CAIO CESAR DE MELO FREIRE

**Caracterização de processos evolutivos de vírus de RNA a partir de padrões deixados nas
filogenias virais**

Tese de doutorado apresentada ao Programa
Interunidades de Pós-Graduação em Bioinformática
da Universidade de São Paulo para obtenção do título
Doutor.

**São Paulo
2014**

CAIO CESAR DE MELO FREIRE

**Caracterização de processos evolutivos de vírus de RNA a partir de padrões deixados nas
filogenias virais**

Tese de doutorado apresentada ao Programa Interunidades de Pós-Graduação em Bioinformática da Universidade de São Paulo para obtenção do título Doutor.

Área de concentração: Bioinformática.

Orientador: Prof. Dr. Paolo Marinho de Andrade Zanotto.

Versão original.

**São Paulo
2014**

Agradecimentos

À minha família, pais, irmãos, esposa e filha de quem eu me ausentei durante a realização deste trabalho, minhas sinceras desculpas.

Aos amigos que venho juntando desde a infância nos tempos de Colégio Salesiano em Natal, na juventude durante a graduação na UFRN e nos percursos da pós-graduação aqui em São Paulo, onde tive o prazer de morar em um lugar chamado Belugas.

Ao Prof. Paolo Zanutto que me acolheu no seu laboratório e me ensinou muito sobre ciência e também sobre a vida.

Aos colegas do LEMB, especialmente Carlota, Titila, Camis e Frank que me receberam nesta família.

Ao programa de pós-graduação em Bioinformática que me deu uma oportunidade de ingressar na USP.

À CAPES que me financiou durante minha pós-graduação.

À FAPESP que financiou os projetos que possibilitaram minha pesquisa.

Ao grupo de pesquisa em arbovirologia do Instituto Pasteur de Dacar, com quem colaborei durante todo meu doutorado. Especialmente para Dr Amadou Alpha Sall que lidera este grupo.

Ao Prof. Manfred Weidman pela colaboração e confiança para investigar a evolução do TBEV na Europa.

À srta Patricia Martorelli, secretária da pós-graduação, que sempre está disponível para nos auxiliar em nossas demandas.

Aos Professores Alan Duhan e Ricardo Vêncio que presidiram a CPG da Bioinformática durante minha estadia no programa.

Aos professores Arthur Gruber, Ariane Machado e Sergio Matioli que participaram do meu exame de qualificação de Mestrado, quando passei para o Doutorado direto.

Aos professores Fabio Nakano, Robson de Souza e Sergio Matioli que participaram da minha banca de qualificação de doutorado.

A todos os professores que participam da minha vida desde os meus dois anos de idade, CEI, Salesiano, UFRN e USP.

“É melhor falar errado a coisa certa do que falar certo a coisa errada”.

Patativa do Assaré

Sumário

Resumo	7
Abstract.....	8
Lista de Abreviações	9
Introdução	11
Flaviviridae	15
Vírus Zika.....	18
Vírus da Encefalite Transmitida por Carrapatos.....	20
Bunyaviridae	22
Vírus da Febre de Grande Fenda Africana.....	24
Objetivos.....	29
Métodos	30
Aquisição de sequências nucleotídicas.....	30
Detecção de recombinação homóloga.....	30
Análises filogenéticas	31
Detecção de rearranjos genômicos.....	31
Análise de seleção	32
Estimativa de distâncias patrísticas.....	32
Análises filodinâmicas.....	32
Reconstruções de caracteres discretos.....	33
Resultados e discussão.....	35
Vírus Zika.....	35
Detecção de recombinação homóloga	35
Análises filogenéticas	37
Análise de seleção	39
Análises filodinâmicas	40
Movimentação de ZIKV para a África ocidental.....	41
Hospedeiros de ZIKV	45
Contexto filodinâmico dos eventos de recombinação em ZIKV.....	47
Padrões de glicosilação no envelope de ZIKV.....	49
Mudança evolucionária correlacionada ao longo das filogenias de ZIKV	53
Implicações biológicas de nossos achados sobre ZIKV	53
Vírus da Encefalite Transmitida por Carrapatos (TBEV).....	54
Análises filodinâmicas	54
Filogeografia	55
Vírus da Febre da Grande Fenda Africana.....	60

Detecção de recombinação homóloga	60
Análises filogenéticas	60
Análises de seleção	64
Análises filodinâmicas	65
Reconstrução de localidades ancestrais	66
Uma epidemia na Mauritânia pela reemergência de uma antiga linhagem	70
Rearranjos genômicos de RVFV	70
Padrões de adjacência em filogenias indicam reagrupamento de segmentos	71
Distâncias patrísticas sugerem diferentes escalas temporais para cada segmento	73
Comparação dos regimes seletivos entre os segmentos	74
Dinâmicas evolucionárias diferentes por segmento	80
Conclusões	83
Bibliografia	84
Anexo I	94
Anexo II	104
Anexo III	115
Anexo IV	125
Anexo V	129

Resumo

No presente trabalho, investigamos a filodinâmica de três modelos virais diferentes, utilizando técnicas baseadas em verossimilhança e inferência bayesiana. Dois desses são flavivírus com genoma de RNA fita simples e senso positivo. O terceiro é um bunyavírus com genoma tri-segmentado de RNA fita simples com senso negativo. Estes diferentes modelos permitiram estudar diferentes mecanismos promotores de diversidade viral, reagrupamento de segmentos genômicos (“shift”) e mutação (“drift”), que atuam em diferentes granularidades. Descrevemos pela primeira vez o espalhamento geográfico das linhagens de vírus Zika (ZIKV) em um nível continental, assim como ocorrência de recombinação e associação entre padrões de glicosilação e vetores. Para o flavivírus da encefalite transmitida por carrapatos (TBEV), investigamos seu espalhamento e encontramos evidências que corroboram a hipótese de circulação viral restrita a focos na Europa central. As análises sobre o vírus da Febre da Grande Fenda Africana (RVFV) apontaram a ocorrência de reagrupamento de segmentos genômicos e também ajudaram a elucidar sua dispersão do leste do continente africano para o oeste, encontrando-se diversas introduções no Senegal e Mauritânia. Aparentemente, este vírus teve a entrada facilitada nesses países por uma região que funciona como um centro de dispersão (“hub”) por ser encontro de rotas migratórias de animais. Ademais, investigamos a ocorrência de rearranjos de segmentos genômicos de RVFV e também estudamos as diferenças nas dinâmicas evolutivas de cada segmento.

Abstract

In this study, we investigated the phylodynamics of three different viral models, using techniques based on maximum likelihood and Bayesian inference methods. Two of these viruses are flaviviruses, whose genomes are formed by a single-stranded positive-sense RNA molecule. The third is a Bunyavirus with tri-segmented single-stranded RNA genome with negative sense. These different models allowed us to investigate two different mechanisms to promote viral diversity, (i) recombination of genomic segments ("shift") and (ii) mutation ("drift"), therefore exploring different levels of granularity of evolutionary process. We described for the first time the geographic spread of Zika virus (ZIKV) strains in a continental level, as well as, the occurrence of recombination and association between glycosylation patterns and vectors. For the other Flavivirus, tick-borne encephalitis virus (TBEV), we investigated its spreading and found evidences to support the hypothesis that viral circulation is very constrained by the foci in central Europe. The analyses about the Rift Valley Fever Virus (RVFV) revealed the occurrence of reassortment of genomic segments and their dispersal from eastern Africa to the west, with several introductions to Senegal and Mauritania. Apparently, the entry of RVFV in these countries was facilitated by the region of Kedougou, where several migratory routes of animals converge. This place maybe works as a hub to spread RVFV for West Africa. Moreover, we also investigated the differences in evolutionary dynamics of each genomic segment of RVFV.

Lista de Abreviações

3'NCR	Região não codificante 3'
AI	Índice de associação
BA	Bandia
BF	Burkina-Faso
BF	Fatores de Bayes
BSP	plots de <i>skyrides</i> bayesianos
BSSVS	seleção variável de busca estocástica bayesiana
C	Proteína do Capsídeo
C	Segmento curto
CF	República Centro-Africana
CHKV	Vírus Chikungunya
CTMC	Cadeia de Markov tempo contínua
DA	Dacar
DE	Dezidougou
DENV	Vírus Dengue
E	Proteína do Envelope
FM	Estados Federados da Micronésia
GTR	Tempo reversível generalizado
HPD	Densidades probabilísticas mais altas no intervalo de 95 %
I	Isoleucina
KE	Kedougou
Km	Quilômetro
L	Segmento longo
LEMB	Laboratório de Evolução Molecular e Bioinformática
lnL	Logarítmico natural da Verossimilhança
M	Segmento médio
MC	Tamanho dos clados monofiléticos
MCC	Máxima credibilidade de clados
MCMC	Algoritmo de Monte Carlo aplicado a Cadeias de Markov
MY	Malásia
N	Proteína do Nucleocapsídeo
Ne.g	Tamanho populacional efetivo vezes o tempo geracional
NG	Nigéria
NS	Proteína não-estrutural
NS1	Proteína não-estrutural 1
NS2	Proteína não-estrutural 2
NS3	Proteína não-estrutural 3
NS4	Proteína não-estrutural 4
NS5	Proteína não-estrutural 5

prM	Proteína Pré-membrana
PS	Valor de parcimônia
PST	Conjunto de árvores filogenéticas obtidas a <i>posteriori</i>
RNA	Ácido Ribonucleico
RVFV	Vírus da Grande Fenda Africana
SA	Saboya
SLAC	Contagem única de ancestral por verossimilhança
SS	Sokala-Sobara
ssRNA	RNA de fita simples
T	Treonina
TBEV	Vírus da Encefalite Transmitida por Carrapatos
UG	Uganda
YFV	Vírus da Febre Amarela
ZIKV	Vírus Zika

Introdução

Uma das preocupações mais fundamentais da Biologia como ciência é a classificação biológica, sendo alvo de preocupação de filósofos e cientistas desde o período clássico, quando já se percebe uma nítida intenção de encapsular os seres vivos em classes para os estudar. Por exemplo, Aristóteles propõe uma teoria de definição em que a melhor maneira de se criar uma definição é encontrar o grupo próximo em que a característica definidora reside, ou seja, pode-se inferir dessa afirmação uma tendência a se agrupar os seres vivos em categorias limitadas por características compartilhadas pelos membros do grupo. Para ter valor discriminatório esta característica deve ser essencial para se distinguir entre as categorias, obedecendo a condição de *differentia*. O sistema classificatório de Aristóteles também se destaca dos seus predecessores pela orientação funcional em que ele se sustenta, além da estrutura hierárquica entre os graus de descrição: gênero e espécie (ARISTOTLE, 2004, 2012).

Apesar da óbvia utilidade de organização didática da categorização biológica, ressalvas devem ser feitas ao uso indiscriminado do termo espécie que segundo Darwin em sua obra prima em 1872 - A Origem das Espécies - parece ser um termo convenientemente aplicado arbitrariamente a um grupo de indivíduos muito semelhantes entre si (DARWIN, 2004). Portanto, um sistema classificatório baseado exclusivamente em compartilhamento de características pode não ser informativo sobre os processos de especiação subjacentes que produziram os padrões observados. A única figura presente na sexta edição de A Origem das Espécies (reproduzida na Figura 1) expõe uma árvore teórica de relações entre organismos hipotéticos, relacionado a existência dos níveis taxonômicos à separação entre as espécies e variedades e sua subsequente divergência e extinção (TASSY, 2011). Este é o motivo pelo qual a árvore de Darwin é considerada um esquema tanto de padrões quanto de processos, padrão de relações de descendência e processo de diversificação. Os processos que ligam o passado ao presente de um sistema, eles podem ser de dois tipos: (i) destruidor ou (ii) conservador da informação (Figura 2), dificultando inferências a respeito deles baseadas em padrões observados no presente (SOBER, 1988).

Figura 1. Árvore filogenética teórica de Darwin. Esta ilustração representa tanto um esquema de padrão (padrão de descendência) quanto de processo (processo de diversificação). Figura retirada e modificada de: (TASSY, 2011).

Figura 2. Esquema ilustrativo sobre a qualidade dos processos que ligam estados passados de um sistema ao presente. No lado esquerdo, observa-se um processo destruidor da informação, enquanto no lado direito conserva-se a informação. Esta figura foi inspirada em: (SOBER, 1988).

Não obstante, os fundamentos do conceito moderno de filogenias podem também ser creditados a Darwin, uma vez que na sexta e última edição de sua obra-prima ele incluiu o termo Filogenia recém

cunhado por Ernst Haeckel em 1866 e também dois conceitos estabelecidos por Lankester em 1870 (Figura 3): (i) homogenia, *i. e.*, semelhança devido a descendência o que conhecemos atualmente por homologia e (ii) homoplasia, *i.e.*, semelhança não relacionada a descendência (TASSY, 2011). Portanto, as bases dos conceitos que embasariam as noções do que conhecemos atualmente por filogenia foram estabelecidas no fim do século XIX, junto à teoria mais aceita de evolução natural proposta por Charles Darwin.

Figura 3. Homologia versus homoplasia. Esta ilustração demonstra relações homólogas e homoplásticas da característica ‘C’. No clado ‘1’ C é homólogo, pois existe também no ancestral dos terminais. Uma situação que se repete no clado 2. No entanto, ao se comparar estes dois clados a característica ‘C’ é homoplástica, uma vez que ela não está presente no ancestral comum mais recente, o nó ‘3’. Portanto, as linhagens ‘1’ e ‘2’ independentemente mudaram da característica ‘B’ para ‘C’. Imagem retirada e modificada de (HALL, 2003)

No que concerne aos vírus, o sistema de classificação mais aceito (Figura 4), proposto por David Baltimore, agrupa-os em famílias de acordo com a natureza do genoma e tipo de estratégia de replicação (BALTIMORE, 1971). Neste sistema os vírus cujos genomas são constituídos por uma única fita de ácido ribonucleico (ssRNA) são classificados de acordo com o senso em positivos (ssRNA+) e negativos (ssRNA-). Estas duas classes, IV e V respectivamente, diferem principalmente na forma como se processa a tradução das proteínas virais. Os ribossomos da célula hospedeira acessam diretamente o ssRNA+ dos vírus do grupo IV, já os vírus do grupo V necessitam transcrever seus ssRNA- em uma molécula complementar com senso positivo, possibilitando seu reconhecimento pelos ribossomos. Comodamente, a evolução rápida de vírus de RNA os torna um modelo apropriado para inferências filodinâmicas, uma vez que estes eventos estão inseridos em uma reduzida escala temporal (HOLMES; GRENFELL, 2009) e datas de isolamento podem ser usadas, com relativa

confiabilidade, como calibradores temporais. Objetivando-se comparar a evolução das duas diferentes classes virais (IV e V), escolhemos como modelo dois flavivírus (ssRNA+) e um bunyavírus (ssRNA-).

Figura 4. Esquema da classificação de Baltimore. Para estudar evolução de vírus de RNA escolhemos representantes de duas classes, flavivírus do grupo IV e buniavírus do grupo V. Figura retirada e modifica de: http://viralzone.expasy.org/all_by_species/254.html.

As mudanças genéticas que acontecem nos vírus de RNA ocorrem basicamente devido a dois mecanismos bem conhecidos (Figura 5): (i) mutações pontuais, *drift*, e (ii) rearranjos de fragmentos genômicos, *shift*. Estes últimos podem ser de dois tipos: recombinação, processo em que ocorre a quebra e junção de partes do material genético de origens distintas ou permuta de segmentos genéticos inteiros, no caso de vírus com genoma segmentado como os Buniavirus (SIMON-LORIERE; HOLMES, 2011).

Figura 5. Mudanças em ampla escala que ocorrem em flavivírus e bunivírus. No lado direito, esquema representativo de recombinação, onde a RNA polimerase troca de fita molde durante a síntese de uma nova cadeia. No lado esquerdo, permuta de segmentos inteiros durante a montagem da partícula viral. O último evento pode ocorrer em bunivírus, cujos genomas são formados por três segmentos. Figura retirada e modificada de (SIMON-LORIERE; HOLMES, 2011).

Flaviviridae

Esta família viral pertence ao grupo IV da classificação de Baltimore (Figura 4), o genoma consiste em uma única fita de RNA de sentido positivo com tamanho aproximado de 11 mil pares de base, encapsidado na extremidade 5'. Ele é flanqueado por duas regiões não codificantes (5' e 3') e possui uma única fase de leitura que codifica uma única poli proteína: 5'-C-prM-E-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5-3' (CHAMBERS et al., 1990; KUNO; CHANG, 2007), clivada em proteína do capsídeo, proteína de membrana, envelope, proteínas não estruturais (Figura 6A). Interessantemente, percebe-se uma tendência dos genes virais a se organizar espacialmente no genoma de acordo com a função desempenhada. Os genes que codificam proteínas estruturais se localizam mais próximos à extremidade 5', seguidos pelos não-estruturais (Figura 6A).

As partículas dos vírus desta família são envelopadas, esféricas e tem aproximadamente 50 nm de diâmetros (Figura 6B). Possuem apenas três proteínas estruturais (Figura 6), capsídeo (C),

membrana (M) e envelope (E). A simetria icosaédrica da partícula parece ser conferida pelo arranjo entre as duas proteínas de superfície, E e M (LINDENBACH; RICE, 2003).

Figura 6. A) Representação esquemática do genoma de flavivírus. B) Ilustração do vírion de flavivírus. A figura A foi retirada e modificada de (GUZMAN et al., 2010) e figura B foi retirada e modificada de http://viralzone.expasy.org/all_by_species/24.html.

As três proteínas estruturais são transcritas primeiro. O capsídeo (C) é formado por uma proteína muito básica de aproximadamente 11kDa. Os resíduos carregados que provavelmente interagem com o RNA viral são agrupados nos *termini* amina e carboxi, sendo separados por domínios hidrofóbicos que mediam a associação com a membrana. Ademais, o C-terminal da proteína C também contém uma âncora hidrofóbica que serve como um peptídeo de sinal para a translocação do precursor da proteína M (prM) no retículo endoplasmático (RE), onde a prM é clivada na sua forma madura M que forma o vírion (Figura 6B). Relata-se também uma função semelhante a uma de leite da fração pr da proteína prM para o envelope viral (LINDENBACH; RICE, 2003). Esta última proteína estrutural, E, é formada por uma proteína de 53 kDa, dimérica (Figura 5B) e N-glicosilada que intermedia a ligação e fusão com a membrana celular do hospedeiro (MODIS et al., 2004). Além de participar da resposta imune humoral protetora pelo estímulo a neutralização mediada por anticorpos (CHEN; MAGUIRE;

MARKS, 1996). Cada subunidade monomérica do envelope é composta de três domínios distintos. O domínio I é uma estrutura de barril β paralelamente orientado a membrana viral. O domínio II é composto por um par de loops descontínuos e um deles que apresenta uma estrutura altamente conservada entre os flavivírus funcionando como um peptídeo de fusão interno, estabilizado por três pontes dissulfeto. Ao domínio III é atribuída a função de interagir com receptores celulares para entrada na célula e também é sugerido que contém resíduos responsáveis pela determinação da resposta do hospedeiro, tropismo e virulência entre flavivírus (REY et al., 1995). Também se encontra nesta região os mais importantes epítomos para a neutralização por anticorpos protetores. Mutações neste domínio têm implicações relevantes no potencial de virulência (WEAVER; VASILAKIS, 2009).

Entre as proteínas não estruturais (Figura 6A), a proteína NS1 é dimérica, com 46 kDa, N-glicosilada, e existe nas formas intra e extracelulares (JACOBS et al., 2000). A proteína NS2A tem 22 kDa e está envolvida nas mudanças entre o empacotamento do RNA e a replicação como também possivelmente em efeitos antagonistas ao interferon (KHROMYKH et al., 2001; MUÑOZ-JORDAN et al., 2003). NS2B tem 14 kDa é ligada a membrana e associada a NS3 forma o complexo de proteases viral, ainda atua como cofator na ativação de NS3 (ERBEL et al., 2006). NS3 é uma proteína multifuncional de 70 kDa com atividade de serina protease, helicase e RNA trifosfatase. Também está envolvida no processamento da poliproteína viral e na replicação do RNA (LI et al., 1999). NS4A e NS4B são pequenas proteínas hidrofóbicas de 16 e 27 kDa respectivamente que inibem a sinalização do interferon (JONES et al., 2005). NS5 é uma grande proteína multifuncional de 103 kDa com atividades de processamento do RNA viral (adição do cap), RNA polimerase dependente de RNA, indução de interleucina oito e localização nuclear (EGLOFF et al., 2002).

A família Flaviviridae se divide em três gêneros: (i) Flavivirus (setor azul na Figura 7), (ii) Hepacivirus (setor rosa na Figura 7) e (iii) Pestivirus (setor verde na Figura 7). Embora os dois últimos gêneros tenham estratégias de replicação semelhantes aos Flavivirus, eles representam grupos antigenicamente distintos e não são transmitidos por artrópodes. Provavelmente, Hepacivirus e Pestivirus divergiram a muito tempo dos Flavivirus (CALISHER; GOULD, 2003). Diante disto, decidimos representar nesta tese somente os vírus protótipos da família, *i.e.*, o gênero Flavivirus. Inclusive, existe uma nítida separação filogenética no interior do gênero Flavivirus entre os vírus que são transmitidos por ácaros (*tick-borne* representados pelo setor vermelho na Figura 7) e por mosquitos (*mosquito-borne* representados pelo setor amarelo na Figura 7). Não obstante, alguns dos membros desta família viral são agentes causais de importantes epidemias, como Dengue (DENV) e Febre Amarela (YFV). Inclusive, o nome da família se refere ao vírus da febre amarela, uma vez que a palavra *flavus* em latim significa amarelo. Neste trabalho de doutorado, investigamos um possível vírus emergente espalhado majoritariamente pela África e sudeste asiático, vírus Zika (ZIKV), e outro

endêmico da Europa central, vírus da encefalite transmitido por carrapatos (TBEV), representando os dois principais grupos de Flavivirus.

Figura 7. Relações filogenéticas na família Flaviviridae a partir da sequência proteica da polimerase viral. O setor azul demarca o gênero Flavivirus; o rosado, Hepacivirus e o verde, Pestivirus. O gênero Flavivirus foi colorido internamente para realçar o espectro de hospedeiros dos vírus: vírus que infectam somente insetos em laranja, sem artrópode vetor conhecido em roxo, transmitidos por mosquitos em amarelo e transmitidos por carrapatos em vermelho. Figura modificada e adaptada de (LOBO et al., 2009).

