantly slower than other flaviviruses (Jenkins et al., 2002; Bryant et al., 2007; Sall et al., 2010) and displays significant rate variation amongst lineages, so that it does not conform to a strict molecular clock (Sall et al., 2010). There is no evidence to support recombination as an important factor in YFV evolution in nature and *in vitro* studies suggests that it is very unlikely in flaviviruses (McGee et al., 2011).

The most striking feature of YFV phylogenies is the strong geographic structuring between and within continents (Fig. 2). The deepest phylogenetic split separates West African sequences from a clade containing more divergent East and Central African sequences (Chang et al., 1995; Wang et al., 1996; Mutebi et al., 2001). The pattern suggests that YFV arose in East or Central Africa (Mutebi et al., 2001) where three genotypes (namely Angola, East/Central Africa, and East Africa) have now been identified and later spread to West Africa where two genotypes have been noted (i.e. West Africa genotype I and II) (Mutebi et al., 2001). This evolutionary pattern also appears to be reflected in the number of repeat sequences (RYF repeats) in the 3′NCR of strains from different regions (Rice et al., 1985; Wang et al., 1996; Mutebi et al., 2004) (see Box 1).

Using the Bayesian coalescent approach employed by the software programme BEAST (Pybus and Rambaut, 2009; Drummond et al., 2012 for review), on a data set of 133 YFV prM/E sequences,