Vírus Zika

O vírus Zika (ZIKV) é um arbovírus membro da família Flaviviridae e gênero Flavivirus. O vírus é geralmente transmitido por mosquitos e seu ciclo natural de transmissão na África envolve majoritariamente vetores silvestres do gênero *Aedes* (*Ae. furcifer*, *Ae. taylori*, *Ae. luteocephalus*, *Ae. vittatus*, *Ae. opock* e *Ae. africanus*) e macacos (Figura 8). Humanos são ocasionalmente infectados, com consequências abrangendo desde a ausência de sintomas a uma síndrome semelhante a gripe por Influenza associada a febre, dor de cabeça, artralgia, mialgia, anorexia e astenia (BEARCROFT, 1956; FAGBAMI, 1979; OLSON et al., 1981; SIMPSON, 1964). Estes achados clínicos são similares aos

encontrados na febre por Dengue (DENV) e Chikungunya (CHKV), sendo necessário o diagnóstico diferencial entre essas doenças.

Figura 8. Ciclo de transmissão de vírus Zika na África. O vírus é mantido em um ciclo silvestre nas florestas, onde infecta principalmente macacos sendo transmitidos por espécies silvestres de mosquitos. Ocasionalmente, mosquitos peridomésticos podem se alimentar de animais na floresta e fazer o elo entre humanos no ambiente próximo a habitações humanas. Em seguida, o vírus pode se estabelecer em um ambiente domiciliar, inclusive já foi mostrado que ZIKV é transmitido por mosquitos que vivem no ambiente doméstico como *Aedes aegypti*. Imagem modificada do original que foi cedida por Oumar Faye do Instituto Pasteur de Dacar.

O ZIKV foi isolado pela primeira vez em 1947 em um macaco sentinela na floresta de Zika em Uganda, no ano seguinte o vírus foi novamente isolado em uma amostra de *Aedes africanus* proveniente do mesmo local (DICK; KITCHEN; HADDOW, 1952). Desde então, inquéritos sorológicos e entomológicos tem reportado esta virose no continente africano e no sudeste asiático (BARBOZA et al., 2008). A primeira epidemia descrita fora da África ocorreu em 2007 na Micronésia, resultando em 99 casos confirmados no período de dois meses (LANCIOTTI et al., 2008). Este surto evidencia a capacidade de febre por Zika se tornar uma doença emergente, a possibilidade de falso diagnóstico com Dengue pode atrapalhar a identificação de uma epidemia (HAYES, 2009). Ademais, evidências recentes sugerem que podem haver outras vias de transmissão que não exijam mosquitos vetores, como a via por contato sexual (FOY et al., 2011). No fim de 2013, outra epidemia fora da África foi reportada na polinésia francesa com mais de 35 mil casos, alertando autoridades de saúde para o potencial de espalhamento pandêmico desta virose que também foi identificada na Nova Caledônia, Ilhas Cook e Ilha de Páscoa (MUSSO; NILLES; CAO-LORMEAU, 2014).

Figura 9. Distribuição geográfica de Zika no mundo. Regiões onde ZIKV foi isolado em humanos estão coloridas em vermelho. Os lugares com apenas evidência sorológica estão marcados em tons pálidos de rosa. Figura retirada e modificada de: http://commons.wikimedia.org/wiki/File:Zika_virus.png.

No oeste da África, o programa de vigilância sorológica e entomológica reporta a circulação periódica de ZIKV desde 1968 (<http://www.pasteur.fr/recherche/banques/CRORA/>). Um total de 606 ocorrências de ZIKV foram notificados no período de 1968 e 2002. Destes, 594 foram reportados em material entomológico, três em primatas não humanos e nove em humanos. Não obstante, pouco se compreende a respeito das relações genéticas entre essas amostras virais. Portanto, apresentamos nesta tese a primeira caracterização filogeográfica dos isolados de ZIKV do oeste da África, enfatizando a circulação regional da virose no Senegal e na Costa do Marfim. Utilizamos sequências nucleotídicas de 42 isolados virais amostrados em mosquitos, humanos e outros mamíferos durante o período de 1947 a 2007.

Vírus da Encefalite Transmitida por Carrapatos

Os vírus de encefalite transmitidos por carrapato (TBEV), *Tick-Borne Encephalitis Virus* em inglês, são também classificados na família Flaviviridae, devido a estrutura e genoma viral característicos. Formam um grupo distinto dos flavivírus transmitidos por mosquitos (DOBLER et al.,

2012). A patologia associada, encefalite transmitida por carrapato (TBEV), é caracterizada por meningite asséptica acompanhada por febre alta, cefaleias, náuseas, êmese e vertigem, e representa um dos mais importantes acometimentos do sistema nervoso central nos países da Europa central e ocidental (KAISER, 2012). Em concordância com o nome, o vírus é primariamente transmitido por carrapatos para hospedeiros vertebrados (Figura 10), principalmente pequenos roedores na Europa central (Figura 11) (WEIDMANN et al., 2006). Humanos e aves podem ser hospedeiros incidentais (DOBLER et al., 2012).

A diversidade de hospedeiros possibilita um complexo ciclo de transmissão, repercutindo nos padrões de espalhamento e endemidade desta virose na Europa (Figura 10 e 11). Aparentemente, o TBEV pode ser carregado a longas distâncias por aves migratórias (HOOGSTRAAL et al., 1963; WALDENSTRÖM; LUNDKVIST, 2007) e mantido endêmico infectando pequenos roedores (KNAP et al., 2012) em áreas delimitadas, focos locais (LABUDA et al., 2002), sob um regime de evolução local (WEIDMANN et al., 2011). Objetivando compreender como estes vírus se dispersam dentro de e entre os focos, sequências parciais dos envelopes virais foram empregadas para investigar diversidade genética local e geral, regimes de dispersão viral na Europa central e investigar evidências de focos locais deste vírus na Europa central.

Figura 10. Ilustração esquemática do ciclo de transmissão de Vírus da Encefalite Transmitida por Carrapatos. Humanos são infectados ocasionalmente quando adentram o habitat de carrapatos e roedores, por exemplo, ao realizar atividades como ecoturismo e manipular alimentos de origem animal. Figura retirada e modificada de (PFEFFER; DOBLER, 2010).

Figura 11. Distribuição de linhagens do Vírus da Encefalite Transmitida por Carrapatos na Eurásia. As áreas foram coloridas de acordo com a linhagem predominante na região analisada. Figura retirada e modificada de http://en.wikipedia.org/wiki/Tick-borne_encephalitis.

Bunyaviridae

Estes vírus apresentam um genoma de RNA de fita simples dividido em três segmentos: longo (L), médio (M) e curto (S do inglês *Small*) (Figura 12). O segmento L codifica uma RNA polimerase dependente de RNA (MÜLLER et al., 1994), o segmento M codificada as glicoproteínas de envelope G_N e G_C e também duas proteínas não-estruturais (NSm) de 14 e 78 kDa (Figura 12), respectivamente (COLLETT et al., 1985). O segmento S, somente em Phlebovirus e Tospovirus, caracteriza-se por uma estratégia de codificação em ambos os sentidos, uma vez que codifica a proteína de nucleocapsídeo (N) no sentido positivo 5'-3' e no sentido negativo 3'-5' a proteína não-estrutural (NSs) (GIORGI et al., 1991) (Figura 12). A última proteína está relacionada com imunidade inata de mamíferos, por interferir nas vias de expressão do gene do Interferon (BILLECOCQ et al., 2004).

Figura 12. Representação esquemática de partícula viral e genoma segmentado de Phlebovirus.

Figura retirada e modificada de (PEPIN et al., 2010).

De acordo com o comitê internacional para taxonomia de vírus (<http://www.ictvonline.org/>), a família Bunyaviridae não é incluída em nenhuma ordem e possui cinco gêneros conhecidos (Figura 13): (i) Hantavirus com 24 espécies, (ii) Tospovirus com nove espécies, (iii) Orthobunyavirus com 48 espécies, (iv) Nairovirus com sete espécies e (v) Phlebovirus com nove espécies. Como exemplo desta família, estudamos o vírus da Febre da Grande Fenda Africana (RVFV) que é a espécie mais conhecida do gênero Phlebovirus e possui uma grande importância médica e veterinária, tendo sido isolado no Quênia pela primeira vez na década de 1930.

Figura 13. Relações filogenéticas na família Bunyaviridae. Os cinco gêneros descritos pelo ICTV estão coloridos de acordo. Esta árvore foi inferida a partir da sequências proteicas da RNA polimerase viral, utilizando o programa FastTree (PRICE; DEHAL; ARKIN, 2010).

Vírus da Febre de Grande Fenda Africana

O vírus da Febre da Grande Fenda Africana, *Rift Valley Fever virus* (RVFV) em inglês, família Bunyaviridae, gênero Phlebovirus, está relacionado a epidemias de elevada magnitude afetando rebanhos e humanos por todo o continente africano. Desde a primeira década do século XX há relatos sobre uma doença semelhante em ovinos. No entanto, o vírus foi isolado somente no ano de 1931, em um uma fazenda de ovelhas localizada na porção queniana da grande fenda africana (DAUBNEY; HUDSON, 1932; FINDLAY; DAUBNEY, 1931). O primeiro surto fora da África foi reportado na península arábica, em 2000 (WORLD HEALTH ORGANIZATION, 2000).

RVFV é primariamente transmitido entre animais e humanos por mosquitos do gênero *Aedes* (Figura 14), podendo causar surtos geralmente caracterizados por numerosos abortos e óbitos em rebanhos caprinos e ovinos (GOULD; HIGGS, 2009; HOOGSTRAAL et al., 1979; LAUGHLIN et al., 1979; MEEGAN, 1979; PEPIN et al., 2010). Os primeiros registros na década de 1930 referem-se a numerosos abortos em ovelhas recém introduzidas no Quênia. Possivelmente, a introdução brusca de

uma grande população de hospedeiros susceptíveis relaciona-se com o intercâmbio do ciclo silvestre para o epidêmico (HANOTTE et al., 2002; PEPIN et al., 2010). Ademais, existem evidências de correlação positiva entre alterações climáticas e emergência de RVFV (Figura 14). Por exemplo, na região do chifre da África, pluviosidade e incidência parecem estar correlacionados (ANYAMBA et al., 2010). Todavia, este padrão não é observado no oeste africano ao sul do Saara, onde o vírus parece estar em um ciclo enzoótico (TRAORÉ-LAMIZANA et al., 2001; ZELLER et al., 1997). Ademais, as alterações climáticas recentes como aumento de pluviosidade e temperatura podem propiciar a disseminação de vetores (GOULD; HIGGS, 2009), o que tem estimulado avaliações de risco do espalhamento desta virose para outros continentes (CHEVALIER et al., 2010; HARTLEY et al., 2011).

Assim como em outros Buniavírus, o genoma tri-segmentado de RVFV permite a ocorrência de reagrupamento (SALL et al., 1999; SIMON-LORIERE; HOLMES, 2011), que pode ser uma maneira de promover diversidade genética (Figura 5). Dado que o sucesso deste evento requer diferentes vírus co-infectando um hospedeiro e RVFV é capaz de infectar uma ampla gama de mamíferos com sucesso, ovinos, caprinos, bovinos, equinos e bufalinos, assim como dípteros vetores (PEPIN et al., 2010), o reagrupamento de segmentos genômicos pode ter sido um evento recorrente na disseminação do vírus ao longo do século XX. Não obstante, a organização segmentada do genoma permite que os segmentos tenham experimentado distintas histórias evolutivas devido a diferentes pressões seletivas (RAMBAUT et al., 2008). Embora tenha sido previamente demonstrado que o reagrupamento genético ocorre na natureza (SALL et al., 1999), estimativas deste evento se tornaram difíceis devido à reduzida disponibilidade de sequências de localidades distintas como o oeste da África. Estudos recentes reportaram ausência de evidências de reagrupamento genômico (BIRD et al., 2007) e algumas inferências filogenéticas foram realizadas sem considerar a ocorrência deste evento, utilizando concatenados de sequências (NDERITU et al., 2011).

Figura 14. Ciclo de transmissão de RVFV. Neste esquema nota-se três níveis de transmissão e o papel dos hospedeiros em cada um deles. No topo da figura, retrata-se o ciclo de transmissão endêmica com transmissão trans-ovariana em mosquitos que responsável pela manutenção de RVFV no ambiente em períodos de estiagem. No centro da figura, a transmissão epizootica em que humanos e ruminantes são infectados após uma explosão populacional de mosquitos vetores, geralmente relacionada a aumento da pluviosidade. No a parte inferior, o ciclo epidêmico, grande transmissão entre humanos, que é raro e pode ocorrer por contato com materiais infectados derivados de animais e também por mosquitos. Figura retirada e modificada de (PFEFFER; DOBLER, 2010).

A doença permaneceu como preocupação veterinária, até causar uma extensiva epidemia no Egito em 1977 (HOOGSTRAAL et al., 1979; LAUGHLIN et al., 1979), onde causou aproximadamente um milhão de infecções humanas, 600 óbitos e severas epizootias (MARTIN et al., 2008). Em seguida estudos adicionais sobre a transmissão de RVFV expuseram outros surtos em países da África subsaariana onde o vírus é endêmico (CHEVALIER; THIONGANE; LANCELOT, 2009; FAYE et al., 2007; MARRAMA et al., 2005; SALL et al., 1998a; TRAORÉ-LAMIZANA et al., 2001; ZELLER et al., 1997). A passagem do ciclo enzoótico (*i.e.*, entre animais silvestres) para a transmissão peridoméstica (*i.e.*, entre animais como gado, ovelhas e cabras) é cogitada como um dos fatores responsáveis por ocasionar epidemias (WEAVER, 2005). Recentemente, surtos foram reportados no Quênia, Somália e Tanzânia (WORLD HEALTH ORGANIZATION, 2007), África do Sul (ARCHER et al., 2011; WORLD HEALTH ORGANIZATION, 2010), Mauritânia (EL MAMY et al., 2011) e Sudão – onde mais de 200 mortes humanas foram reportadas (WORLD HEALTH ORGANIZATION, 2008). Além disso, RVFV foi relatado pela primeira vez fora da África, na península arábica, atingindo Iêmen e Arábia Saudita (WORLD HEALTH ORGANIZATION, 2000). Na Figura 15, pode-se verificar a ocorrência de RVFV na África.

Figura 15. Distribuição de RVFV na África. Em azul, países onde a doença é endêmica ou apresentaram epidemias da febre da grande fenda africana. Em verde, países com casos esporádicos. Figura retirada e modificada de http://en.wikipedia.org/wiki/Rift_Valley_fever.

utilizando métodos baseados em filogenia molecular, esforçaram-se em elucidar os mecanismos subjacentes à distribuição e dispersão na África (BIRD et al., 2007, 2008; GROBBELAAR et al., 2011; NDERITU et al., 2011; SALL et al., 1997, 1998a, 1998b). Estes estudos revelaram que: (i) os genomas de RVFV apresentam reduzida diversidade genética abrangendo de 1% a 5% de distâncias genéticas (BIRD et al., 2007), (ii) mudam seguindo taxas evolucionárias entre $2,8E-4$ a $3,9E-4$ substituição/sítio/ano (PEPIN et al., 2010), (iii) pode ser classificado em diversas linhagens (BIRD et al., 2007; GROBBELAAR et al., 2011), (iv) pode ser agrupado por origem geográfica (SALL et al., 1998a), (v) reagrupamento genômico, ocorre na natureza (SALL et al., 1999) e (vi) o conteúdo de diversidade genética coalesce aproximadamente há 120 a 130 anos atrás (BIRD et al., 2007). No entanto, excetuando-se a região do chifre da África (PEPIN et al., 2010), o conhecimento sobre evolução de RVFV no interior em escala regional é limitado. Esta lacuna de informação demonstra a necessidade de se investigar os determinantes da circulação regional de RVFV no oeste do continente africano.

Objetivos

Objetivo Geral

Este trabalho pretende estudar mecanismos promotores de diversidade genética em diferentes linhagens virais cujos genomas são constituídos por ácido ribonucleico (RNA), assim como as consequências demográficas destes eventos diversificadores.

Objetivos Específicos

- I. Identificar eventos de mudanças em grande escala nestes genomas, recombinação em flavivírus e rearranjos entre segmentos em bunyavírus.
- II. Estimar a escala temporal em que ocorreu a separação das linhagens virais.
- III. Descrever eventos de migração do vírus Zika (ZIKV) pelo continente africano em um contexto temporal.
- IV. Estudar as consequências dos padrões de glicosilação da proteína do envelope de ZIKV para a transmissão por vetores.
- V. Elucidar o modo de espalhamento do vírus da encefalite transmitida por carrapatos (TBEV) pela Europa Central.
- VI. Descrever a migração do vírus da febre da grande fenda africana (RVFV) para o oeste da África.
- VII. Estudar as consequências dos rearranjos de segmentos de RVFV sobre as dinâmicas virais inferidas.

Métodos

Aquisição de sequências nucleotídicas

Este estudo foi realizado em parceria com o grupo de arbovirologia do Instituto Pasteur de Dacar (<http://www.pasteur.sn/>), o qual foi encarregado da coleta, catalogação, extração e amplificação do material genético viral. Durante a vigência do programa de auxílio à pesquisa *Viral Genetic Diversity Network* (VGDN), o grupo do Laboratório de Evolução Molecular e Bioinformática (LEMB) da Universidade de São Paulo foi encarregado do sequenciamento dos amplificadores de nucleotídeos e para preencher a lacuna das análises bioinformáticas foi realizado este trabalho de doutorado. Portanto, o presente estudo foi realizado com sequências de linhagens virais do Oeste da África, assim como se adicionou sequências representativas de outras localidades disponíveis em GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>). As análises com RVFV foram feitas com sequências parciais dos segmentos longo (129 nt), médio (718 nt) e curto (601 nt); as análises com ZIKV foram realizadas com sequências parciais do gene do envelope (753 nt), NS5 (701 nt) e 3'NCR (537 nt). Ademais, incluímos 10 genomas de ZIKV no estudo para as análises de recombinação. As sequências provenientes do GenBank foram mantidas integralmente e os alinhamentos foram completados com “missing data”, visto que utilizamos apenas métodos filogenéticos baseados em máxima verossimilhança, portanto este tipo de configuração não prejudica as análises subsequentes (WIENS, 2003, 2006). Ademais, o grupo do departamento de virologia da Universidade de Göttingen, como parte de nosso esforço colaborativo, forneceu sequências parciais dos envelopes de TBEV (~1570 nt), assim como a localização dos isolados geográficos. Na última parte do trabalho em que investigamos as dinâmicas evolucionárias de RVFV, utilizamos todas as sequências disponíveis no GenBank para este vírus, de onde extraímos informação a respeito de país e data de isolamento para cada sequência com um script em Perl escrito pelo aluno.

Deteção de recombinação homóloga

Inicialmente, as sequências foram alinhadas com o programa MUSCLE v3.7 (EDGAR, 2004), em seguida foram editadas manualmente com o programa SeaView v4.3.3 (GOUY; GUINDON; GASCUEL, 2010). A detecção de possíveis eventos de recombinação homóloga, identificação de sequências recombinantes e localização de pontos de quebra foram investigadas por meio dos métodos (RDP, GenConv, Chimaera, MaxChi, Bootscan, SiScan e 3Seq) implementados no programa RDP4 beta48 (MARTIN et al., 2010), utilizando-se intervalo de confiança de 95% e correção múltipla de Bonferroni para se evitar resultados falsos positivos. Sequências recombinantes potencialmente

causam erros sistemáticos de filogenia (POSADA; CRANDALL, 2002; SCHIERUP; HEIN, 2000), portanto foi estabelecida a retirada destas sequências das análises filogenéticas subsequentes. Em seguida, investigou-se a ocorrência de pontos de quebra nos genomas de ZIKV, utilizando-se uma abordagem baseada em Phylo-HMM que detecta alterações nas topologias de árvores filogenéticas sobre as colunas do alinhamento múltiplo, que podem ser devido à recombinação ou a outros eventos de rearranjos genômicos, implementada em Rec-HMM (WESTESSON; HOLMES, 2009).

Análises filogenéticas

O conteúdo de sinal filogenético dos bancos de sequências foi avaliado com auxílio do algoritmo de mapeamento de verossimilhança (STRIMMER; VON HAESLER, 1997) implementado em TREE-PUZZLE v5.2 (SCHMIDT et al., 2002) que quantifica os quartetos filogenéticos bem resolvidos nos bancos de dados. Árvores filogenéticas foram obtidas utilizando o programa GARLI v2.0 (ZWICKL, 2006) que emprega um algoritmo genético estocástico para estimar simultaneamente a melhor topologia, comprimentos de ramos e parâmetros do modelo de substituição empregado que maximizam o valor logarítmico da Verossimilhança (lnL), dez replicações independentes foram executadas sob o modelo de substituição *general time reversible* (GTR) com taxa de variação gama e uma proporção de sítios invariantes (GTR+ Γ +I). A replicação com lnL mais alto foi selecionada. Objetivando-se avaliar o suporte estatístico dos ramos das árvores, foram empregadas mil amostragens não paramétricas de *bootstrap*.

Deteção de rearranjos genômicos

Considerando a natureza trissegmentada do genoma de RVFV, investigou-se a ocorrência de rearranjo entre os segmentos genômicos de linhagens virais distintas. Inicialmente, aplicou-se uma abordagem baseada no algoritmo de Monte Carlo aplicado a cadeias de Markov (MCMC) para gerar um conjunto de árvores obtidas durante a estacionariedade do MCMC (PST) com MrBayes v3.2 (HUELSENBECK; RONQUIST, 2001; RONQUIST; HUELSENBECK, 2003). A estacionariedade do MCMC de duas execuções independentes do programa cada uma com quatro cadeias, sendo três quentes e uma fria foi obtida após 20 milhões de gerações e as árvores foram amostradas a cada 5000 gerações. Em seguida, utilizou-se o programa GiRaF v1.01 (NAGARAJAN; KINGSFORD, 2011) que busca clados incompatíveis entre conjuntos de árvores filogenéticas obtidas de diferentes segmentos genômicos. Para essa análise foram utilizadas as árvores amostradas durante a estacionariedade do algoritmo de Monte Carlo implementado no programa MrBayes v3.2, clados que apareceram em menos de 5 % das árvores amostradas foram descartados da análise e foram considerados positivos clados que

tiveram posições distintas em mais de 70 % da amostra de árvores testadas. Também utilizamos a mesma abordagem para investigar se existiam incongruências entre as filogenias estimadas com E e NS5 de ZIKV.

Análise de seleção

Para inferir as pressões de seleção agindo sobre os genes analisados, estimamos a diferença entre as taxas de substituição não sinônimas (dN) e sinônimas (dS) por sítio codificante usando algoritmo de contagem ancestral de única verossimilhança (SLAC) disponível em HyPhy v0.99 (POND; FROST; MUSE, 2005), assumindo um nível de confiança de 95% ($\alpha = 0,05$). Valores da diferença ω (onde ω representa $dN-dS$) maiores que zero são indicativos de seleção direcional positiva ($\omega > 0$), enquanto valores menores que zero indicam seleção negativa purificadora. A quantidade de sítios sob seleção em cada segmento de RVFV foi comparada com o teste exato de Fisher do programa R (<http://www.r-project.org/>). A força de seleção de cada segmento, quantidade de $dN-dS$, foi comparada com o teste de Kruskal-Wallis, também do R.

Estimativa de distâncias patrísticas

Para se estudar as distâncias patrísticas dos segmentos de RVFV, 100 árvores de verossimilhança foram estimadas independentemente com o Garli, usando GTR+ Γ +I assumindo que as condições dos modelos mais simples seriam satisfeitas ao estimar as probabilidades de transição entre os estados com o modelo mais complexo. Em seguida, selecionamos a árvore com maior valor de verossimilhança (V). As distâncias patrísticas foram então calculadas a partir desta árvore com o R-pacote adephylo v1.16 (JOMBART; BALLOUX; DRAY, 2010). As diferenças entre as distâncias patrísticas foram analisadas com o teste de comparação múltipla de Kruskal-Wallis do R-pacote pgirmess v 1.59 (GIRAUDOUX, 2013).

Análises filodinâmicas

Uma vez que as datas de isolamento das amostras virais estavam disponíveis, a partir das melhores árvores de máxima verossimilhança as taxas de substituição por sítio por ano (μ) foram estimadas com R8s v1.71 (SANDERSON, 2003) utilizando o método de verossimilhança penalizada (SANDERSON, 2002) que emprega um algoritmo semi-paramétrico utilizando diferentes taxas de substituição em cada ramo com uma penalidade não paramétrica que penaliza o modelo quando as taxas mudam muito rápido de ramo para ramo. Em seguida, árvores filogenéticas com a máxima credibilidade de clados (MCC) foram inferidas usando uma abordagem baseada em MCMC implementada em BEAST v1.6.2 (DRUMMOND; RAMBAUT, 2007) empregando GTR+ Γ +I e relógio

molecular relaxado não correlato, utilizando uma distribuição probabilística lognormal (DRUMMOND et al., 2006) com o μ previamente estimado. Ademais, foi investigada a variação no tamanho populacional efetivo vezes o tempo geracional ($Ne.g$) para se inferir a demografia viral por meio de plots de *skyrides* bayesianos (BSP) (MININ; BLOOMQUIST; SUCHARD, 2008). A convergência do MCMC foi obtida após quatro execuções independentes com 50 milhões de gerações cada uma, que foram suficientes para se obter uma amostragem apropriada da estacionaridade do MCMC que foi inspecionada com auxílio do programa Tracer v 1.5 (disponível em <http://tree.bio.ed.ac.uk/software/>), considerando-se suficiente quando os tamanhos amostrais efetivos (ESS) dos parâmetros atingiram valores superiores a 200.

Reconstruções de caracteres discretos

Objetivando inferir a história do espalhamento geográfico, a distribuição em hospedeiros e herança de sítios de glicosilação, foi utilizado o modelo de distribuição discreta em que foi atribuído um carácter representativo do estado nos nós terminais para se reconstruirmos o estado ancestral nos nós internos durante a inferência filogenética, utilizando o algoritmo de seleção variável de busca estocástica bayesiana (BSSVS) (LEMEY et al., 2009), implementado no programa BEAST v 1.8. Para as reconstruções filogeográficas do RVFV foram utilizadas granularidades diferentes de acordo com o nível de descrição. Para tal, atribuiu-se um estado de país quando se analisou o espalhamento na África, assim como foi atribuído um estado de província ao se avaliar a dispersão nos territórios do Senegal e da Mauritânia. Contudo as análises com ZIKV foram realizadas considerando-se a descrição no nível de províncias na Costa do Marfim e no Senegal e uma vez que os isolados virais dos demais países são representados por amostras de apenas uma localidade discriminou-se-os no nível de países. Ademais, estados ancestrais baseados no sítio de glicosilação na proteína do envelope e no hospedeiro de isolamento foram inferidos nas genealogias de ZIKV. Avaliou-se a associação entre os fenótipos virais de ZIKV com as filogenias inferidas, por meio da discrepância entre as associações observadas e o esperado por amostragem aleatória após 10000 replicações independentes, medida pelo valor de parcimônia (PS), o índice de associação (AI) e o tamanho dos clados monofiléticos (MC), usando-se o programa BaTS (PARKER; RAMBAUT; PYBUS, 2008). Mudanças evolucionárias correlacionadas entre fenótipos de ZIKV (padrão de glicosilação e hospedeiro vetor) foram investigadas ao longo das filogenias deste vírus, empregando-se um abordagem baseada em verossimilhança para se testar a adaptabilidade de modelos distinto, um em que os caracteres evoluem independentemente e outro de forma dependente (BARKER; PAGEL, 2005), usando o programa BayesTraits (<http://www.evolution.rdg.ac.uk/>). Ademais, foi aplicada uma metodologia de filogeografia semelhante para se inferir os movimentos ancestrais de migração de TBEV na Europa central, discriminando as

localidades de acordo com a província de isolamento na República Checa (Boemia central, sul da Boemia, norte da Boemia, Norte da Morávia) e para os demais países (Áustria, Alemanha e Eslováquia) em que as amostras originam-se de uma única região utilizamos a discriminação em termos de países. Adicionalmente, objetivando investigar as relações entre distâncias genéticas e geográficas, calculamos as distâncias patrísticas, utilizando o programa Patristic (FOURMENT; GIBBS, 2006), a partir da árvore filogenética com melhor valor de verossimilhança obtida após 100 execuções independentes do programa GARLI. As distâncias geográficas entre dois locais de isolamento foram estimadas utilizando um script em Perl. As correlações entre as matrizes de distância, genética e geográfica, foram inspecionadas com o teste de Mantel implementado no pacote APE v3.0.2 (PARADIS; CLAUDE; STRIMMER, 2004) no ambiente R (<http://www.r-project.org/>). As relações entre as distâncias também foram investigadas com um correlograma de Mantel (CARREL et al., 2010), que foi gerado com correção múltipla de Bonferroni e coeficiente de correlação de Spearman (ρ), utilizando o pacote VEGAN v2.05 (OKSANEN et al., 2012). A taxa de evolução (μ) que melhor se adaptou a cada segmento de RVFV foi escolhida baseada nos testes de critério de informação do Akaike através de cadeias de Markov (AICM) e também do *path-sampling* (BAELE et al., 2012)

Resultados e discussão

Vírus Zika

Os resultados descritos nesta seção foram majoritariamente publicados no artigo “Molecular evolution of Zika virus during its emergence in the 20th century” no jornal PLoS Neglected Tropical Diseases (doi: 10.1371/journal.pntd.0002636) que segue no Anexo I.

Detecção de recombinação homóloga

As análises preliminares utilizando RDP com os genomas disponíveis de ZIKV detectaram eventos de recombinação por métodos diferentes (Tabela 1). Apesar de alguns trabalhos demonstrarem a ocorrência de recombinação em flavivírus (HOLMES; WOROBEY; RAMBAUT, 1999; TOLOU et al., 2001), estes eventos não são considerados frequentes nesta família viral (SIMON-LORIERE; HOLMES, 2011; TAUCHER; BERGER; MANDL, 2010). Em seguida, objetivando investigar mais exhaustivamente este achado, empregamos um método mais exaustivo baseado em uma variação do modelo oculto de Markov, PHYLO-HMM, que avalia a probabilidade de mudança filogenética percorrendo as colunas do alinhamento múltiplo de sequências (WESTESSON; HOLMES, 2009), estas análises também indicaram recombinação, assim como a estimativa de pontos de quebra nas posições 1044, 1095, 5181, 5238, 9007, 9132, 9580 e 9631 em relação ao genoma da amostra ArD1362 (Figura 16).

Os resultados de ambos métodos sugeriram pontos de quebra nas regiões dos genes do Envelope e da polimerase NS5. Em seguida, os possíveis recombinantes foram detectados novamente por um método distinto, utilizando RDP, entre as sequências parciais dos genes. Foi detectado um único evento de recombinação nas sequências parciais do gene E com ponto de quebra estimado nas posições 1299 e 1947 da poliproteína viral, abrangendo as amostras ArA986, HD78788, ArA27101, ArA27290, ArA27096, ArA27443, ArA27407, ArA27106 e ArA982 pelos métodos com valores significativos de p (Bootscan, 1,31E-5; Maxchi, 2,85E-3; Chimaera, 1,59E-3; SiSscan, 1,79E-29; 3Seq, 6,85E-19). Assim como, um único evento de recombinação foi estimado nas sequências parciais do NS5 com ponto de quebra nas posições 9142 e 9654 da poliproteína viral, abrangendo as amostras ArD158084, ArB1362 e ArD157995 pelos métodos com valores significativos de p (Bootscan, 9,93E-9; Maxchi, 3,32E-7; Chimaera, 3,32E-7; SiSscan, 5,27E-28 e 3Seq, 7,65E-24). Estes recombinantes preditos foram excluídos das análises filogenéticas subsequentes para se evitar erros sistemáticos de filogenia (POSADA; CRANDALL, 2002; SCHIERUP; HEIN, 2000). Ademais, a análise da incongruência filogenética entre o conjunto de árvores obtidas do posterior (PST) apontou discordância

entre três nós terminais (ArD128000, ArA1465 e ArD142623) nas árvores analisadas, logo estes três também foram retirados dos concatenados de sequências. Considerando as limitações devido a amostragem restrita do estudo, estes achados apontam que ZIKV pode ter experimentado taxas de recombinação incomuns na natureza para um flavivírus, todavia estudos adicionais são requeridos para se avaliar mais apropriadamente os efeitos sobre espalhamento viral, manutenção do ciclo enzoótico e potenciais repercussões de cunho epidemiológico.

Tabela 1. Detecção de eventos de recombinação em genomas de ZIKV

Início	Fim	Recombinantes	RDP	GENECONV	Bootscan	Maxchi	Chimaera	SiScan	3Seq
3119	4261	ArD142623	5,66E-38	1,34E-37	6,35E-42	2,79E-17	1.69E-16	5,54E-14	NS [#]
93*	805	ArD128000	1,76E-33	5,25E-34	1,42E-33	5,69E-05	3.79E-13	1,37E-14	4,83E-11
1352	1835	ArD157995	2,10E-33	1,07E-28	3,98E-18	1,10E-09	9.02E-10	2,43E-10	3,84E-09
5161	5222	ArD157995	2,39E-28	1,19E-22	NS [#]	1,33E-09	2.39E-09	NS [#]	NS [#]
8271	9002	ArD142623	3,05E-18	5,69E-28	3,25E-21	2,91E-13	4.27E-11	2,22E-21	NS [#]
5096	5555*	ArD128000	3,83E-25	3,76E-23	3,13E-25	6,35E-10	2.82E-10	4,62E-11	2,01E-03
5096	5555*	ArD7117	3,83E-25	3,76E-23	3,13E-25	6,35E-10	2.82E-10	4,62E-11	2,01E-03
3003	3530	ArD128000	1,12E-23	1,32E-21	9,45E-24	1,91E-08	9.25E-09	4,18E-09	2,89E-05
847	1272	ArD158084	1,34E-15	5,94E-13	3,99E-16	1,85E-08	8.28E-07	7,55E-04	NS [#]
847	1272	ArD157995	1,34E-15	5,94E-13	3,99E-16	1,85E-08	8.28E-07	7,55E-04	NS [#]
847	1272	ArD158095	1,34E-15	5,94E-13	3,99E-16	1,85E-08	8.28E-07	7,55E-04	NS [#]
5631*	6498	ArD142623	5,58E-23	2,31E-22	1,12E-24	1,21E-13	3.17E-12	1,99E-19	NS [#]
1*	634	ArD142623	5,71E-24	1,35E-21	1,33E-14	7,65E-05	9.08E-05	2,04E-15	NS [#]
9560	9802	ArD142623	1,77E-12	1,25E-11	4,63E-13	8,56E-05	2.75E-06	NS [#]	NS [#]
9003*	9176	ArD157995	1,77E-11	2,15E-06	2,78E-06	6,08E-04	NS [#]	NS [#]	NS [#]
9003*	9176	ArD158084	1,77E-11	2,15E-06	2,78E-06	6,08E-04	NS [#]	NS [#]	NS [#]
9003*	9176	ArD158095	1,77E-11	2,15E-06	2,78E-06	6,08E-04	NS [#]	NS [#]	NS [#]
7172	7406	ArD142623	2,87E-06	1,68E-09	9,18E-08	1,12E-06	3.02E-06	9,75E-07	NS [#]

*O programa não foi capaz de determinar a posição putativa do ponto de quebra, provavelmente devido à sobreposição de um outro evento de recombinação subsequente.

[#]O valor de p utilizando-se este método não foi significativo para a detecção do evento listado. O método PhylPro foi ocultado da tabela acima por não ter reportado valores de p significativos durante a detecção dos eventos apresentados.

Figura 16. Plotagem da distribuição da probabilidade de mudança de estados calculada por recHMM. As linhas tracejadas em verde fluorescente indicam a posição dos pontos de quebra em relação ao genoma da amostra ArD1362.

Análises filogenéticas

O mapeamento da verossimilhança das sequências de ZIKV sugeriu que os genes da proteína não estrutural 5 e do envelope são mais adequados a reconstrução filogenética (Tabela 2) que a sequência nucleotídica da região 3'NCR. As árvores reconstruídas a partir das sequências de ZIKV (E, NS5 e concatenados desses dois genes) demonstram que existem duas linhagens principais circulando no continente africano (Figura 17), concordando com resultados de trabalhos anteriores que empregaram um menor número de amostras e foram capazes de categorizar as linhagens de ZIKV em dois grupos principais, o clado da cepa protótipo MR766, isolada em Uganda no ano de 1947 e o clado nigeriano (LANCIOTTI et al., 2008). Assim como o agrupamento entre as linhagens asiáticas corrobora a noção que a mais recente epidemia reportada de febre por Zika na Micronésia pode ter sido causada pela expansão de uma antiga linhagem asiática e não pela introdução recente de uma linhagem africana no arquipélago (HADDOW et al., 2012). Ademais, observou-se que os isolados virais de Costa do Marfim e Senegal se distribuem entre os dois grupos principais indicando que estes países experimentaram introduções independentes durante a dispersão deste vírus pelo continente africano (Figura 17).

Tabela 2. Mapeamento de Verossimilhança das regiões genômicas de ZIKV.

Região genômica	Quartetos resolvidos (%)	Quartetos não resolvidos (%)*	Quartetos parcialmente resolvidos (%)
Envelope	93,3	3,2	3,5
NS5	93,5	3,4	3,3
3' NCR	77,7	16,4	6,0

*O percentual de quartetos não resolvidos é um indicador da resolução filogenética dos dados sob análise. Se o percentual é alto, a resolução filogenética é baixa. Portanto, os resultados sugeriram que o gene NS5 é o mais adequado para a inferência filogenética devido à menor incerteza associada, apesar de que não existe diferença substancial entre o percentual de não resolvidos de NS5 e envelope.

Tabela 3. Sítios codificantes negativamente selecionados detectados em ZIKV.

Posição do aa	Ω	p-valor
12	-3,734029	0,00001
62	-5,000000	0,004115
79	-5,000000	0,004951
212	-5,1873	0,0044
212	-5,4860	0,0035
223	-5,1873	0,0044
265	-5,1351	0,0045
350	-5,0000	0,0043

*A posição do aminoácido foi referida ao primeiro códon da sequência do isolado micronésio ECMN2007, cujo número de acesso é EU545988.

†Sítios negativamente selecionados ($\omega < 0$) identificados pela diferença significativa entre as taxas de mutação não sinônima (dN) e sinônima (dS) por sítio utilizando contagem ancestral de única verossimilhança, implementado em HyPhy.

Figura 17. Árvore filogenética inferida a partir de sequências concatenadas de ZIKV usando métodos baseado em Máxima Verossimilhança. Nos nós filogenéticos são mostrados valores de *bootstrap* acima de 70 %, os isolados virais da Costa do Marfim (marcados em laranja) e do Senegal (marcados em verde) são destacados nos dois clados propostos por (LANCIOTTI et al., 2008).

Análise de seleção

Diversos sítios negativamente selecionados foram encontrados (Tabela 3), em concordância com o modo de transmissão, evento que ocorre durante alimentação sanguinolenta de mosquitos, assim como a natureza aguda da infecção, favorecendo o acúmulo de mutações sinônimas e sítios negativamente selecionados (HANADA; SUZUKI; GOJOBORI, 2004). Por outro lado, a alternância entre hospedeiros artrópodes a mamíferos pode impor severos gargalos de seleção purificadora ao

vírus, diminuindo a diversidade genética viral (HOLMES, 2003; JERZAK et al., 2008; KUNO; CHANG, 2005).

Análises filodinâmicas

As médias estimadas e as densidades probabilísticas mais altas no intervalo de 95 % (HPD) das taxas de substituição por sítio por ano estimadas para as regiões genômicas do envelope (E), polimerase (NS5) e não codificante a 3' (3'NCR), foram, respectivamente, $2,135\text{E-}3$ ($2,040\text{E-}3$ a $2,3221\text{E-}3$), $7,1789\text{E-}4$ ($6,9466\text{E-}4$ a $7,4170\text{E-}4$) e $1,1285\text{E-}3$ ($2,7089\text{E-}4$ a $2,5040\text{E-}3$) substituições por sítio por ano. Os resultados sugeriram uma origem recente para as amostras de ZIKV do estudo, próxima ao início do século passado. As datas estimadas para as raízes das árvores filogenéticas inferidas e os respectivos HPDs, a partir das sequências parciais do gene E, NS5 e 3'NCR, foram respectivamente 107 anos (70 a 156 anos), 90 anos (67 a 120 anos), 80 anos (44 a 129 anos). Apesar das diferentes raízes estimadas, observamos haver sobreposição entre os intervalos de confiança. Além disso, assumimos que a variação na diversidade genética no tempo, medida como o tamanho populacional efetivo vezes o tempo de geração ($Ne.g$) estimada com o método de MCMC, correlaciona-se com a demografia de ZIKV, sendo proporcional ao número de infecções no tempo. Os BSP para todas as regiões genômicas estimadas refletem uma flutuação similar no tamanho populacional viral, marcada por um aumento no $Ne.g$ entre 1940 a 1960 e uma diminuição após essa data (Figura 18), até uma aparente estabilidade próximo a 1990 que seria coerente com o ciclo enzoótico que caracteriza esta virose, no qual a população viral é mantida em um nível basal, uma vez que epidemias causadas por ZIKV não são comuns, portanto não se espera observar intensas flutuações no tamanho populacional de ZIKV.

Figura 18. Plotagens de *skyrides* bayesianos (BSP) das duas regiões genômicas de ZIKV. O tamanho populacional efetivo vezes o tempo de geração ($Ne.g$) aproxima o número de infecções no tempo. Os plots indicam complexas oscilações com altos títulos de atividade viral iniciando aproximadamente em 1940 e culminando próximo a 1960, seguido por um decaimento regular. A estabilização próxima ao presente é consistente com a atividade enzoótica, desde que o vírus não é comumente mantido em populações humanas.

Movimentação de ZIKV para a África ocidental

A nossa amostragem nos permitiu inferir a ligação mais provável entre as localidades de isolamento das linhagens de ZIKV. Nossos resultados indicaram que o vírus emergiu em Uganda por volta de 1920, provavelmente entre 1892 e 1943. Este achado é coerente com a primeira descrição de ZIKV em 1947 (DICK; KITCHEN; HADDOW, 1952). Ademais, no contexto da África ocidental nós fomos capazes de encontrar duas introduções independentes de ZIKV para esta porção do continente (Figuras 19 e 20). A primeira destas introduções que passou pela Costa do Marfim (CI) e Senegal (SN) foi relacionada às linhagens do clado MR766 (linhas amarelas na Figura 19), que provavelmente veio de Uganda próximo a 1940 para Dezidougou (CI). Desta localidade, espalhou-se para Kedougou (SN) em 1985 e para Sokala-Sobara (CI) em 1995.

A segunda introdução na África ocidental foi relacionada ao clado nigeriano de ZIKV (linhas verdes na Figura 19), nesta o vírus foi movido inicialmente de Uganda para a República Centro-Africana e Nigéria próximo a 1935. Da Nigéria, o vírus provavelmente foi disperso para Saboya (SN) por volta de 1950 e de lá novamente foi para Dezidougou (CI) e Bandia (SN) próximo a 1960. Deste último lugar, ZIKV foi introduzido em Kedougou (SN) perto de 1965 e de lá foi para Burkina Faso por volta de 1980 e para Dacar (SN) em torno de 1985.

Uma linhagem adicional de ZIKV também originada em Uganda foi dispersada para a Malásia em torno de 1945 e Micronésia por volta de 1960 (linhas pretas na Figura 19), formando clado asiático de ZIKV (HADDOW et al., 2012). Esta última linhagem está relacionada ao espalhamento recente de ZIKV pelo oceano pacífico (CAO-LORMEAU et al., 2014; MUSSO; NILLES; CAO-LORMEAU, 2014).

Adicionalmente, o presente estudo foi capaz de estimar uma data de separação entre as linhagens africanas e asiáticas bem como inferir a origem africana das cepas de ZIKV isoladas na Ásia. A movimentação de ZIKV pelo continente africano e para a Ásia foi sumarizada no mapa da Figura 19 e na árvore filogenética da Figura 20. Os resultados foram suportados pelo desvio significativo da aleatoriedade da distribuição dos caracteres de localidade na amostra de árvores filogenéticas obtida durante a estacionaridade do MCMC (Tabela 4), evidenciados pelos índices PS e AI significativamente diferentes da amostra aleatória que suporta a distribuição global dos estados nas filogenias. Os índices de MC para cada estado informam a respeito do quanto um estado é monofilético nas filogenias, logo apenas os isolados virais de Kedougou e Dezidougou desviam da aleatoriedade quanto à monofilia.

Tabela 4. Significância da ocorrência de estados de localidade em nós terminais das filogenias de ZIKV.

Estatística*	Média observada	Intervalo de confiança (95%)	Média esperada	Intervalo de confiança (95%)	Valor de $p^{\dagger}$
AI	0,93	0,61-1,25	2,98	2,45-3,48	0
PS	13,00	13-13	18,87	17,23-19,96	0
MC estado K	7	7-7	2,86	2,01-4,58	0
MC estado D	6	6-6	1,50	1-2,26	1.00E-4
MC estado L	1	1-1	1	1-1	1
MC estado O	1	1-1	1	1-1	1
MC estado C	1	1-1	1	1-1	1
MC estado S	1	1-1	1	1-1	1
MC estado B	1	1-1	1	1-1	1
MC estado Y	1	1-1	1	1-1	1
MC estado A	1	1-1	1	1-1	1
MC estado U	1	1-1	1,01	1-1,01	1
MC estado M	1	1-1	1	1-1	1

*AI, Índice de associação, PS, valor de parcimônia e MC, tamanho de clado monofilético.

¶Valores de p inferiores 0,05 não são considerados estatisticamente significativos.

Figura 19. Mapa representando o espalhamento geográfico de ZIKV no Continente Africano e para a Ásia. As linhas dirigidas conectam as localidades de partida e destino (evidenciadas pelas setas) das linhagens virais. O tempo estimado para o ancestral comum mais recentes das cepas de diferentes localidades foi exposto nos retângulos e pode ser interpretado como o mais antigo ano possível de introdução da linhagem no local. As siglas no mapa significam UG, Uganda; CF, República Centro-Africana; NG, Nigéria; BF, Burkina Faso; MY, Malásia; FM, Micronésia; KE, Kedougou (Senegal); DA, Dacar (Senegal); BA, Bandia (Senegal); SA, Saboya (Senegal); SS, Sokala-Sobara (Costa do Marfim); DE, Dezidougou (Costa do Marfim).

Figura 20. Árvore filogenética com máxima credibilidade de clados resumando as reconstruções de estados geográficos ao longo de uma árvore escalonada temporariamente. Nos ramos são mostradas as reconstruções de estados ancestrais indicando a mais provável localidade de uma linhagem no tempo ao longo de ramos coloridos de acordo com o estado. Ademais, próximo aos nós são expostos o valor de probabilidade a *posteriori*. As localidades foram codificadas como U (Uganda), C (República Centro-Africana), M (Malásia), Y (Micronésia), L (Burkina Faso), N (Nigéria), K (Kedougou), D (Dezidougou), O (Sokala-Sobara), S (Saboya), B (Bandia) e A (Dacar).

Hospedeiros de ZIKV

As análises de reconstrução de estados ancestrais com hospedeiros das linhagens de ZIKV sugeriram que não existe um padrão simples de distribuição de hospedeiros nas genealogias inferidas (Figura 21). Este achado pode ser coerente com o *status* enzoótico da virose (KUNO; CHANG, 2005), como ocorre com ZIKV na África, onde uma diversa gama mosquito vetores pode ocupar nichos ecológicos semelhantes e compartilhar linhagens virais, inclusive este fato pode ser subsídio para a ocorrência de recombinação uma vez que a coinfeção de um hospedeiro por linhagens virais diferentes é uma condição *sine qua non* para o sucesso de um evento de recombinação.

Apesar de existirem evidências que a distribuição de caracteres de hospedeiros nos nós terminais das genealogias inferidas não é aleatória ($PS < 0,05$), embora o valor de p de AI tenha sugerido que essa distribuição se aproxima de uma obtida por amostragem aleatória (Tabela 5), uma única espécie de hospedeiros apresentou significativa monofilia *Aedes dalzieli*, sugerindo um papel relevante deste vetor zoofílico no oeste da África (DIALLO et al., 1999). Este resultado pode ser informativo para os serviços de vigilância epidemiológica e de saúde pública uma vez que competência vetorial é uma das condições requeridas para a emergência de uma arbovirose (WEAVER; REISEN, 2010).

Tabela 5. Significância da ocorrência de estados de hospedeiros de isolamento em nós terminais das filogenias de ZIKV.

Estatística*	Média observada	Intervalo de Confiança (95%)	Média esperada	Intervalo de Confiança (95%)	Valor de $p^{\dagger}$
AI	3,45	3,04-3,84	3,90	3,44-4,24	0,06
PS	26,05	25-27	28,36	26,27-30,11	0,03
MC estado M	1,00	1-1	1,04	1-1,2	1
MC estado A	1,00	1-1	1,09	1-1,71	1
MC estado O	1,00	1-1	1,02	1-1,05	1
MC estado L	1	1-1	1,14	1-1,92	1
MC estado G	1	1-1	1	1-1	1
MC estado V	1,05	1-1	1,05	1-1,24	1
MC estado E	1	1-1	1,01	1-1,04	1
MC estado F	1,01	1-1	1,22	1-2,00	1
MC estado T	1,00	1-1	1,01	1-1,01	1

MC estado D	6	6-6	1,50	1.00-2,28	1,00E-4
MC estado H	1	1-1	1,04	1-1,24	1

*AI, Índice de associação, PS, valor de parcimônia e MC, tamanho de clado monofilético.

¶Valores de p inferiores 0,05 não são considerados estatisticamente significativos.

Figura 21. Cladograma mostrando a distribuição de caracteres de hospedeiros de isolamento (L, *Ae. luteocephalus*; D, *Ae. dalzieli*; V, *Ae. vittatus*; T, *Ae. taylora*; F, *Ae. furcifer*; A, *Ae. africanus*; E, *Ae. aegypti*; O, *Ae. opock*; G, *Muci graham*; M, macaco; H, humano). Os ramos são coloridos de acordo com o caractere ancestral inferido e próximos aos nós são mostrados valores de probabilidade a posteriori dos nós nas árvores filogenéticas.

Contexto filodinâmico dos eventos de recombinação em ZIKV

Uma vez que tínhamos datas de isolamento das amostras e estimamos uma taxa de evolução do vírus, estimamos a data dos eventos de recombinação (Figura 22). Nossos resultados sugeriram que algumas linhagens de ZIKV amostradas em Dezidougou (CI) em 1990 (ArA27101, ArA27290, ArA27096, ArA27443, ArA27407 e ArA27106) com recombinação na região do envelope (E) compartilharam um ancestral comum por volta de 1962 (1951 – 1967, 95% HPD). Assim como, a linhagem ArA982, também isolada em Dezidougou em 1999, e seu grupo-irmão ArA986, amostrada na província vizinha de Sokala-Sobara em 1999, compartilharam um ancestral comum por volta de 1992 (1981 – 1996, 95% HPD). Diante destes resultados podemos especular que a recombinação no envelope pode ter provido alguma vantagem adaptativa para a manutenção e/ou estabelecimento de ZIKV no ciclo enzoótico que caracteriza esta região da Costa do Marfim. A outra linhagem identificada com E recombinante (HD78788), isolado de um caso humano na capital do Senegal em 1991, compartilhou um ancestral comum por volta de 1984 (1976 – 1988, 95% HPD) com outras linhagens de Kedougou também no Senegal.

Por outro lado, as três linhagens recombinantes no gene da NS5 não formaram clado ao longo da árvore, apesar de dois deles, ArD157995 e ArD158084, terem sido isolados em Kedougou (SN) em mosquitos da espécie *Aedes dalzieli* em 2001. O outro foi isolado em Bouboui na República Centro-Africana em 1968 de *Aedes africanus*. Estes resultados apontaram a recombinação no NS5 não foi um evento ancestral que foi mantido nas linhagens que sucederam, como possivelmente aconteceu no E das linhagens de Dezidougou. A distribuição preferencial das linhagens recombinantes ao longo das filogenias foi suportada por p -valores significativos de AI e $PS \leq 2.00E-4$ (Tabela 6). Ademais, os padrões de adjacência dos recombinantes de E e NS5 foram confirmados também por MC (Tabela 6).

Figura 22. Árvore filogenética com máxima credibilidade de clados resumando as reconstruções de estados de recombinação ao longo de uma árvore escalonada temporariamente. Os ramos foram coloridos de acordo com o estado do recombinante. Ademais, próximo aos nós são expostos o valor de probabilidade *a posteriori*.

Tabela 6. Significância da ocorrência de estados de recombinação em nós terminais das filogenias de ZIKV.

Estatística*	Média observada	Intervalo de Confiança (95%)	Média esperada	Intervalo de Confiança (95%)	Valor de $p^{\dagger}$
AI	0,65	0,23-0,99	1,91	1,36-2,42	2,00E-4
PS	6,78	6,0-7,0	11,19	9,85-11,99	2,00E-4
MC estado NR	6,00	6,0-6,0	4,63	2,93-7,86	0,30
MC estado ER	6,00	6,0-6,0	1,51	1,00-2,29	2,00E-4
MC estado NS5R	1,00	1,00-1,00	1,05	1,0-1,25	1,00

*NR significa não recombinante; ER, recombinante no envelope; e NS5R, recombinante no NS5.

Padrões de glicosilação no envelope de ZIKV

Nossas análises sugeriram diversos sítios de glicosilação na proteína E (Figura 23). Nós detectamos em algumas sequências do E um provável motivo (Asn-X-Thr), que indica a presença de um sítio N-glicosilado no resíduo Asn-154, corroborando os achados de (HADDOW et al., 2012; LANCIOTTI et al., 2008). Este resíduo é localizado em uma α -hélice na estrutura do E (seta amarela na Figura 23A e esfera amarela na Figura 23B). Nossas inferências sobre sítios de glicosilação no E também apontaram vários sítios O-glicosilados (seta vermelhas na Figura 23A e esferas vermelhas na Figura 23B) e nenhum sítio C-manosilado. Dentre os sítios O-glicosilados, encontramos um provável sítio tipo-mucina no resíduo Thr-170 em todas as linhagens de ZIKV estudadas, também encontramos outros sítios deste tipo em outras sequências estudadas nos resíduos Thr-245 e Thr-381 (Figura 23A). Ademais, encontramos prováveis sítios de fixação GlcNAc nos resíduos Ser-142, Ser-227, Thr-231, Ser-304, Thr-366 e Thr-381 de algumas linhagens (Figura 23A).

Dada a importância do sítio N-glicosilado no resíduo Asn-154 da proteína E para infectividade e montagem das partículas virais de flavivírus (HANNA et al., 2005; LEE et al., 2010; MONDOTTE et

al., 2007) e o fato de que nós observamos variações no motivo que define o sítio (deleções e substituições 156 Thr/Iso), nós investigamos a correlação entre a conservação deste motivo (Asn-X-Thr) e filogenias de ZIKV. Nossos resultados sugeriram que a aquisição deste sítio de glicosilação foi um evento recorrente na história de ZIKV, baseados nas mudanças observadas de Isoleucina para Treonina e vice-versa na árvore da Figura 24. Nossas inferências a respeito da distribuição deste polimorfismo nas filogenias virais são baseadas em *p*-valores de AI e PS $\leq 7.00E-4$ (Tabela 7). No entanto, nossas conclusões são limitadas pelas passagens seriais em cérebro de camundongo das cepas mais antigas de ZIKV (HADDOW et al., 2012), que podem ter resultado na perda deste sítio de glicosilação como se observa no vírus do Oeste do Nilo (CHAMBERS et al., 1998).

Tabela 7. Significância da ocorrência de estados de sítios de glicosilação em nós terminais das filogenias de ZIKV.

Estatística*	Média observada	Intervalo de Confiança (95%)	Média esperada	Intervalo de confiança (95%)	Valor de <i>p</i>[¶]
AI	0,89	0,49-1,28	2,45	1,81-3,06	0
PS	10,56	10-11	15,89	13,49-18,02	7,00E-4
MC estado I	6	6-2	2,85	2,00-4,62	0,03
MC estado T	3,89	2-6	2,33	1,49-3,36	0,02
MC estado A	1,00	1-1	1,09	1-1,66	1

*AI, Índice de associação, PS, valor de parcimônia e MC, tamanho de clado monofilético. I para Isoleucina; T, Treonina; e A, ausente.

[¶]Valores de *p* inferiores 0,05 não são considerados estatisticamente significativos.

Figura 24. Cladograma mostrando a distribuição de caracteres de sítios de glicosilação (I, Isoleucina; T, Treonina; A, ausente). Os ramos são coloridos de acordo com o caractere ancestral inferido e próximos aos nós são mostrados valores de probabilidade *a posteriori* dos nós nas árvores filogenéticas.

Mudança evolucionária correlacionada ao longo das filogenias de ZIKV

Uma vez que foi previamente demonstrado que a ausência do sítio N-glicosilado no Envelope aumenta a infectividade viral em células de *Aedes albopictus* (HANNA et al., 2005; LEE et al., 2010) e o gene para o envelope das linhagens de ZIKV isoladas em *A. dalzieli* – o único vetor com MC significativo – apresentaram ausência deste sítio, investigou-se a correlação entre essas duas características. Crucialmente, o modelo em que ausência do sítio N-glicosilado e colonização de *Aedes dalzieli* evoluíram de maneira correlacionada foi favorecido (BF = 47,004). Ademais, estes achados podem estar correlacionados ao ciclo enzoótico do ZIKV no oeste da África e o comportamento zoofílico do *A. dalzieli* (DIALLO et al., 1999), cujas fêmeas fazem repasto sanguíneo de um amplo espectro de hospedeiros.

Implicações biológicas de nossos achados sobre ZIKV

Nossas análises apontaram que ZIKV pode ter experimentado diversos eventos de recombinação, fato que é incomum entre flavivírus (SIMON-LORIERE; HOLMES, 2011). Os eventos recorrentes de perda e ganho do sítio de N-glicosilado na proteína E podem estar relacionados à infectividade em células de mosquitos (LEE et al., 2010), assim como a correlação significativa entre a ausência deste sítio de glicosilação em linhagens de ZIKV de *A. dalzieli* fornece evidências indiretas sobre a adaptação de algumas linhagens deste vírus nesta espécie de mosquitos que pode fazer a transferência de ZIKV do ambiente silvestre para o peridoméstico, uma vez que esta espécie de mosquito costuma se alimentar de gado (DIALLO et al., 1999). Em suma, nossos resultados corroboram a ideia que três linhagens de ZIKV se originaram no leste da África, provavelmente em Uganda, e migraram para o oeste da África e para a Ásia.

Vírus da Encefalite Transmitida por Carrapatos (TBEV)

Os resultados descritos abaixo são parte do artigo “Molecular phylogeography of tick-borne encephalitis virus in Central Europe” publicado no Journal of General Virology (doi: 10.1099/vir.0.054478-0) que segue no Anexo II.

Análises filodinâmicas

Objetivando obter um melhor entendimento sobre os movimentos das linhagens de TBEV na Europa central, investigamos sequências de Envelope de 102 isolados, representando distintos *foci* na região. Dois esquemas alternativos foram comparados para se estimar mais apropriadamente a taxa de substituição (μ) expressa em substituições por sítio por ano (s/s/a). O teste de Fatores de Bayes (BF) indicou que uma taxa menor ($4,07\text{E-}5$ s/s/a) teve um BF levemente maior (1,635) que a proposição com uma taxa maior ($1,69\text{E-}4$ s/s/a), sugerindo que estes vírus estão circulando na região a aproximadamente 350 anos (Figura 25). Ademais, os resultados apontaram que as linhagens de TBEV se dividem em dois grupos principais, setores verde e amarelo na Figura 25 e experimentaram uma forte seleção purificadora no gene E (Tabela 8).

Tabela 8. Sítios codificantes sob seleção purificadora no E de TBEV da Europa central

Posição do aminoácido*	ω	p-valor
283	-4.80	7.09E-03
294	-4.08	1.24E-02
298	-5.00	4.12E-03
311	-3.00	3.70E-02
334	-4.00	1.23E-02
348	-3.00	3.70E-02
372	-4.71	7.09E-03
417	-9.13	3.65E-02
421	-3.00	3.70E-02
434	-3.53	2.27E-02
447	-4.00	1.23E-02
455	-3.00	3.70E-02
477	-3.00	3.70E-02
500	-4.71	6.44E-03
516	-5.88	1.82E-03
533	-3.53	2.27E-02
551	-3.00	3.70E-02
582	-5.00	4.12E-03
592	-4.71	6.44E-03
603	-3.53	2.27E-02
604	-10.59	1.20E-05
616	-3.33	2.82E-02

618	-3.53	2.44E-02
660	-3.53	2.27E-02
676	-3.00	3.71E-02
683	-3.53	2.27E-02
718	-10.70	1.20E-05
719	-4.04	1.23E-02
732	-3.03	3.70E-02
735	-3.82	3.39E-02
763	-3.03	3.70E-02

*A posição do aminoácido é relative a primeira posição do códon da sequência da poliproteína do isolado Toro-2003, cujo número de acesso no GenBank é ABD62793.2.

[†]Diferença entre os sítios não-sinônimos (*dN*) pelos (*dS*) sinônimos por sítio.

Filogeografia

Para investigar padrões de espalhamento, testamos diferentes modelos de espalhamento, um que pressupõe uma difusão contínua (LEMEY et al., 2010) e outro que o espalhamento ocorre por eventos discretos de migração (LEMEY et al., 2009). A difusão contínua foi testada sob duas condições, taxas de difusão dentre localidades homogêneas (difusão browniana) ou se as taxas seguem uma distribuição de Cauchy. O cenário com saltos discretos entre localidades foi testado usando-se uma cadeia de Markov tempo contínua (CTMC) e um caso especial desse modelo: seleção variável com busca estocástica bayesiana (BSSVS). Os BF estimados para os modelos discretos foram significativamente maiores (Tabela 9).

Tabela 9. Fatores de Bayes estimados entre os diferentes modelos testados nas inferências filogeográficas.

	lnP (modelo/dado)	S.E.	BSSVS	CTMC	Browniano	Cauchy
BSSVS	-5408.819	+/- 0.402	0	-1.062	11.679	11.478
CTMC	-5407.758	+/- 0.348	1.062	0	12.74	12.539
Browniano	-5420.498	+/- 0.999	-11.679	-12.74	0	-0.201
Cauchy	-5420.297	+/- 0.879	-11.478	-12.539	0.201	0

Ademais, o teste de correlação de Mantel entre as matrizes de distância patristica e geográfica indicou que a relação entre elas é significativa ($p = 9.9999E-6$). O correlograma de Mantel não apresentou homogeneidade em todas as distâncias geográficas (Figura 26). Foi encontrada correlação significativamente positiva somente em distâncias inferiores a 100 quilômetros (Km). No entanto, a correlação foi significativamente negativa em distâncias entre 100 e 200 Km e novamente próxima aos

300 Km. Estes achados sugerem que os *foci* de TBEV parecem ser coesos em um raio inferior a 100 Km, aparentemente vírus não consegue se difundir continuamente para distâncias maiores. Em suma, os resultados indicaram dois modos distintos de dispersão de TBEV: (i) em uma escala local, TBEV pode ser dispersado continuamente e (ii) em uma escala regional o espalhamento de TBEV pode ser explicado por eventos de migração “em saltos”, devido a correlação negativa entre distâncias genéticas e geográficas.

As análises de migração indicaram o espalhamento de TBEV da Alemanha para outras áreas na Europa no início do século XVIII e que o vírus foi trazido de volta no século XX (Figura 27). Como recentemente confirmado, a movimentação do TBEV tende a seguir uma distribuição clinal orientado primariamente do leste para o oeste (HEINZE; GOULD; FORRESTER, 2012; ZANOTTO et al., 1995). No entanto, nós encontramos evidências de fluxo migratório em ambas direções, o que pode ter sido possível porque empregamos uma análise mais detalhada, em uma escala geográfica numérica maior, novos padrões podem emergir. Considerando, a história complexa deste vírus na Europa central onde comércio, tráfico de pessoas e animais, desenvolvimento econômico e guerras ocorrem a séculos, estas migrações podem muito bem ser entendidas em um contexto local de atividade antrópica.

Figura 25. Árvore filogenética com máxima credibilidade de clados sumarizando as reconstruções de estados geográficos ao longo de uma árvore escalonada temporariamente. Os ramos estão coloridos de acordo com o estado ancestral inferido. Os valores de probabilidade *a posteriori* são mostrados em negro próximo aos nós e os valores de probabilidade do estado inferido são mostrados coloridos próximos aos ramos.

Figura 26. Correlograma de Mantel mostrando as relações entre as matrizes de distância patrística e geográfica. No eixo vertical, encontra-se o coeficiente de correlação de Spearman (ρ). No eixo horizontal, as distâncias geográficas. A linha tracejada representa ausência de correlação ($\rho \leq 0$), os pontos acima dela correlação positiva e os abaixo correlação negativa, preenchidos em negro quando significantes ($p \leq 0,05$).

Figura 27. Movimentação detalhada do TBEV na Europa central desde o século XVIII. Setas representam direção das migrações. Os lugares de isolamento foram mostrados pelas letras maiúsculas: G para Alemanha; A, Áustria; K, Eslováquia; N, Boêmia do norte; C, Boêmia central; S, Boêmia do sul; B, Morávia do sul e N para Morávia do norte.

Vírus da Febre da Grande Fenda Africana

O nosso primeiro trabalho em parceria com o grupo do Instituto Pasteur sobre este agente viral foi publicado no jornal PLoS ONE em 2012 (doi: 10.1371/journal.pone.0035216), constituindo uma parte substancial dos achados apresentados na seção subsequente e segue no anexo III. Ainda a continuação dessa colaboração possibilitou uma outra publicação no jornal Emerging Infectious Diseases (doi: 10.3201/eid2002.130996), cujos achados também estão apresentados abaixo e no anexo IV. Ademais, descrevemos abaixo a terceira parte deste trabalho sobre as dinâmicas diferentes dos três segmentos virais, atualmente em revisão no periódico PLoS Neglected Tropical Diseases, cujo manuscrito segue no anexo V.

Deteção de recombinação homóloga

Após nossas análises exaustivas em busca de recombinação intrassegmento com o programa RDP, não encontramos evidências deste evento de mudança genética em ampla escala, concordando com estudos anteriores (BIRD et al., 2007; PEPIN et al., 2010), os quais advogam haver poucas evidências da ocorrência deste evento entre linhagens de Vírus da Febre da Grande Fenda Africana (VFGFA).

Análises filogenéticas

O conteúdo de sinal filogenético das sequências de VFGFA foi investigado após se reconstruir 50 mil quartetos para cada segmento genômico utilizando-se o método de mapeamento de verossimilhança implementado no programa TREE-PUZZLE. Os resultados (Tabela 10) apontaram que o segmento médio (M) tem o maior conteúdo de quartetos bem resolvidos filogeneticamente, sugerindo que este segmento é mais informativo do que os outros dois. As árvores de verossimilhança de RVFV, assim como as árvores MCC, apontaram que as linhagens virais isoladas no Senegal e na Mauritânia se organizaram em distintos grupos nas topologias, agrupando-se preferencialmente com isolados do leste da África, assim como previamente foi observado na epidemia no território da Mauritânia em 2003 (FAYE et al., 2007), sugerindo que estes países experimentaram múltiplas introduções durante o século XX. Ademais, o padrão recorrente de agrupamento entre as linhagens virais destes países como grupos irmãos sugeriu que RVFV não experimentou limitações para ser disperso através das fronteiras destes países (Figuras 28, 29 e 30).

Tabela 10. Mapeamento de Verossimilhança dos segmentos genômicos de RVFV.

Segmento genômico	Quartetos resolvidos (%)	Quartetos não resolvidos (%)*	Quartetos parcialmente resolvidos (%)
Longo	85,5	10,1	4,4
Médio	89,6	7,5	2,9
Curto	84,6	11,6	3,8

*O percentual de quartetos não resolvidos é um indicador da resolução filogenética dos dados sob análise. Se o percentual é alto, a resolução filogenética é baixa.

Figura 28. Árvore filogenética com máxima credibilidade de clados (MCC) inferida a partir das sequências do segmento curto (C) de RVFV. Relações entre 167 cepas de RVFV isoladas de diferentes localidades e países. Amostras de Maurîtânia e Senegal são destacadas por círculos amarelos e verdes respectivamente. O agrupamento recorrente das cepas destes países sugere introduções múltiplas de RVFV nestes países durante o século XX. Valores de probabilidade *a posteriori* maiores que 90 % são mostrados próximos aos nós da árvore.

Figura 29. Árvore filogenética com máxima credibilidade de clados (MCC) inferida a partir das sequências do segmento médio (M) de RVFV. Relações entre 128 cepas de RVFV isoladas de diferentes localidades e países. Amostras de Mauritânia e Senegal são destacadas por círculos amarelos e verdes respectivamente. O agrupamento recorrente das cepas destes países sugere introduções múltiplas de RVFV nestes países durante o século XX. Valores de probabilidade *a posteriori* maiores que 90 % são mostrados próximos aos nós da árvore.

Figura 30. Árvore filogenética com máxima credibilidade de clados (MCC) inferida a partir das sequências do segmento longo (L) de RVFV. Relações entre 126 cepas de RVFV isoladas de diferentes localidades e países. Amostras de Mauritània e Senegal são destacadas por círculos amarelos e verdes respectivamente. O agrupamento recorrente das cepas destes países sugere introduções múltiplas de RVFV nestes países durante o século XX. Valores de probabilidade *a posteriori* maiores que 90 % são mostrados próximos aos nós da árvore.

Análises de seleção

Os resultados obtidos com os três segmentos genômicos apontaram a existência de diversos sítios sob forte seleção negativa, indicado por $\omega < 0$, (Tabela 12) sugerindo que polimorfismos deletérios em genes funcionalmente importantes possam ter sido removidos. Por outro lado, a ausência de sítios positivamente selecionados indicados por $\omega > 0$, expôs a inexistência de seleção direcional, o que caracteriza fenótipos bem adaptados (DICKERSON, 1971; NEI; SUZUKI; NOZAWA, 2010). Estes achados corroboram a existência de diversos gargalos de transmissão impostos pelo ciclo natural do vírus cujo modo de infecção aguda e transmissão por vetor favorecem o acúmulo de mutações sinônimas (HANADA; SUZUKI; GOJOBORI, 2004). Ademais, a alternância entre um diverso conjunto de artrópodes vetores e hospedeiros mamíferos mantém pressões seletivas requeridas para a estabilidade genômica, como a integridade da proteína não estrutural, codificada pelo segmento S, que aparentemente exige que RVFV infecte alternadamente células de artrópodes e mamíferos (MOUTAILLER et al., 2011).

Tabela 12. Sítios codificantes negativamente selecionados detectados em RVFV.

Segmento genômico	Posição do aminoácido*	$\omega^{\dagger}$	<i>p</i> -valor
Longo	559	-5,0648	0,0085
Longo	860	-5,0648	0,0085
Longo	956	-5,0648	0,0085
Longo	1091	-6,3310	0,0030
Longo	1332	-5,0648	0,0085
Longo	1870	-5,0648	0,0085
Médio	64	-5,1841	0,0093
Médio	172	-54,8731	0,0072
Médio	225	-6,9101	0,0021
Médio	282	-4,5937	0,0071
Médio	439	-5,0000	0,0041
Médio	971	-5,5281	0,0071
Médio	1115	-11,6919	0,0047
Curto	116	-5,0000	0,0047
Curto	124	-4,5667	0,0073
Curto	127	-8,0000	0,0002
Curto	148	-4,5667	0,0073
Curto	151	-5,9157	0,0048

*A posição do aminoácido foi referida ao primeiro códon da sequência do isolado mauritaniano OS8. O número de acesso do segmento longo, médio e curto é, respectivamente, DQ375395.1, DQ380185.1 e DQ380177.1.

[†]Sítios negativamente selecionados ($\omega < 0$) identificados pela diferença significativa entre as taxas de mutação não sinônima (*dN*) e sinônima (*dS*) por sítio utilizando contagem ancestral de única verossimilhança, implementado em HyPhy.

Análises filodinâmicas

As médias estimadas e as densidades probabilísticas mais altas no intervalo de 95 % das taxas de substituição por sítio por ano estimadas para os segmentos longo, médio e curto foram, respectivamente, 1,351E-4 (1,343E-4 a 1,389E-4), 2,336E-4 (2,073E-4 a 2,065E-4) e 2,106E-4 (1,867E-4 a 2,352E-4) substituições por sítio por ano. Assumimos que a variação na diversidade genética no tempo, medida como o tamanho populacional efetivo vezes o tempo de geração ($Ne.g$) estimada com o método de MCMC, correlaciona-se com a demografia de RVFV, sendo proporcional ao número aproximado de infecções no tempo. Os BSP para todos os segmentos genômicos do vírus apresentaram um aumento no $Ne.g$ na primeira metade do século passado seguido por um contínuo decaimento na segunda metade do século (Figura 31). Ademais, os BSP concordam com uma origem recente de RVFV (BIRD et al., 2007, 2008), e intensa atividade viral na primeira metade do século XX (PEPIN et al., 2010), seguido por uma ampla dispersão no continente africano. Ademais, a redução no tamanho da população viral ou dos eventos de infecção são consistentes com a atividade enzoótica próxima ao presente. Estes achados corroboram o espalhamento de RVFV entre populações amplamente susceptíveis seguindo a introdução do gado no leste do continente africano primariamente no Quênia durante a primeira metade do século XX para abastecer a atividade pecuária (JOHNSON, 1993; PEPIN et al., 2010). Em suma, os elevados níveis de $Ne.g$ no passado concordaram com o fato que RVFV experimentou intensa atividade no século passado causando intensos surtos e sugeriram

uma tendência recente a atividade enzoótica.

Figura 31. Gráficos de *skyrides* bayesianos (BSP) dos três segmentos genômicos. O tamanho populacional efetivo vezes o tempo de geração ($Ne.g$) aproxima o número de infecções no tempo. As curvas mostram complexas oscilações com alta atividade viral iniciando aproximadamente em 1930 e culminando entre 1940 e 1960, seguido por um decaimento regular a partir de 1970 até o presente. A estabilização próxima ao presente é consistente com a atividade enzoótica, uma vez que o vírus não é comumente mantido em populações humanas.

Reconstrução de localidades ancestrais

Apesar de as reconstruções filogeográficas para os três segmentos genômicos terem sido coerentes entre si (congruência dada por similaridade de padrões de adjacência topológicas), optamos por apresentar somente as feitas com o segmento S por estas apresentarem valores superiores de probabilidade de existência do estado no nó filogenético inferido nas árvores MCC, portanto escolhemos representar a história mais provável de movimento de RVFV na África. As reconstruções filogeográficas revelaram um padrão complexo e movimentação viral através de grandes distâncias no continente africano (Figura 32). As datas nos retângulos nos mapas indicam o TMRCA estimado entre a linhagem da localidade em que a seta de inicia e a linhagem da localidade para qual a seta aponta, assumindo-se que o ano estimado é o mais remoto possível para a introdução daquela linhagem em uma determinada localidade. Observou-se que o espalhamento de RVFV pode ter acontecido gradualmente de uma localidade para outra quando se observa as inferências feitas para a região do Oeste da África (Figura 33), onde a amostragem mais densa permite um detalhamento mais refinado. Os resultados apontaram no mínimo cinco introduções distintas no Senegal e na Mauritânia durante o século passado (Figura 32). A primeira entrada de RVFV estimada na região (linha azul na figura 32 e 33) foi relacionada a cepas de Zimbábue, África do Sul e Uganda possivelmente a aproximadamente 80 anos atrás. Estas inferências são consistentes como inquéritos sorológicos feitos pelo Instituto Pasteur de Dacar que sugere que a zona litorânea do oeste africano pode ter conhecido RVFV durante a primeira metade do século XX (comunicação pessoal feita pelo nosso colaborador, Dr Amadou Alpha Sall). A segunda introdução na África ocidental (linha preta nas Figuras 32 e 33) foi associada com isolados virais da África do Sul Zimbábue compartilhando um ancestral comum cerca de 72 anos atrás. Esta introdução foi relacionada com amostras de Diawara (SN) e Hodh El Garbi (MR), relacionadas com o surto reportado em Diawara em 1998 (MARRAMA et al., 2005). A terceira introdução na região estudada (linha verde nas Figuras 14 e 15) consistiu de linhagens que se originaram em Zimbábue no início do século passado, então moveram-se para a República Centro-Africana aproximadamente 70 anos atrás e de lá chegaram ao Senegal e Mauritânia cerca de 47 anos atrás e na Guiné cerca de 38 anos atrás. Amostras dessa introdução foram isoladas em Barkedji (SN) em 1993 e em Rosso (MR) em 1987 e na Guiné em 1981 e 1984. A quarta introdução (linha preta sólida nas Figuras 32 e 33) foi evidenciada por uma única linhagem isolada em Kedougou em 1983 associada com uma antiga linhagem sul-africana que foi movida para o Egito a aproximadamente 47 anos atrás e de Madagascar a aproximadamente 40 anos. A introdução mais recente (quinta introdução, linha vermelha nas Figuras 32 e 33) foi devido a uma onda de RVFV que se espalhou para lugares distantes na África, foi relacionada a uma linhagem do Sul da África que se moveu 47 anos atrás de Zimbábue e Quênia, em seguida alcançando Madagascar, Somália, Tanzânia, Angola, Senegal, Mauritânia e Arábia Saudita

durante os últimos 28 anos. Um dos aspectos mais interessantes desta onda de RVFV foi que o vírus recorrentemente reemergiu no Quênia de onde ele foi dissipado para outras localidades.

A história das introduções, mudanças e espalhamento de RVFV nos territórios de Senegal e Mauritânia é mostrada com maior detalhamento na Figura 33. O mapa na Figura 15 sumariza reconstruções filogeográficas independentes. De acordo com os resultados, a primeira introdução possivelmente se originou no sul da África, ainda na década de 1930 aportou em Barkedji (SN). Subsequentemente, foi movida para Rosso (MR) por volta de 1970 e Kolda (SN) por volta de 1980. A linhagem de Rosso foi então dispersada para as províncias de Garack (MR), Nkick (MR) e Keur Macene (MR) ainda na década de 80. Mais tarde em 1985, uma linhagem de Nkick em seguida foi movida para Terg no norte da Mauritânia. A linhagem de Kolda (SN) no fim da década de 1980 foi movida para o leste da Mauritânia, aportando em Ayoun El Atross, deste para Guimi (MR) próximo a 2000. A segunda introdução, possivelmente da África central, alcançou Diawara (SN) próximo a 1940. Então, moveu-se para Hodh El Garbi na Mauritânia próximo a 1970. Assim como a primeira (linha azul na Figura 33) e terceira (linha verde na Figura 3) introdução, a quinta (linha vermelha na Figura 33) também aparentemente entrou na região por Bakedji e então foi dispersada para Hseytine (MR) e Lazaret (MR). Em seguida, foi espalhado de Hseytine para Kiffa (MR) a de Lazaret para Matar Lajam e Dar Naim e do último para Tijiga durante a última década. A descrição do espalhamento da quarta e quinta introduções foi exposta no parágrafo anterior.

Figura 32. Representação em larga escala do espalhamento geográfico de RVFV na África e Arábia Saudita. As linhas dirigidas conectam as localidades de partida e de destino (evidenciadas pelas setas) das linhagens virais. As introduções distintas em Senegal e Mauritânia foram representadas por cores diferentes. O tempo estimado para o ancestral comum mais recentes das cepas de diferentes países foi exposto nos retângulos e pode ser interpretado como o mais antigo ano possível de introdução da linhagem na localidade.

Figura 33. Espalhamento geográfico de RVFV no interior de Mauritânia e Senegal. As linhas dirigidas conectam as localidades de partida e de destino (evidenciadas pelas setas) das linhagens virais. As introduções distintas em Senegal e Mauritânia foram representadas por cores diferentes. O tempo estimado para o ancestral comum mais recente das cepas de diferentes localidades foi exposto nos retângulos e pode ser interpretado como o mais antigo ano possível de introdução da linhagem no local.

Uma epidemia na Mauritânia pela reemergência de uma antiga linhagem

Durante a realização de nossa pesquisa uma epidemia de RVFV foi deflagrada na Mauritânia no fim de 2010 (EL MAMY et al., 2011), os nossos colaboradores foram a campo, coletaram amostras e sequenciaram porções dos segmentos M e S. Nossas análises filogenéticas foram capazes de revelar que esta epidemia foi causada pela reemergência de uma antiga linhagem (Figura 34) (FAYE et al., 2014), segunda introdução nos resultados descritos acima e em (SOUMARÉ et al., 2012), que já circulava no oeste da África desde a década de 1940 (Figura 33). Esta porção do trabalho tem um elevado valor agregado, uma vez que contribuiu para esclarecer as origens de uma epidemia corrente.

Figura 34. Árvores de máxima verossimilhança mostrando as relações filogenéticas das linhagens amostradas durante a epidemia de RVFV na Mauritânia em 2010. A) Árvore inferida para o segmento curto. B) Árvore inferida para o segmento médio. As linhagens da epidemia estudada estão marcadas por círculos laranja.

Rearranjos genômicos de RVFV

Rearranjos genômicos (*reassortment* em inglês) podem ser um meio importante de diversificação genética para vírus de genoma RNA segmentado em geral e para os bunyavírus em particular, inclusive possibilitando o surgimento de vírus patogênicos, a partir de outros não patogênicos (BOWEN et al., 2001). A ocorrência e viabilidade destes eventos foi demonstrada tanto em condições de laboratório (TURELL et al., 1990), assim como na natureza (SALL et al., 1999).

Ademais, interações intermoleculares específicas entre os três segmentos genômicos virais parecem ser necessárias para o correto empacotamento dos segmentos para a geração de vírions infecciosos (TERASAKI et al., 2011).

Padrões de adjacência em filogenias indicam reagrupamento de segmentos

Para se comparar as adjacências filogenéticas entre as linhagens de RVFV inferidas a partir de cada segmento, nós analisamos os conjuntos de árvores filogenéticas geradas com MrBayes e conseguimos identificar duas linhagens com histórias filogenéticas discordantes para os segmentos L, M e S. Nossos resultados confirmaram achados prévios que afirmavam que o isolado 2000700608 de uma vaca no Quênia foi formado por reagrupamento do segmento M (BIRD et al., 2008) (Tabela 12 e nomes verdes na Figura 35). Também revelamos um novo evento de reagrupamento em uma linhagem tanzaniana TANTan00107 (Tabela 12 e nomes vermelhos na Figura 35), isolada de um material humano. Desde que sequências dos segmentos de TANTan00107 foram utilizadas concatenadas (NDERITU et al., 2011), este último achado demonstrou a importância de se estudar os segmentos em separado para prevenir erros sistemáticos durante inferências filogenéticas.

Uma vez que nós não tínhamos sequências dos três segmentos para todas as amostras de RVDV, investigamos a ocorrência de reagrupamento entre pares de segmentos (L e M; L e S; M e S), usando conjuntos de com mais *taxa* que os da seção anterior. Esta nova etapa de análise nos possibilitou identificar cinco linhagens com história filogenética discrepantes entre os segmentos M e S (Figura 35 e Tabela 12). Três destas linhagens (76370, SA75 e 3574, nomes laranja, roxo e azul na Figura 35B), isoladas no Zimbábue e África do Sul (Tabela 12), formam um grupo *reassortant* bem suportado que reaparece em filogenias independentes para os segmentos M e S. As outras duas linhagens *reassortants* foram isoladas na Mauritânia (211HMMRRO1987 e 11ANMMRHG1998; nomes rosa e ciano na Figura 35B) e foram isolados em ovelha e humanos (Tabela 13).

Figura 35. Eventos de reagrupamento de segmentos genômicos descobertos pelas incongruências entre as filogenias para os segmentos de RVFV. A) Reagrupamento entre os segmentos L e M. B) Reagrupamento entre os segmentos M e S. Somente os nomes dos *taxa reassortants* são mostrados e coloridos para facilitar a identificação. A ordem dos *taxa* é a mesma entre cada par de árvores.

Tabela 13. Linhagens de RVFV com história filogenética discrepantes

Linhagem	País	Ano	Hospedeiro	Nº de acesso L	Nº de acesso M	Nº de acesso S
2007000608	Quênia	2007	Gado	EU574023	EU574049	EU574079
211HMMRRO1987	Mauritânia	1987	Humano	Indisponível	JN995301	JN995253
11ANMMRHG1998	Mauritânia	1998	Ovelha	Indisponível	JN995312	JN995264
SA75	África do Sul	1975	Humano	DQ375428	DQ380189	DQ380175
3574	África do Sul	1974	Ovelha	JF784386	JF784387	JF784388
TANTan00107	Tanzânia	2007	Humano	HM586959	HM586970	HM586981
76370	Zimbábue	1970	Gado	DQ375426	DQ380188	DQ380165

Distâncias patrísticas sugerem diferentes escalas temporais para cada segmento

Uma vez que o genoma segmentado permite separar histórias evolucionárias por segmento (CARPI; HOLMES; KITCHEN, 2010; RAMBAUT et al., 2008), nós analisamos a quantidade de divergência genética por segmento pela estimativa das distâncias patrísticas das genealogias virais (Figura 36), que podem mudar mais do que as distâncias genéticas, portanto mais sensíveis em revelar discordâncias (FOURMENT; GIBBS, 2006). Estas análises mostraram que existem diferenças significativas entre os segmentos ($\chi^2 = 136,21$ g.l.=2 e p -valor $< 2,2E-16$ obtidos em um teste de Kruskal-Wallis). Ademais, as comparações entre os pares mostraram que as distâncias patrísticas são significativamente diferentes entre eles, pois os valores de diferença (689,73; 513,84 e 1203.57 para L e M; L e S; M e S, respectivamente) obtidos do teste de Kruskal-Wallis com comparação múltipla foram superiores ao valor crítico de 247,76. Estes resultados indicaram que os segmentos de RVFV tem diferentes quantidades de mudanças genéticas, sugerindo a existência de escalas temporais distintas para os segmentos.

Figura 36. Distribuição das distâncias patrísticas por segmento de RVFV. Distâncias patrísticas estimadas a partir da melhor árvore de máxima verossimilhança para cada segmento. As diferenças nas distribuições de picos nas frequências de distâncias patrísticas corroboram a noção de que os segmentos de RVFV estão em escalas temporais diferentes.

Comparação dos regimes seletivos entre os segmentos

Diferenças no regime de seleção entre segmentos pode explicar as diferenças observadas nas distâncias patrísticas. Portanto, nós investigamos novamente a seleção agindo no nível do códon sobre cada segmento, para se comparar a intensidade da seleção sobre cada segmento nós selecionamos nesta seção apenas as linhagens, cujas sequências para todos os genes estavam disponíveis. Aplicando um nível de confiança (α) de 5%, encontramos apenas sítios sob seleção purificadora ($\omega < 0$) em todos os genes analisados (Tabela 14). As comparações entre as quantidades de sítios sob seleção em cada segmento, 98 de 2066, 58 de 1197 e 20 de 510 para L, M e S, respectivamente, mostraram que não existe diferenças significativas entre as proporções (p -valor=0,74 em um teste exato de Fisher com $\alpha=0,05$). Ademais, nós comparamos a intensidade da seleção sobre cada segmento com os ω estimados para os sítios de cada um deles, aplicando um teste de Kruskal-Wallis, e não encontramos diferenças

significativas ($\chi^2 = 3,92$, g.l. = 2 and p -valor = 0,14). Portanto, nossos resultados não suportam diferenças aparentes nos regimes de seleção atuantes sobre os segmentos de RVFV. Nós também estimamos o D de Tajima para cada segmento e encontramos que eles são significativamente negativos e similares para todos os segmentos (L: D= -3,19, $p = 0,0013$; M: D= -3,18, $p = 0,0014$; S: D= -3,32, $p=0,0009$). Estes valores negativos de D de Tajima não mudaram após excluir os *reassortants* das análises e também são indicadores de seleção negativa. Uma vez que os regimes seletivos são similares entre os segmentos e as distâncias patrísticas são significativamente diferentes, nós concluímos que os segmentos de RVFV estão evoluindo em escalas temporais diferentes.

Tabela 14. Sítios codificantes detectados sob seleção purificadora com significância estatística em RVFV.

Posição na proteína	dN - dS	p -valor	Gene
27	-11.695148	0.029228	Polimerase
47	-3.360744	0.028409	Polimerase
77	-3.360744	0.028409	Polimerase
90	-3.076402	0.037037	Polimerase
105	-3.076402	0.038415	Polimerase
119	-4.10187	0.012346	Polimerase
146	-4.10187	0.012346	Polimerase
165	-3.223782	0.033645	Polimerase
168	-4.976072	0.00478	Polimerase
178	-3.360744	0.028409	Polimerase
201	-3.076402	0.037037	Polimerase
230	-3.360744	0.028409	Polimerase
253	-3.076402	0.037037	Polimerase
307	-3.360744	0.030955	Polimerase
308	-3.360744	0.028409	Polimerase
337	-4.10187	0.012346	Polimerase
355	-4.471884	0.009213	Polimerase
401	-5.601241	0.002645	Polimerase
416	-3.076402	0.037037	Polimerase
472	-3.201003	0.032878	Polimerase
484	-10.352989	0.033017	Polimerase

502	-3.076402	0.037037	Polimerase
533	-4.480992	0.008669	Polimerase
536	-3.076402	0.037037	Polimerase
538	-3.360744	0.028409	Polimerase
572	-3.360744	0.030955	Polimerase
588	-3.340391	0.030099	Polimerase
612	-3.076402	0.037304	Polimerase
697	-3.360744	0.028409	Polimerase
722	-3.360744	0.028409	Polimerase
745	-4.10187	0.013252	Polimerase
793	-5.127337	0.004115	Polimerase
799	-3.360744	0.028409	Polimerase
805	-3.360744	0.028409	Polimerase
834	-5.601241	0.002645	Polimerase
869	-3.076402	0.037037	Polimerase
885	-3.360744	0.029748	Polimerase
887	-3.360744	0.028409	Polimerase
930	-4.480992	0.008669	Polimerase
933	-3.360744	0.029748	Polimerase
961	-4.10187	0.012346	Polimerase
989	-3.351575	0.029999	Polimerase
1004	-3.350899	0.030017	Polimerase
1065	-6.721489	0.000958	Polimerase
1077	-3.360744	0.030955	Polimerase
1084	-3.352174	0.029983	Polimerase
1126	-3.360744	0.029748	Polimerase
1129	-3.076402	0.037037	Polimerase
1135	-4.09127	0.013376	Polimerase
1177	-4.268004	0.010533	Polimerase
1206	-3.076402	0.037037	Polimerase
1209	-3.076402	0.037233	Polimerase
1261	-4.480992	0.008669	Polimerase
1269	-3.076402	0.037037	Polimerase
1270	-3.360744	0.028409	Polimerase

1276	-3.354489	0.029921	Polimerase
1299	-3.076402	0.037037	Polimerase
1300	-4.274351	0.010543	Polimerase
1306	-4.480992	0.008669	Polimerase
1314	-3.360744	0.028409	Polimerase
1319	-3.076402	0.037037	Polimerase
1361	-2.95185	0.043591	Polimerase
1363	-3.076402	0.037037	Polimerase
1378	-3.360744	0.028409	Polimerase
1387	-3.076402	0.037037	Polimerase
1400	-6	0.001372	Polimerase
1428	-4.360459	0.037499	Polimerase
1453	-3.277281	0.028409	Polimerase
1484	-3	0.037037	Polimerase
1534	-3.80291	0.037279	Polimerase
1536	-3.277281	0.028409	Polimerase
1567	-3	0.037037	Polimerase
1569	-10.531736	0.03165	Polimerase
1572	-4	0.012346	Polimerase
1581	-3.277281	0.028409	Polimerase
1600	-3.076402	0.037037	Polimerase
1622	-10.00052	0.03418	Polimerase
1687	-4.480992	0.008669	Polimerase
1695	-4.10187	0.012346	Polimerase
1744	-3.360744	0.028409	Polimerase
1755	-3.360744	0.028409	Polimerase
1771	-3.076402	0.037992	Polimerase
1779	-3.360744	0.028409	Polimerase
1780	-3.209825	0.032652	Polimerase
1817	-3.076402	0.037037	Polimerase
1823	-3.076402	0.037037	Polimerase
1844	-4.480992	0.008669	Polimerase
1890	-3.340541	0.030095	Polimerase
1927	-3.076402	0.037037	Polimerase

1932	-4.10187	0.012346	Polimerase
1960	-11.666434	0.0293	Polimerase
1971	-3.339665	0.030119	Polimerase
1993	-5.127337	0.004119	Polimerase
1997	-3.360744	0.028409	Polimerase
2001	-3.360744	0.028409	Polimerase
2017	-3.210327	0.032687	Polimerase
2029	-3.360744	0.028409	Polimerase
2045	-3.345811	0.029953	Polimerase
28	-3.542475	0.026868	Glicoproteína
30	-3.549784	0.026745	Glicoproteína
59	-3.118959	0.037064	Glicoproteína
64	-3.345186	0.03024	Glicoproteína
115	-3.118959	0.037037	Glicoproteína
165	-3.118959	0.037361	Glicoproteína
172	-45.589561	0.007503	Glicoproteína
179	-3.549784	0.026745	Glicoproteína
198	-4.158612	0.012938	Glicoproteína
200	-3.549784	0.025122	Glicoproteína
202	-3.118959	0.037037	Glicoproteína
205	-3.118959	0.040258	Glicoproteína
211	-3.118959	0.037037	Glicoproteína
225	-5.916307	0.002155	Glicoproteína
326	-3	0.037037	Glicoproteína
346	-0.002874	0.001164	Glicoproteína
349	-3.409388	0.026901	Glicoproteína
362	-3.411248	0.026857	Glicoproteína
429	-4.541597	0.009044	Glicoproteína
447	-3.039437	0.037037	Glicoproteína
458	-4.052583	0.012421	Glicoproteína
482	-3.459277	0.026745	Glicoproteína
490	-3.039437	0.039266	Glicoproteína
509	-3.039437	0.03861	Glicoproteína
515	-4.158612	0.013495	Glicoproteína

519	-3.118959	0.037037	Glicoproteína
520	-3.118959	0.037037	Glicoproteína
522	-4.421283	0.009663	Glicoproteína
557	-4.158612	0.012346	Glicoproteína
591	-3.118959	0.037037	Glicoproteína
601	-4.158612	0.012346	Glicoproteína
624	-3.315963	0.03082	Glicoproteína
648	-3.34292	0.030552	Glicoproteína
670	-3.638786	0.045267	Glicoproteína
672	-3.315963	0.03082	Glicoproteína
705	-4.158612	0.012346	Glicoproteína
738	-3.546653	0.026853	Glicoproteína
769	-3.118959	0.037037	Glicoproteína
781	-3.118959	0.037105	Glicoproteína
793	-3.549784	0.025122	Glicoproteína
807	-3.118959	0.037037	Glicoproteína
824	-4.158612	0.01264	Glicoproteína
893	-3.549784	0.025122	Glicoproteína
923	-3.315963	0.03082	Glicoproteína
958	-3.118959	0.037037	Glicoproteína
971	-3.549784	0.025122	Glicoproteína
1001	-3.118959	0.037037	Glicoproteína
1051	-3.118959	0.037037	Glicoproteína
1090	-4.158612	0.012346	Glicoproteína
1115	-11.0793	0.010552	Glicoproteína
1116	-3.118959	0.037037	Glicoproteína
1127	-4.158612	0.012424	Glicoproteína
1128	-3.118959	0.037208	Glicoproteína
1141	-3.549784	0.025122	Glicoproteína
1159	-3.118959	0.037037	Glicoproteína
1184	-3.118959	0.037037	Glicoproteína
1191	-3.118959	0.037037	Glicoproteína
1192	-3.118959	0.037037	Glicoproteína
56	-3.377157	0.025962	Nucleocapsídeo

88	-0.000133	0.001823	Nucleocapsídeo
116	-5	0.004642	Nucleocapsídeo
119	-3	0.039921	Nucleocapsídeo
127	-3	0.037116	Nucleocapsídeo
142	-4	0.012346	Nucleocapsídeo
148	-3.377157	0.025962	Nucleocapsídeo
151	-3.751373	0.036666	Nucleocapsídeo
170	-3	0.037037	Nucleocapsídeo
192	-3.492527	0.04557	Nucleocapsídeo
254	-4.787689	0.007513	Nucleocapsídeo
76	-3	0.03928	Não estrutural
90	-3.340128	0.029765	Não estrutural
95	-3	0.037115	Não estrutural
97	-3	0.037037	Não estrutural
186	-3	0.037037	Não estrutural
209	-9.601229	0.034718	Não estrutural
226	-4	0.012346	Não estrutural
232	-3	0.037037	Não estrutural
240	-3	0.037037	Não estrutural

Dinâmicas evolucionárias diferentes por segmento

Eventos de reagrupamento podem desacoplar dinâmicas evolucionárias de segmentos genômicos (CARPI; HOLMES; KITCHEN, 2010; RAMBAUT et al., 2008). Diante deste fato, nós estimamos as taxas evolucionárias (μ) para cada segmento independentemente e comparamos os valores de verossimilhança com os testes de critério de informação do Akaike através de cadeias de Markov (AICM) e também do *path-sampling* (BAELE et al., 2012) para escolher a μ que melhor se adapta aos dados (Tabela 15).

Tabela 15. Comparação entre taxas de evolução diferentes para cada segmento

Segmento	AICM	σ	Diferença entre AICM	log (V) marginal do PS	μ	95 % HPD μ
Longo	41130	0.455	4.107	-13207.78	2.31E-04	(1.88E-4 a 2.72E-4)
Longo	41135	0.259	-4.107	-18826.17	3.28E-04	(2.56E-4 a 3.98E-4)
Médio	27870	0.298	6.654	-14206.09	3.80E-04	(2.53E-4 a 5.05E-4)
Médio	27876	0.199	-6.654	-14208.76	2.00E-04	(1.52E-4 a 2.4E-4)

Curto	11867	0.31	17.718	-5687.28	3.36E-04	(2.42E-4 a 4.26E-4)
Curto	11885	0.217	-17.718	-5413.06	2.10E-04	(1.66E-4 a 2.47E-4)

AICM: critério de informação do Akaike através de cadeias de Markov. μ : taxa de evolução em substituições por sítio por ano. Log(V): logaritmo do valor de verossimilhança. σ : erro padrão.

A melhor μ para o segmento L foi 2,31E-4 substituições por sítio por ano (s/s/a), variando de 1,88E-4 a 2,72E-4 (dentro do 95% HPD). Para o segmento M, a μ média de 3,80E-4 s/s/a (HPD= 2,53E-4 a 5,05E-4 s/s/a) se adaptou melhor ao dados. Finalmente, para o segmento S a melhor m foi de 3,36E-4 s/s/a (HPD = 2,42E-4 a 4,26E-4 s/s/a). Em suma, a taxa de evolução média ótima para o segmento M é 1,65 vezes maior que a taxa para o segmento L e 1,45 vezes maior para o segmento S (Figura 37A). As diferenças nas taxas implicam em uma raiz da árvore mais antiga para o segmento L e raízes sobrepostas para os segmentos M e S (Figura 37B).

Os tamanhos populacionais efetivos ($N_e.g$) estimados separadamente para cada segmento no tempo experimentaram episódios de anticorrelação o que sugere algum tipo de desacoplamento nas dinâmicas evolutivas inferidas para cada segmento (Figura 37C). As anticorrelações foram investigadas com a estimativa dos coeficientes de Kendall (τ), cujos valores negativos ($\tau < 0$) indicam correlação negativa entre as distribuições estudadas. Estas últimas análises apontaram que nos períodos de 1950 a 1970 e de 1990 a 2010 os $N_e.g$ estimados para os segmentos foram anticorrelacionados (Figura 37D). Interessantemente, as raízes das linhagens *reassortants* (intervalos verticais na Figura 37C) coincidem com os períodos em que se observa as anticorrelações significativas.

Figura 36. Estimativas filodinâmicas para a melhor taxa evolucionária estimada para cada segmento. A) Distribuição das taxas evolucionárias (μ) estimadas durante a convergência do MCMC. B) Distribuição das raízes estimadas (TMRCA) para cada segmento. C) BSP mostrando a flutuação do tamanho populacional efetivo ($N_e.g$) no tempo estimada para cada segmento. Os intervalos verticais coloridos indicam as idades dos nós estimados para os *reassortants* detectados (vermelho para 2007000608 e TANTan00107; cinza para SA75, 3574 e 76370). D) Correlação entre os para de $N_e.g$ em intervalos de 20 anos. Coeficientes de correlação de Kendall (τ) < 0 indicam anticorrelação nos intervalos e p -valores < 0,05, significância estatística.

Conclusões

O presente trabalho compara a biologia evolutiva entre duas classes (IV e V) de vírus com genoma de RNA. Um dos aspectos interessantes deste trabalho é que vírus de RNA de fita negativa de genoma segmentado, como o RVFV, dispõe de um mecanismo de rearranjo (*reassortment* em inglês) de segmentos, gerando diversidade genética por meio de *shift*, assim como por ajustes adaptativos finos, *drift*, causados por mutações pontuais que podem provocar mudanças em sequências de aminoácidos ou regulatórias. É interessante notar que não observamos seleção positiva em RVFV sugerindo que nossas amostras remetem a linhagens em um ótimo adaptativo. Não obstante, nossos resultados demonstram ter havido reagrupamento de segmentos genômicos durante a evolução recente de RVFV e que as dinâmicas evolutivas são diferentes, não sendo possível responsabilizar seleção atuando em nível de códons.

Por sua vez, os vírus de RNA de fita única senso positivo, como os flavivírus não apresentam comumente rearranjos genômicos significativos causados por recombinação, mas sim *drift* causado por mutações pontuais. Portanto, os achados de evidência de recombinação em ZIKV são importantes e podem fornecer subsídios para o melhor entendimento de um flavivírus em um ciclo silvestre, além disso descrevemos em algum detalhe a evolução em sítios de glicosilação no envelope viral o que pode ter relevantes repercussões para a compreensão da capacidade deste vírus de colonizar diversos hospedeiros.

A elucidação dos padrões migratório de TBEV na Europa central nos permitiu estudar um modelo viral com um padrão muito distinto de espalhamento, caracterizado por atividade focal. O vírus cresce primeiro em uma localidade, foco, ficando restrito a um raio de espalhamento, em seguida o vírus é levado para outra localidade onde tem uma explosão local novamente e assim por diante. Esta peculiaridade do TBEV nos permitiu aplicar uma abordagem baseada em comparação de distâncias para investigar os padrões de espalhamento de TBEV.

Em suma, os diferentes modelos virais nos possibilitaram estudar mecanismos evolutivos diferentes, alterações genômicas grandes: (i) rearranjos de segmentos genômicos de RVFV e (ii) recombinação homóloga de ZIKV. Ainda estudamos os padrões de mutações pontuais (*drift*) em cada um dos três modelos estudados, no caso do TBEV este é o único mecanismo detectado de mudança em nossos estudos. Os regimes atuantes sobre os três vírus são semelhantes, sendo todos caracterizados pela presença de muitos sítios sob seleção purificadora em suas proteínas, o que é coerente com as passagens alternadas entre hospedeiros artrópodes e vertebrados que caracterizam os ciclos de transmissão dos três modelos virais. Ademais, o ciclo de transmissão alternado impõe muitos gargalos para mudanças serem observadas nas populações virais.

Bibliografia

- ANYAMBA, A. et al. Prediction, assessment of the Rift Valley fever activity in East and Southern Africa 2006-2008 and possible vector control strategies. **The American journal of tropical medicine and hygiene**, v. 83, n. 2 Suppl, p. 43–51, ago. 2010.
- ARCHER, B. N. et al. Outbreak of Rift Valley fever affecting veterinarians and farmers in South Africa, 2008. **South African Medical Journal = Suid-Afrikaanse Tydskrif Vir Geneeskunde**, v. 101, n. 4, p. 263–266, abr. 2011.
- ARISTOTLE. **History of Animals**. reprint ed. [s.l.] Kessinger Publishing, 2004, 2004. p. 304 pages
- ARISTOTLE. **The Organon**. reprint ed. [s.l.] Roger Bishop Jones, 2012, 2012. p. 428 pages
- BAELE, G. et al. Accurate Model Selection of Relaxed Molecular Clocks in Bayesian Phylogenetics. **Molecular biology and evolution**, p. 1–6, 22 out. 2012.
- BALTIMORE, D. Expression of animal virus genomes. **Bacteriological reviews**, v. 35, n. 3, p. 235–41, set. 1971.
- BARBOZA, P. et al. [Emerging viral infections in South East Asia and the Pacific region]. **Médecine et maladies infectieuses**, v. 38, n. 10, p. 513–23, out. 2008.
- BARKER, D.; PAGEL, M. Predicting functional gene links from phylogenetic-statistical analyses of whole genomes. **PLoS computational biology**, v. 1, n. 1, p. e3, jun. 2005.
- BEARCROFT, W. G. Zika virus infection experimentally induced in a human volunteer. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 50, n. 5, p. 442–8, set. 1956.
- BILLECOCQ, A. et al. NSs protein of Rift Valley fever virus blocks interferon production by inhibiting host gene transcription. **Journal of virology**, v. 78, n. 18, p. 9798–806, set. 2004.
- BIRD, B. H. et al. Complete genome analysis of 33 ecologically and biologically diverse Rift Valley fever virus strains reveals widespread virus movement and low genetic diversity due to recent common ancestry. **Journal of virology**, v. 81, n. 6, p. 2805–16, mar. 2007.
- BIRD, B. H. et al. Multiple virus lineages sharing recent common ancestry were associated with a Large Rift Valley fever outbreak among livestock in Kenya during 2006-2007. **Journal of virology**, v. 82, n. 22, p. 11152–66, nov. 2008.
- BOWEN, M. D. et al. A reassortant bunyavirus isolated from acute hemorrhagic fever cases in Kenya and Somalia. **Virology**, v. 291, n. 2, p. 185–90, 20 dez. 2001.
- CALISHER, C. H.; GOULD, E. A. Taxonomy of the virus family Flaviviridae. **Advances in virus research**, v. 59, p. 1–19, jan. 2003.
- CAO-LORMEAU, V.-M. et al. Zika virus, French polynesia, South pacific, 2013. **Emerging infectious diseases**, v. 20, n. 6, p. 1085–6, jun. 2014.

CARPI, G.; HOLMES, E. C.; KITCHEN, A. The evolutionary dynamics of bluetongue virus. **Journal of molecular evolution**, v. 70, n. 6, p. 583–92, jun. 2010.

CARREL, M. A et al. Spatiotemporal structure of molecular evolution of H5N1 highly pathogenic avian influenza viruses in Vietnam. **PloS one**, v. 5, n. 1, p. e8631, jan. 2010.

CHAMBERS, T. J. et al. Flavivirus genome organization, expression, and replication. **Annual review of microbiology**, v. 44, p. 649–88, jan. 1990.

CHAMBERS, T. J. et al. West Nile virus envelope proteins: nucleotide sequence analysis of strains differing in mouse neuroinvasiveness. **The Journal of general virology**, v. 79 (Pt 10, p. 2375–80, out. 1998.

CHEN, Y.; MAGUIRE, T.; MARKS, R. M. Demonstration of binding of dengue virus envelope protein to target cells. **Journal of virology**, v. 70, n. 12, p. 8765–72, dez. 1996.

CHEVALIER, V. et al. Rift Valley fever--a threat for Europe? **Euro surveillance : bulletin européen sur les maladies transmissibles = European communicable disease bulletin**, v. 15, n. 10, p. 19506, 11 mar. 2010.

CHEVALIER, V.; THIONGANE, Y.; LANCELOT, R. Endemic transmission of Rift Valley Fever in Senegal. **Transboundary and emerging diseases**, v. 56, n. 9-10, p. 372–4, dez. 2009.

COLLETT, M. S. et al. Complete nucleotide sequence of the M RNA segment of Rift Valley fever virus. **Virology**, v. 144, n. 1, p. 228–45, 15 jul. 1985.

DARWIN, C. **A Origem Das Espécies**. São Paulo: Martin Claret, 2004. p. 640

DAUBNEY, R.; HUDSON, J. R. Rift Valley Fever. **The Lancet**, v. 6, n. 02, p. 611–612, 6 jul. 1932.

DIALLO, M. et al. Vectors of Chikungunya virus in Senegal: current data and transmission cycles. **The American journal of tropical medicine and hygiene**, v. 60, n. 2, p. 281–6, fev. 1999.

DICK, G. W. A.; KITCHEN, S. F.; HADDOW, A. J. Zika virus. I. Isolations and serological specificity. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 46, n. 5, p. 509–20, set. 1952.

DICKERSON, R. E. The structures of cytochrome c and the rates of molecular evolution. **Journal of molecular evolution**, v. 1, n. 1, p. 26–45, jan. 1971.

DOBLER, G. et al. Epidemiology and distribution of tick-borne encephalitis. **Wiener medizinische Wochenschrift (1946)**, v. 162, n. 11-12, p. 230–8, jun. 2012.

DRUMMOND, A. J. et al. Relaxed phylogenetics and dating with confidence. **PLoS biology**, v. 4, n. 5, p. e88, maio 2006.

DRUMMOND, A. J.; RAMBAUT, A. BEAST: Bayesian evolutionary analysis by sampling trees. **BMC Evolutionary Biology**, v. 7, p. 214, 2007.

EDGAR, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic acids research**, v. 32, n. 5, p. 1792–7, jan. 2004.

- EGLOFF, M.-P. et al. An RNA cap (nucleoside-2'-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. **The EMBO journal**, v. 21, n. 11, p. 2757–68, 3 jun. 2002.
- EL MAMY, A. B. O. et al. Unexpected Rift Valley fever outbreak, northern Mauritania. **Emerging infectious diseases**, v. 17, n. 10, p. 1894–6, out. 2011.
- ERBEL, P. et al. Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus. **Nature structural & molecular biology**, v. 13, n. 4, p. 372–3, abr. 2006.
- FAGBAMI, A. H. Zika virus infections in Nigeria: virological and seroepidemiological investigations in Oyo State. **The Journal of hygiene**, v. 83, n. 2, p. 213–9, out. 1979.
- FAYE, O. et al. Rift Valley fever outbreak with East-Central African virus lineage in Mauritania, 2003. **Emerging infectious diseases**, v. 13, n. 7, p. 1016–23, jul. 2007.
- FAYE, O. et al. Reemergence of rift valley Fever, mauritania, 2010. **Emerging infectious diseases**, v. 20, n. 2, p. 300–3, fev. 2014.
- FINDLAY, G. M.; DAUBNEY, R. The Virus of Rift Valley Fever or Enzoötic Hepatitis. **The Lancet**, v. 24, n. 6, p. 1350–1, jun. 1931.
- FOURMENT, M.; GIBBS, M. J. PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. **BMC evolutionary biology**, v. 6, p. 1, jan. 2006.
- FOY, B. D. et al. Probable non-vector-borne transmission of Zika virus, Colorado, USA. **Emerging infectious diseases**, v. 17, n. 5, p. 880–2, maio 2011.
- GIORGI, C. et al. Sequences and coding strategies of the S RNAs of Toscana and Rift Valley fever viruses compared to those of Punta Toro, Sicilian Sandfly fever, and Uukuniemi viruses. **Virology**, v. 180, n. 2, p. 738–53, fev. 1991.
- GIRAUDOUX, P. **pgirmess: Data analysis in ecology**, 2013.
- GOULD, E. A; HIGGS, S. Impact of climate change and other factors on emerging arbovirus diseases. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 103, n. 2, p. 109–21, fev. 2009.
- GOUY, M.; GUINDON, S.; GASCUEL, O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. **Molecular biology and evolution**, v. 27, n. 2, p. 221–4, fev. 2010.
- GROBBELAAR, A. A et al. Molecular epidemiology of Rift Valley fever virus. **Emerging infectious diseases**, v. 17, n. 12, p. 2270–6, dez. 2011.
- GUZMAN, M. G. et al. Dengue: a continuing global threat. **Nature reviews. Microbiology**, v. 8, n. 12 Suppl, p. S7–16, dez. 2010.
- HADDOW, A. D. et al. Genetic Characterization of Zika Virus Strains: Geographic Expansion of the Asian Lineage. **PLoS Neglected Tropical Diseases**, v. 6, n. 2, p. e1477, 28 fev. 2012.

HALL, B. K. Descent with modification: the unity underlying homology and homoplasy as seen through an analysis of development and evolution. **Biological reviews of the Cambridge Philosophical Society**, v. 78, n. 3, p. 409–33, ago. 2003.

HANADA, K.; SUZUKI, Y.; GOJOBORI, T. A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. **Molecular biology and evolution**, v. 21, n. 6, p. 1074–80, jun. 2004.

HANNA, S. L. et al. N-linked glycosylation of west nile virus envelope proteins influences particle assembly and infectivity. **Journal of virology**, v. 79, n. 21, p. 13262–74, nov. 2005.

HANOTTE, O. et al. African pastoralism: genetic imprints of origins and migrations. **Science (New York, N.Y.)**, v. 296, n. 5566, p. 336–9, 12 abr. 2002.

HARTLEY, D. M. et al. Potential effects of Rift Valley fever in the United States. **Emerging infectious diseases**, v. 17, n. 8, p. e1, ago. 2011.

HAYES, E. B. Zika virus outside Africa. **Emerging infectious diseases**, v. 15, n. 9, p. 1347–50, set. 2009.

HEINZE, D. M.; GOULD, E. A.; FORRESTER, N. L. Revisiting the clinal concept of evolution and dispersal for the tick-borne flaviviruses by using phylogenetic and biogeographic analyses. **Journal of virology**, v. 86, n. 16, p. 8663–71, ago. 2012.

HOLMES, E. C. Patterns of intra- and interhost nonsynonymous variation reveal strong purifying selection in dengue virus. **Journal of virology**, v. 77, n. 20, p. 11296–8, out. 2003.

HOLMES, E. C.; GRENFELL, B. T. Discovering the phylodynamics of RNA viruses. **PLoS computational biology**, v. 5, n. 10, p. e1000505, out. 2009.

HOLMES, E. C.; WOROBEY, M.; RAMBAUT, A. Phylogenetic evidence for recombination in dengue virus. **Molecular biology and evolution**, v. 16, n. 3, p. 405–9, mar. 1999.

HOOGSTRAAL, H. et al. Ticks (Ixodidae) on birds migrating from Europe and Asia to Africa 1959–61. **Bulletin of the World Health Organization**, v. 28, n. 2, p. 235–62, jan. 1963.

HOOGSTRAAL, H. et al. The Rift Valley fever epizootic in Egypt 1977–78. 2. Ecological and entomological studies. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 73, n. 6, p. 624–9, jan. 1979.

HUELSENBECK, J. P.; RONQUIST, F. MRBAYES: Bayesian inference of phylogenetic trees. **Bioinformatics (Oxford, England)**, v. 17, n. 8, p. 754–5, ago. 2001.

JACOBS, M. G. et al. Dengue virus nonstructural protein 1 is expressed in a glycosyl-phosphatidylinositol-linked form that is capable of signal transduction. **FASEB journal : official publication of the Federation of American Societies for Experimental Biology**, v. 14, n. 11, p. 1603–10, ago. 2000.

JERZAK, G. V. S. et al. Genetic diversity and purifying selection in West Nile virus populations are maintained during host switching. **Virology**, v. 374, n. 2, p. 256–60, 10 maio 2008.

JOHNSON, K. Emerging Viruses in Context: An Overview of Viral Hemorrhagic Fevers. In: MORSE, S. (Ed.). **Emerging Viruses**. New York: Oxford University Press, 1993. p. 317.

JOMBART, T.; BALLOUX, F.; DRAY, S. adephylo: new tools for investigating the phylogenetic signal in biological traits. **Bioinformatics (Oxford, England)**, v. 26, n. 15, p. 1907–9, 1 ago. 2010.

JONES, M. et al. Dengue virus inhibits alpha interferon signaling by reducing STAT2 expression. **Journal of virology**, v. 79, n. 9, p. 5414–20, maio 2005.

KAISER, R. Tick-borne encephalitis: Clinical findings and prognosis in adults. **Wiener medizinische Wochenschrift (1946)**, v. 162, n. 11-12, p. 239–43, jun. 2012.

KHROMYKH, A. A. et al. Coupling between replication and packaging of flavivirus RNA: evidence derived from the use of DNA-based full-length cDNA clones of Kunjin virus. **Journal of virology**, v. 75, n. 10, p. 4633–40, maio 2001.

KNAP, N. et al. Patterns of tick-borne encephalitis virus infection in rodents in slovenia. **Vector borne and zoonotic diseases (Larchmont, N.Y.)**, v. 12, n. 3, p. 236–42, mar. 2012.

KORTEKAAS, J. et al. Creation of a nonspreading Rift Valley fever virus. **Journal of virology**, v. 85, n. 23, p. 12622–30, dez. 2011.

KUNO, G.; CHANG, G.-J. J. Biological transmission of arboviruses: reexamination of and new insights into components, mechanisms, and unique traits as well as their evolutionary trends. **Clinical microbiology reviews**, v. 18, n. 4, p. 608–37, out. 2005.

KUNO, G.; CHANG, G.-J. J. Full-length sequencing and genomic characterization of Bagaza, Kedougou, and Zika viruses. **Archives of virology**, v. 152, n. 4, p. 687–96, jan. 2007.

LABUDA, M. et al. Tick-borne encephalitis virus foci in Slovakia. **International journal of medical microbiology : IJMM**, v. 291 Suppl , p. 43–7, jun. 2002.

LANCIOTTI, R. S. et al. Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. **Emerging infectious diseases**, v. 14, n. 8, p. 1232–9, ago. 2008.

LAUGHLIN, L. W. et al. Epidemic Rift Valley fever in Egypt: observations of the spectrum of human illness. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 73, n. 6, p. 630–3, jan. 1979.

LEE, E. et al. Both E protein glycans adversely affect dengue virus infectivity but are beneficial for virion release. **Journal of virology**, v. 84, n. 10, p. 5171–80, maio 2010.

LEMEY, P. et al. Bayesian phylogeography finds its roots. **PLoS computational biology**, v. 5, n. 9, p. e1000520, set. 2009.

LEMEY, P. et al. Phylogeography takes a relaxed random walk in continuous space and time. **Molecular biology and evolution**, v. 27, n. 8, p. 1877–85, ago. 2010.

LI, H. et al. The serine protease and RNA-stimulated nucleoside triphosphatase and RNA helicase functional domains of dengue virus type 2 NS3 converge within a region of 20 amino acids. **Journal of virology**, v. 73, n. 4, p. 3108–16, abr. 1999.

- LINDENBACH, B. D.; RICE, C. M. Molecular biology of flaviviruses. **Advances in virus research**, v. 59, p. 23–61, jan. 2003.
- LOBO, F. P. et al. Virus-host coevolution: common patterns of nucleotide motif usage in Flaviviridae and their hosts. **PloS one**, v. 4, n. 7, p. e6282, jan. 2009.
- MARRAMA, L. et al. Domestic transmission of Rift Valley Fever virus in Diawara (Senegal) in 1998. **The Southeast Asian Journal of Tropical Medicine and Public Health**, v. 36, n. 6, p. 1487–1495, nov. 2005.
- MARTIN, D. P. et al. RDP3: a flexible and fast computer program for analyzing recombination. **Bioinformatics (Oxford, England)**, v. 26, n. 19, p. 2462–3, 1 out. 2010.
- MARTIN, V. et al. The impact of climate change on the epidemiology and control of Rift Valley fever. **Revue Scientifique Et Technique (International Office of Epizootics)**, v. 27, n. 2, p. 413–426, ago. 2008.
- MEEGAN, J. M. The Rift Valley fever epizootic in Egypt 1977-78. 1. Description of the epizzotic and virological studies. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 73, n. 6, p. 618–23, jan. 1979.
- MININ, V. N.; BLOOMQUIST, E. W.; SUCHARD, M. A. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. **Molecular biology and evolution**, v. 25, n. 7, p. 1459–71, jul. 2008.
- MODIS, Y. et al. Structure of the dengue virus envelope protein after membrane fusion. **Nature**, v. 427, n. 6972, p. 313–9, 22 jan. 2004.
- MONDOTTE, J. A. et al. Essential role of dengue virus envelope protein N glycosylation at asparagine-67 during viral propagation. **Journal of virology**, v. 81, n. 13, p. 7136–48, jul. 2007.
- MOUTAILLER, S. et al. Host alternation is necessary to maintain the genome stability of rift valley fever virus. **PLoS neglected tropical diseases**, v. 5, n. 5, p. e1156, maio 2011.
- MÜLLER, R. et al. Rift Valley fever virus L segment: correction of the sequence and possible functional role of newly identified regions conserved in RNA-dependent polymerases. **The Journal of general virology**, v. 75 (Pt 6), p. 1345–52, jun. 1994.
- MUÑOZ-JORDAN, J. L. et al. Inhibition of interferon signaling by dengue virus. **Proceedings of the National Academy of Sciences of the United States of America**, v. 100, n. 24, p. 14333–8, 25 nov. 2003.
- MUSSO, D.; NILLES, E. J.; CAO-LORMEAU, V.-M. Rapid spreading of emerging Zika virus in the Pacific area. **Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases**, p. 1–5, 7 jun. 2014.
- NAGARAJAN, N.; KINGSFORD, C. GiRaF: robust, computational identification of influenza reassortments via graph mining. **Nucleic acids research**, v. 39, n. 6, p. e34, mar. 2011.
- NDERITU, L. et al. Sequential Rift Valley fever outbreaks in eastern Africa caused by multiple lineages of the virus. **The Journal of infectious diseases**, v. 203, n. 5, p. 655–65, 1 mar. 2011.

- NEI, M.; SUZUKI, Y.; NOZAWA, M. The neutral theory of molecular evolution in the genomic era. **Annual review of genomics and human genetics**, v. 11, p. 265–89, 22 set. 2010.
- OKSANEN, J. et al. **vegan: Community Ecology Package**. Disponível em: <<http://cran.r-project.org/package=vegan>>.
- OLSON, J. G. et al. Zika virus, a cause of fever in Central Java, Indonesia. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 75, n. 3, p. 389–93, jan. 1981.
- PARADIS, E.; CLAUDE, J.; STRIMMER, K. APE: Analyses of Phylogenetics and Evolution in R language. **Bioinformatics (Oxford, England)**, v. 20, n. 2, p. 289–90, 22 jan. 2004.
- PARKER, J.; RAMBAUT, A.; PYBUS, O. G. Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. **Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases**, v. 8, n. 3, p. 239–46, 2008.
- PEPIN, M. et al. Rift Valley fever virus(Bunyaviridae: Phlebovirus): an update on pathogenesis, molecular epidemiology, vectors, diagnostics and prevention. **Veterinary research**, v. 41, n. 6, p. 61, 2010.
- PFEFFER, M.; DOBLER, G. Emergence of zoonotic arboviruses by animal trade and migration. **Parasites & Vectors**, v. 3, n. 1, p. 35, 2010.
- POND, S. L. K.; FROST, S. D. W.; MUSE, S. V. HyPhy: hypothesis testing using phylogenies. **Bioinformatics (Oxford, England)**, v. 21, n. 5, p. 676–9, 1 mar. 2005.
- POSADA, D.; CRANDALL, K. A. The effect of recombination on the accuracy of phylogeny estimation. **Journal of molecular evolution**, v. 54, n. 3, p. 396–402, mar. 2002.
- PRICE, M. N.; DEHAL, P. S.; ARKIN, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. **PloS one**, v. 5, n. 3, p. e9490, jan. 2010.
- RAMBAUT, A. et al. The genomic and epidemiological dynamics of human influenza A virus. **Nature**, v. 453, n. 7195, p. 615–9, 29 maio 2008.
- REY, F. A. et al. The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. **Nature**, v. 375, n. 6529, p. 291–8, 25 maio 1995.
- RONQUIST, F.; HUELSENBECK, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. **Bioinformatics (Oxford, England)**, v. 19, n. 12, p. 1572–4, 12 ago. 2003.
- SALL, A. A. et al. Variability of the NS(S) protein among Rift Valley fever virus isolates. **The Journal of general virology**, v. 78 (Pt 11, p. 2853–8, nov. 1997.
- SALL, A. A. et al. Molecular epidemiology and emergence of Rift Valley fever. **Memórias do Instituto Oswaldo Cruz**, v. 93, n. 5, p. 609–14, 1998a.
- SALL, A. A. et al. Genetic reassortment of Rift Valley fever virus in nature. **Journal of virology**, v. 73, n. 10, p. 8196–200, out. 1999.

SALL, A. AL et al. Origin of 1997-98 Rift Valley fever outbreak in East Africa. **Lancet**, v. 352, n. 9140, p. 1596–7, 14 nov. 1998b.

SANDERSON, M. J. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. **Molecular biology and evolution**, v. 19, n. 1, p. 101–9, jan. 2002.

SANDERSON, M. J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. **Bioinformatics (Oxford, England)**, v. 19, n. 2, p. 301–2, 22 jan. 2003.

SCHIERUP, M. H.; HEIN, J. Consequences of recombination on traditional phylogenetic analysis. **Genetics**, v. 156, n. 2, p. 879–91, out. 2000.

SCHMIDT, H. A et al. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. **Bioinformatics**, v. 18, n. 3, p. 502–504, 1 mar. 2002.

SIMON-LORIERE, E.; HOLMES, E. C. Why do RNA viruses recombine? **Nature reviews. Microbiology**, v. 9, n. 8, p. 617–626, 4 jul. 2011.

SIMPSON, D. I. Zika virus infection in man. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 58, n. 4, p. 335–8, jul. 1964.

SOBER, E. **Reconstructing the past: Parsimony, Evolution, and Inference**. Cambridge: The MIT Press, 1988. p. 265

SOMARÉ, P. O. L. et al. Phylogeography of Rift Valley Fever virus in Africa reveals multiple introductions in Senegal and Mauritania. **PloS one**, v. 7, n. 4, p. e35216, 23 jan. 2012.

STRIMMER, K.; VON HAESLER, A. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. **Proceedings of the National Academy of Sciences of the United States of America**, v. 94, n. 13, p. 6815–9, 24 jun. 1997.

TASSY, P. Trees before and after Darwin. **Journal of Zoological Systematics and Evolutionary Research**, v. 49, n. 2, p. 89–101, 15 maio 2011.

TAUCHER, C.; BERGER, A.; MANDL, C. W. A trans-complementing recombination trap demonstrates a low propensity of flaviviruses for intermolecular recombination. **Journal of virology**, v. 84, n. 1, p. 599–611, jan. 2010.

TERASAKI, K. et al. Mechanism of tripartite RNA genome packaging in Rift Valley fever virus. **Proceedings of the National Academy of Sciences of the United States of America**, v. 108, n. 2, p. 804–9, 11 jan. 2011.

TOLOU, H. J. et al. Evidence for recombination in natural populations of dengue virus type 1 based on the analysis of complete genome sequences. **The Journal of general virology**, v. 82, n. Pt 6, p. 1283–90, jun. 2001.

TRAORÉ-LAMIZANA, M. et al. Arbovirus surveillance from 1990 to 1995 in the Barkedji area (Ferlo) of Senegal, a possible natural focus of Rift Valley fever virus. **Journal of Medical Entomology**, v. 38, n. 4, p. 480–492, jul. 2001.

TURELL, M. J. et al. Generation and transmission of Rift Valley fever viral reassortants by the mosquito *Culex pipiens*. **The Journal of general virology**, v. 71 (Pt 10, p. 2307–12, out. 1990.

WALDENSTRÖM, J.; LUNDKVIST, Å. Migrating birds and tickborne encephalitis virus. **Emerging infectious ...**, v. 13, n. 8, p. 1215–1218, 2007.

WEAVER, S. C. Host range, amplification and arboviral disease emergence. **Archives of virology. Supplementum**, n. 19, p. 33–44, jan. 2005.

WEAVER, S. C.; REISEN, W. K. Present and future arboviral threats. **Antiviral research**, v. 85, n. 2, p. 328–45, fev. 2010.

WEAVER, S. C.; VASILAKIS, N. Molecular evolution of dengue viruses: contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease. **Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases**, v. 9, n. 4, p. 523–40, jul. 2009.

WEIDMANN, M. et al. Tick-borne encephalitis virus in *Clethrionomys glareolus* in the Czech Republic. **Vector borne and zoonotic diseases (Larchmont, N.Y.)**, v. 6, n. 4, p. 379–81, jan. 2006.

WEIDMANN, M. et al. Relation of genetic phylogeny and geographical distance of tick-borne encephalitis virus in central Europe. **The Journal of general virology**, v. 92, n. Pt 8, p. 1906–16, ago. 2011.

WESTESSON, O.; HOLMES, I. Accurate detection of recombinant breakpoints in whole-genome alignments. **PLoS computational biology**, v. 5, n. 3, p. e1000318, mar. 2009.

WIENS, J. J. Missing Data, Incomplete Taxa, and Phylogenetic Accuracy. **Systematic Biology**, v. 52, n. 4, p. 528–538, 1 ago. 2003.

WIENS, J. J. Missing data and the design of phylogenetic analyses. **Journal of biomedical informatics**, v. 39, n. 1, p. 34–42, fev. 2006.

WORLD HEALTH ORGANIZATION. **WHO | Saudi Arabia and Yemen: First cases of Rift Valley fever reported outside Africa, 2000**. Disponível em: <<http://www.who.int/csr/outbreaknetwork/saudiarabia/en/>>. Acesso em: 11 ago. 2011.

WORLD HEALTH ORGANIZATION. **WHO | Rift Valley Fever in Kenya, Somalia and the United Republic of Tanzania**. Disponível em: <http://www.who.int/csr/don/2007_05_09/en/>. Acesso em: 11 ago. 2011.

WORLD HEALTH ORGANIZATION. **WHO | Rift Valley Fever in Sudan - update 5**. Disponível em: <http://www.who.int/csr/don/2008_01_22/en/index.html>. Acesso em: 11 ago. 2011.

WORLD HEALTH ORGANIZATION. **WHO | Rift Valley fever in South Africa**. Disponível em: <http://www.who.int/csr/don/2010_03_30a/en/>. Acesso em: 10 ago. 2011.

ZANOTTO, P. M. et al. An arbovirus cline across the northern hemisphere. **Virology**, v. 210, n. 1, p. 152–9, 20 jun. 1995.

ZELLER, H. G. et al. Enzootic activity of Rift Valley fever virus in Senegal. **The American Journal of Tropical Medicine and Hygiene**, v. 56, n. 3, p. 265–272, mar. 1997.

ZWICKL, D. **Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion.** [s.l.] The University of Texas at Austin., 2006.

Molecular Evolution of Zika Virus during Its Emergence in the 20th Century

Oumar Faye^{1,3}, Caio C. M. Freire^{2,3}, Atila Iamarino², Ousmane Faye¹, Juliana Velasco C. de Oliveira², Mawlouth Diallo¹, Paolo M. A. Zanotto², Amadou Alpha Sall^{1*}

¹ Institut Pasteur de Dakar, Dakar, Senegal, ² Laboratory of Molecular Evolution and Bioinformatics, Department of Microbiology, Biomedical Sciences Institute, University of Sao Paulo, Sao Paulo, Brazil

Abstract

Zika virus (ZIKV) is a mosquito-borne flavivirus first isolated in Uganda in 1947. Although entomological and virologic surveillance have reported ZIKV enzootic activity in diverse countries of Africa and Asia, few human cases were reported until 2007, when a Zika fever epidemic took place in Micronesia. In the context of West Africa, the WHO Collaborating Centre for Arboviruses and Hemorrhagic Fever at Institut Pasteur of Dakar (<http://www.pasteur.fr/recherche/banques/CRORA/>) reports the periodic circulation of ZIKV since 1968. Despite several reports on ZIKV, the genetic relationships among viral strains from West Africa remain poorly understood. To evaluate the viral spread and its molecular epidemiology, we investigated 37 ZIKV isolates collected from 1968 to 2002 in six localities in Senegal and Côte d'Ivoire. In addition, we included strains from six other countries. Our results suggested that these two countries in West Africa experienced at least two independent introductions of ZIKV during the 20th century, and that apparently these viral lineages were not restricted by mosquito vector species. Moreover, we present evidence that ZIKV has possibly undergone recombination in nature and that a loss of the N154 glycosylation site in the envelope protein was a possible adaptive response to the *Aedes dalzieli* vector.

Citation: Faye O, Freire CCM, Iamarino A, Faye O, de Oliveira JVC, et al. (2014) Molecular Evolution of Zika Virus during Its Emergence in the 20th Century. *PLoS Negl Trop Dis* 8(1): e2636. doi:10.1371/journal.pntd.0002636

Editor: Brian Bird, Centers for Disease Control and Prevention, United States of America

Received: March 4, 2013; **Accepted:** November 27, 2013; **Published:** January 9, 2014

Copyright: © 2014 Faye et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was made possible by the Institut Pasteur de Dakar in Senegal, FAPESP (Fundação de Amparo a Pesquisa do Estado de São Paulo, Brazil) projects #00/04205-6 (VIGDN program), #08/17013-6 and #10/19341-4. CCMF has a CAPES studentship, AI (project #12/04818-5) and JVCdO received FAPESP scholarships and PMAZ holds a CNPq-PQ grant. Funding was also provided by NIH grant R01-AI069145. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: asall@pasteur.sn

These authors contributed equally to this work.

Introduction

Zika virus (ZIKV) is a mosquito-borne flavivirus, a member of the Spondweni serocomplex, whose natural transmission cycle involves mainly vectors from the *Aedes* genus (*A. fuscifera*, *A. taylori*, *A. luteocephalus* and *A. africanus*) and monkeys [1], while humans are occasional hosts. Clinical pictures range from asymptomatic cases to an influenza-like syndrome associated to fever, headache, malaise and cutaneous rash [2,3]. Likewise, direct contact is also considered a potential route of transmission among humans, probably during sexual intercourse [4]. The first isolation of ZIKV was in 1947 from the blood of a sentinel Rhesus monkey No. 766, stationed in the Zika forest, near the Lake Victoria in Uganda, and in 1948 ZIKV was also isolated in the same forest from a pool of *A. africanus* mosquitoes [5]. Thereafter, serological and entomological data indicated ZIKV infections in the African continent in Nigeria in 1971 and 1975 [6], Sierra Leone in 1972 [7], Gabon in 1975 [8], Uganda in 1969 and 1970 [9], Central African Republic in 1979 [10], Senegal from 1988 to 1991 [11] and Côte d'Ivoire in 1999 [12]. Recently, ZIKV was detected in Senegal in 2011 and 2012 (unpublished data). In addition, ZIKV infections in Asia were reported in Pakistan [13], Malaysia [14], Indonesia in 1977 and 1978 [15], Micronesia in 2007 [16,17] and Cambodia in 2010

[18]. Although ZIKV was repeatedly isolated, only 14 human cases were reported before April 2007, when a Zika fever (ZF) epidemic occurred in Yap island in Micronesia, where 49 confirmed cases and 73% of the residents older than 3 years provided serologic evidence for recent ZIKV infection [16]. This outbreak showcased the potential of ZF as an emerging disease, which could be misdiagnosed as dengue fever, as happened during the beginning of the Micronesian outbreak [16,17].

The ZIKV genome consists of a single-stranded positive sense RNA molecule with 10794 kb of length with 2 flanking non-coding regions (5' and 3' NCR) and a single long open reading frame encoding a polyprotein: 5'-C-prM-E-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5-3', that is cleaved into capsid (C), precursor of membrane (prM), envelope (E) and seven non-structural proteins (NS) [19,20]. The E protein (≈53 kDa) is the major virion surface protein. E is involved in various aspects of the viral cycle, mediating binding and membrane fusion [21]. The NS5 protein (≈103 kDa) is the largest viral protein whose C-terminal portion has RNA-dependent RNA polymerase (RdRP) activity and the N-terminus is involved in RNA capping by virtue of its processing due to methyl transferase activity [21]. The 3'NCR of the ZIKV genome contains about 428 nucleotides, including 27 folding patterns [20] that may be involved in the recognition by cellular or

Author Summary

Zika fever is a mosquito-borne illness caused by a flavivirus. Human infections with Zika virus (ZIKV) could cause fever, malaise and cutaneous rash. Despite several ZIKV reports since 1947 when it was first isolated at Zika forest in Uganda, molecular evolution of ZIKV as an emerging agent remains poorly understood. Moreover, despite several ZIKV reports from Africa and Asia, few human cases were notified until 2007 when an epidemic took place in Micronesia. In West Africa, surveillance programs have reported periodic circulation of the virus since 1968. To help fill the gap in understanding ZIKV evolution, 43 ZIKV samples were analyzed. We focused on: (i) adaptive genetic changes including protein glycosylation patterns, (ii) phylogenetic relationship among isolates and their spatiotemporal patterns of spread across Africa and Asia and, (iii) dispersion among vertebrate reservoirs and invertebrate vector species. Our results indicated that ZIKV may have experienced recombination in nature and that, after it emerged from Uganda in the early of the 20th century, it moved to West Africa and Asia in the first half of the century, without any clear preference for host and vector species.

viral factors, translation, genome stabilization, RNA packaging, or cyclization [21]. Although diverse studies have contributed greatly to our understanding of the evolutionary biology of flaviviruses in general [22–25], few studies have addressed ZIKV evolutionary biology [17,26]. Those studies report three main ZIKV lineages, one from Asia and two from Africa. Aiming to fill this gap and gain better insights ZIKV molecular evolution in the 20th century, we investigated 43 ZIKV strains, sampled from 1947 to 2007 in Africa and Asia, to describe phylogenetic relationships, selective influences, recombination events, phylodynamics, phylogeography, host correlations with viral lineages and glycosylation patterns.

Methods

Ethical statements

Samples used in this study are part of the Institute Pasteur in Dakar collection (WHO Collaborating Centre for Arboviruses and/or Hemorrhagic Fever Reference and Research). Monkey and human strains (AnD 30332 and HD 78788) were obtained respectively in 1979 and 1991 in Senegal during routine surveillance. None of the data was directly derived from human or animal samples but rather from cell culture supernatant. Therefore all the samples were anonymous and only reference numbers were used during the analysis that originated this study.

Virus isolates

ZIKV strains were provided by CRORA at the Institute Pasteur of Dakar. The strains were obtained from mosquitoes, humans and other mammals isolated in Burkina Faso, Central African Republic, Côte d'Ivoire and Senegal in West Africa (Table S1). Viral stocks were prepared by inoculating viral strains into *Aedes pseudoscutellaris* clone 61 monolayer in Leibovitz 15 (L-15) growth medium (GibcoBRL, Grand Island, NY, USA) supplemented with 5% fetal bovine serum (FBS) (GibcoBRL, Grand Island, NY, USA), 10% Tryptose Phosphate and antibiotics (Sigma, GmbH, Germany). Viral infection was confirmed after seven days of propagation by an indirect immunofluorescence assay (IFA) using

specific hyper-immune mouse ascitic fluid, as described previously [27]. Cultures supernatants were collected for virus RNA isolation.

RNA extraction

RNA was extracted from ZIKV stocks using the QIAamp RNA Viral Kit (Qiagen, Hilden, Germany) according to the manufacturer's recommendations. RNA was eluted in 50 µl of AVE buffer and stored at −80°C until use.

RT-PCR amplification

For cDNA synthesis, 10 µl of viral RNA was mixed with 1 µl of each of a reverse primer (2 pmol), 1 µl of deoxynucleotide triphosphate (dNTP) (10 mM each dNTP) and the mixture was heated at 65°C for 5 min. Reverse transcription was performed in 20 µl mixture containing mixed of 2.5 U RNasin (Promega, Madison, USA) 5 U of Superscript III reverse transcriptase (Invitrogen, Carlsbad, USA) and incubated at 55°C for 50 min, followed by 70°C for 15 min. PCR products were generated independently using the primers Unifor/Unirev, FD3/FU1 and VD8/EMF1 to amplify partial E, NS5 and NS5/3'NC region respectively [28]. Five microliters of cDNA were mixed with 10× buffer, 5 µl of each primer, 5 µl of dNTPs 10 mM, 3 µl of MgCl₂, and 0.5 µl of Taq polymerase (Promega, Madison, USA).

Nucleotide sequencing

PCR products of the expected size were purified from agarose gels with the QiaQuick Gel Extraction Kit (Qiagen, Hilden, Germany) as specified by the manufacturer. Both strands of each PCR product were sequenced directly with the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit V3.1 on an Applied Biosystems 3100 DNA Analyzer (Applied Biosystem, Foster City, CA, USA) at the Laboratory of Molecular Evolution and Bioinformatics, Biomedical Sciences Institute, University of Sao Paulo, Brazil. We deposited thirty two 753 bp-long sequences from the E gene (Accession numbers: KF383015–KF383046), thirty one of NS5 (708 bp) (Accession numbers: KF38304–KF383114), thirty seven of 3'NCR (537 bp) (Accession numbers: KF383047–KF383083) and six genomes (10274 bp) (Accession numbers: KF383115–KF383120) in GenBank (www.ncbi.nlm.nih.gov/genbank/) from thirty eight viral strains (Table S1). Additional sequences representing strains from Kedougou in Senegal, Nigeria, Malaysia, the Ugandan prototype MR766, the strain related to Micronesian outbreak in 2007 and the Spondweni virus were obtained from GenBank, with the following accession numbers, respectively: HQ234501, HQ234500, HQ234499, NC_012532, EU545988 and DQ859064.1 (Table S1).

Recombination detection

Prior to the analyses, all sequences were aligned with MUSCLE v3.7 [29] and manually edited with SeaView v4.3.3 [30]. To prevent potential biases during phylogenetic inference due to recombination, we first analyzed the sequences of available ZIKV genomes with RDP v4.4.8 program [31] that incorporates RDP [32], GENECONV [33], Chimera [34], MaxChi [35], Bootscan [36], SiScan [37] and 3Seq [38] methods to uncover evidence for recombination events. Only events with *p*-values ≤ 0.01 that were confirmed by four or more methods were considered, using the Bonferroni correction to prevent false positive results [39], as implemented in the RDP program [31]. In addition, the occurrence of recombination in genomes was also investigated with the Rec-HMM program that estimates breakpoints based on the Phylo-HMM approach, which models tree topology changes over the columns of a multiple alignment [40]. Moreover, potential

intra-gene recombination was also inspected with RDP using individual gene datasets, and the incompatibility among phylogenies inferred from genes (NS5 and E) was also investigated with GiRaF v1.01 [41] that searches incompatible clades among posterior set of trees (PST) obtained from different genomic regions with threshold of 70% for incompatible clades. The PST was obtained during Monte Carlo Markov chain (MCMC) stationarity using four chains, one 'cold' and three 'heated', after 20 million of generations, sampling every 5000 generations using MrBayes v3.2.1 [42].

Phylogenetic analyses

The phylogenetic signal content of the sequence datasets to phylogenetic reconstruction was investigated by Likelihood mapping method [43], implemented in TREE-PUZZLE v5.2 [44]. The concordance among gene (E and NS5) datasets without recombinant sequences was further assessed using a permutation test with significance level (α) of 0.05 after 10000 permutations, implemented in the Congruence among Distance Matrices (CADM) package [45]. Phylogenetic trees were generated by Maximum Likelihood (ML) criterion using GARLI v2.0 [46] that uses a stochastic algorithm to estimate simultaneously the best topology, branch lengths and substitution model parameters that maximize the log Likelihood (lnL). The confidence of ML trees was assessed by the convergence of lnL scores from ten independent replicates. We used a substitution model based on general time reversible (GTR) model with gamma-distributed rate variation (Γ) and a proportion of invariant sites (I). Support for the topology was obtained after 1000 non-parametric bootstrap replicates with GARLI. Then, we summarized the bootstrap trees into one consensus tree to access bootstrap values, using Dendropy v3.10.1 [47].

Selection analyses

To infer the selection pressures acting on each gene of ZIKV, we estimated the difference between the non-synonymous (dN) and synonymous (dS) rates per codon site using the single likelihood ancestor counting (SLAC) algorithm available in HyPhy v0.99 [48], assuming a significance level of 1% ($\alpha=0.01$). In the HyPhy output, values of ω are expressed as $\omega=dN-dS$. Therefore, ω smaller than zero ($\omega<0$) indicate purifying, negative selection.

Prediction of glycosylation sites

Potential glycosylation sites that may have adaptive value were previously described in ZIKV proteins [17,20,26]. Thus, we investigated partial E sequences to detect potential glycosylation sites using NetCGlyc v1.0 [49], NetOGlyc v3.1 [50], YinOYang v1.2 [51] and NetNGlyc v1.0 [51,52] methods that employ algorithms based in neural networks to predict, respectively, C-mannosylated, mucin-type O-linked, N-acetylglucosamine (GlcNAc) and N-linked glycosylation sites. To infer the structural position of the predicted glycosylation sites, we modeled the tridimensional structures of E regions of viral polypeptide of the Micronesian strain (GenBank accession number ACD75819). We used the homologous sequences from Japanese Encephalitis virus (PDB code 3p54), West Nile virus (PDB code 2i69) and Dengue virus type 3 (PDB code 1uzg). The amino acids sequences were aligned using MUSCLE v3.7 [29], a total of 1000 independent models were generated and optimized using Modeller v9.10 [53], and the best models were validated with PROCHECK v3.5.4 [54].

Phyldynamic analysis

Maximum Clade Credibility (MCC) trees were inferred using a Markov Chain Monte Carlo (MCMC) Bayesian approach

implemented on the program BEAST v1.6.2 [55] under GTR + Γ + I and a relaxed (uncorrelated lognormal) molecular clock [56]. MCMC convergence was obtained for four independent runs with 50 million generations, which were sufficient to obtain a proper sample for the posterior at MCMC stationarity, assessed by effective sample sizes (ESS) above 200 inspected using Tracer v1.5 (<http://tree.bio.ed.ac.uk/software/tracer/>). Furthermore, using the concatenated sequences of E and NS5 genes, we employed a discrete model attributing state characters representing isolation locality, animal source, recombination and N-linked glycosylation on E protein of each of the strains with the Bayesian Stochastic Search Variable (BSSVS) algorithm [57], implemented in BEAST. This method estimates the most probable state at each node in the MCC trees, allowing us to reconstruct plausible ancestral states on these nodes. Moreover, we represented the viral migration in Google Earth (<http://www.google.com/earth/>), using the SPREAD v1.0.3 program [58]. We evaluated the correlation among viral states and inferred phylogenies from PST by the parsimony score (PS), association index (AI) and monophyletic clade size (MC), with BaTS v1.0 [59] after 10000 null replications. In addition, we investigated the occurrence of correlated evolutionary change among ZIKV phenotypes (glycosylation pattern and vector host) along PST, employing a ML approach to test the fit of the two evolutionary models, one where the two traits evolve independently on the phylogenetic tree (independent model), and one where they evolve in a correlated way (dependent model) [60], using BayesTraits program (<http://www.evolution.rdg.ac.uk/>). To evaluate model suitability to ZIKV data, we estimated the marginal likelihoods for both models after 1000 bootstrap replications and compared Bayes factors (BF) between models [61], using Tracer v1.5.

Results/Discussion

Recombination among ZIKV strains

The primary analyses with RDP suggested 13 recombination events in ZIKV complete genomes (Table S2). Rec-HMM also detected genomic breakpoints with confidence in the following alignment positions: 1044 to 1095, 5181 to 5238, 9007 to 9132 and 9580 to 9631 (Figure S1). Since the results obtained by both methods revealed breakpoints in the E and NS5 genomic regions, we investigated these evidences with RDP on partial gene sequences. We found a single event in E sequences with estimated breakpoints near to the 414th and 1065th site of E gene reaching nine viral strains: ArA986, HD78788, ArA27101, ArA27290, ArA27096, ArA27443, ArA27407, ArA27106 and ArA982. These results were found by Bootscan, Maxchi, Chimaera, SiScan and 3Seq methods and supported by significant p -values of 1.31E-5, 2.85E-3, 1.59E-3, 1.79E-29 and 6.85E-19, respectively. Likewise, only one recombination event was detected in NS5 sequences with estimated breakpoints near sites 1581 and 2152 of the NS5 gene from strains ArD158084, ArB1362 and ArD157995. These findings were confirmed by Bootscan, Maxchi, Chimaera, SiScan and 3Seq methods and supported by significant p -values of 9.93E-9, 3.32E-7, 3.32E-7, 5.27E-28 and 7.65E-24, respectively. These potential recombinant sequences were excluded from further analyses to avoid inferential biases [62,63]. To perform the phylogenetic analysis we concatenated E and NS5 sequences and replaced inferred recombinant fragments with missing data. This is in line with the use of Maximum Likelihood approaches, which is fairly robust to the introduction of gaps [64,65]. In addition, we found incompatibilities between E and NS5 phylogenies using GiRaF. The three discordant strains (ArD128000, ArA1465 and ArD142623) were excluded, and we used 40 (31 from E and 36

from NS5) concatenated sequences for phylogenetic analysis. Moreover, we also found that the two remaining datasets for E and NS5 have no conflicting phylogenetic signal, as estimated by a CADM test (p -value = 9.99E-5 and α = 0.05). Given the limited sampling that we investigated, these results indicate that ZIKV may be experiencing recombination in the field, which is uncommon among flaviviruses [66]. These findings remain to be properly evaluated and assessed related to their effects on viral spread, zoonotic maintenance and epidemiologic potential. The possibility that our findings could be a consequence of cross contamination among isolates seems highly improbable given the extreme precautions that were taken. RNA extraction and reverse transcription were done separately for each isolate under BSL-II cabinets, sequenced several times leading to identical sequences, even when processed in different laboratories in Sao Paulo, Brazil, and Dakar, Senegal.

Phylogenetic analysis

We first investigated the phylogenetic signal content in our data by reconstructing 50000 quartets for each gene segment using the likelihood mapping method (see methods section). Our results indicated that NS5 and E datasets had high phylogenetic signal content given their lower percentage of unresolved quartets (3.2% and 3.4%, respectively), while 3'NCR showed less signal (16.4% of unresolved quartets) and was not considered. The ML trees for E (data not shown), NS5 (data not shown) and the two concatenated genes (Figure 1) reinforced that ZIKV strains could be classified in three major clusters [17]. Accordingly, the African strains were arranged into two groups: the MR766 prototype strain cluster (yellow sector on Figure 1) and the Nigerian cluster (green sector on Figure 1); and the Micronesian and Malaysian strains clustered together forming the Asian clade (Figure 1), in agreement with [26]. For West Africa, the strains from Côte d'Ivoire and Senegal were found in both African clusters, suggesting that at least two distinct lineages of ZIKV circulated in these countries. Interestingly, we found that the position of the Senegalese cluster, comprising viruses isolated from 1998 to 2001 associated with *A. dalzielii*, branching as a sister group of HD78788 isolated in Senegal in 1991, was not simply explained by recombination (with both Giraf and RDP) or poor rooting of the tree, since it did not depend on the inclusion (Figure 1) or exclusion (Figure S2) of the Spondweni, which is a *bonafide* outgroup. It was observed 65% of the time during a highly stringent maximum likelihood (ML) analysis with GARLI, not taking into account dates of isolation, but crucially it had a posterior probability of one during Bayesian Inference (BI) that do take into account dates of isolation. Although we cannot rule out systematic topological errors, BI was certainly better informed than ML, since RNA viruses evolve fast, making their times of isolation important parameters for phylogenetic inference. Moreover, since we did not find compositional or codon usage biases in those sequences and in agreement with the consistent BI results, we could not rule out that the long branch length observed was not due to a detected increase of almost 10 fold increase in the rate of change along that lineage, which was not caused by detectable positive selection, as evaluated using HyPhy.

Selection analyses

Selection analyses of E and NS5 genes uncovered several sites under strong negative selection indicated by $\omega < 0$. This suggests frequent purging of deleterious polymorphisms in functionally important genes. Likewise, the lack of positively selected sites, indicated by $\omega > 0$, is typical of highly adapted phenotypes and shows no detectable directional change on the available data. Our

findings were expected, as the infection and transmission modes of ZIKV allow the accumulation of synonymous mutations and negatively selected sites [67]. The alternation between arthropod vector and mammal hosts may impose several barriers to non-synonymous mutations in important genes [68].

Phylogenetic analyses

The μ and the highest posterior densities (HPD with 95% of confidence interval) estimated with Beast for E, NS5 and 3'NCR genomic regions were, respectively, 2.135E-3 (2.04E-3 to 2.33221E-3), 7.1789E-4 (6.9466E-4 to 7.417E-4) and 1.1285E-3 (2.708E-4 to 2.504E-3) substitutions per site per year, which are similar to μ estimated other flaviviruses [69]. As evolutionary rates are the result of spontaneous mutations followed by selection, differences *per gene* are expected and in accordance with their biological role, given that the NS5 is a polymerase and the E is a surface protein. In addition, the root date estimates and 95% HPDs of the phylogenetic trees for E, NS5 and 3'NCR genomic regions were, respectively, 1900 (1851 to 1937), 1927 (1887 to 1940) and 1923 (1874 to 1959). These dates suggest a recent origin for the ZIKV strains (included in this study) near to the beginning of the 20th century.

Movement of ZIKV

Based on our samples we inferred the most likely geographical pathway connecting ZIKV lineages. These results indicated that ZIKV emerged in Uganda around 1920, most probably between 1892 and 1943. This inference is in line with the first known ZIKV isolation in Uganda in 1947 [5]. We found two independent ZIKV introductions into West Africa from the Eastern portion of the continent (Figures 2 and S2A, and kml file in Dataset S1). The first viral introduction into Côte d'Ivoire (CI) and Senegal (SN) was related to the MR766 cluster (yellow lines in Figure 2), which possibly moved from Uganda around 1940 into Dezidougou (CI). From there, this lineage probably spread to Kedougou in Senegal (SN) around 1985 and to Sokala-Sobara (CI) around 1995. The second introduction was related to a Nigerian cluster (green lines in Figure 2), when ZIKV probably moved from Uganda to the Central African Republic and Nigeria around 1935. From Nigeria, the virus probably spread to Saboya (SN) around 1950 and from there to Dezidougou (CI) and Bandia (SN) around 1960. From Bandia, ZIKV was introduced into Kedougou (SN) around 1965 and from there to Burkina Faso around 1980 and to Dakar (SN) around 1985. Moreover, an additional ZIKV lineage from Uganda probably spread to Malaysia around 1945 and from there, the virus reached Micronesia around 1960, forming the Asian cluster [26]. The correlation between viral location (coded as character states) and phylogenies was strongly supported by significant AI and PS values, p -values ≤ 1.00 E-4 (Dataset S2). Thus, assuming an origin of ZIKV in Uganda, our findings revealed at least two independent exits from East Africa in agreement with the two previously proposed African clades [17] and also pointed to a viral migratory flow from Eastern Africa to Asia. Although our sampling was the most comprehensive to this date, our conclusions about ZIKV movement are informed conjectures at best on the most plausible hypotheses on ZIKV spreading patterns, which are limited by the inherent biases of this type of analyses.

Animal sources of ZIKV

The association of the animal sources with viral lineages (Figure S2B) suggested that ZIKV dispersed widely among distinct animal species without a clear pattern of preference, maybe associated to the enzootic cycle of ZIKV in Africa, whose natural cycle allows a

Figure 1. Maximum likelihood phylogenetic tree inferred for concatenated of sequences from Envelope and NS5 genes of Zika virus. Consensus tree summarized after 1000 non-parametric bootstrap replicates, with support values greater than 60% shown in the nodes. The cluster the Ugandan MR766 prototype strain was highlighted by the yellow sector and the Nigerian cluster was highlighted by the green sector. The strains from Senegal and Côte d'Ivoire are shown in green and orange, respectively. The tree has been rooted with the Spondweni lineage isolated in South Africa used as outgroup to root the tree.
doi:10.1371/journal.pntd.0002636.g001

broad range of hosts [70]. Nevertheless, we found significant MC (p -value $\approx 1.00 \times 10^{-4}$, Dataset S2) for ZIKV strains isolated from *A. dalzielii*, suggesting a possible important role of this zoophilic vector [71] in West Africa. This association was found to be robust to the exclusion of vertebrate host from the analysis. The plausibility of the putative recombination events we detected (Table S2), could in part be explained by mosquitoes taking sequential blood meals from distinct animal species harboring distinct ZIKV lineages, which is in line with ours and others host range findings [70]. Also, when analyzing the increase of ZIKV activity in Kedougou, (where most of the viruses analyzed herein were collected), we

noticed that ZIKV activity is much more frequent, with an interval of 1–2 years, compared to the 5 to 8 years cycle of dengue and yellow fever virus. Hence from 1972 to 2002, ZIKV emerged in 1973, 1976, 1979, 1980 and 1981. Such frequent activity can also be an opportunity of co-circulation and mixing of multiple genotypes present in the forest and that may favor recombination among them.

A phylodynamic context for recombination events

The occurrence of recombination among ZIKV strains in time-scaled phylogenetic trees suggested that some ZIKV lineages

Figure 2. Geographic spread of ZIKV in Africa and Asia. The directed lines connect the most probable sources and target localities of viral lineages (shown by arrows), with widths proportional to the posterior probabilities and values shown in red. Only plausible routes with probabilities above 50% are shown. The distinct introductions into Senegal and Côte d'Ivoire were represented by different colors. The estimated time to the most recent common ancestor of strains from different countries are shown with 95% posterior time intervals in parenthesis and could be interpreted as the oldest possible year of introduction of that lineage at that locality.
doi:10.1371/journal.pntd.0002636.g002

Figure 3. Mapping of predicted glycosylation sites on envelope protein of ZIKV. A) Alignment of E protein showing predicted glycosylation sites. Red arrows point to O-linked glycosylation sites (Ser or Thr residues) and the yellow arrow points to the N-linked glycosylation site (Asn-X-Thr motif). B) Tridimensional structure of E protein. Red beads indicate O-linked glycosylation sites and the yellow bead indicates the unique N-linked glycosylation site.

doi:10.1371/journal.pntd.0002636.g003

sampled in Dezidougou (CI) in 1990 (ArA27101, ArA27290, ArA27096, ArA27443, ArA27407 and ArA27106) with recombinant E (Figure S2C) shared a common ancestor around 1962 (ranging from 1951 to 1967 HPD with 95% of confidence interval). Likewise, the strain ArA982 was also isolated at Dezidougou in 1999 and its sister-group ArA986, which shared a common ancestor with the former around 1992 (ranging from 1981 to 1996 HPD with 95% of confidence interval), was sampled in the neighbor province Sokala-Sobara (CI) in 1999. Together these results indicated that recombination in envelope protein could be an important trend among the enzootic cycle of ZIKV at this region in Côte d'Ivoire, as ZIKV lineages did not show a clear pattern of host preference and recombination requires the infection of the same host by more than one viral strain. Besides, the other E recombinant strain (HD78788), isolated from a human case at Dakar (SN) in 1991, shared a common ancestor around 1984 (ranging from 1976 to 1988 HPD with 95% of confidence interval) with ZIKV strains from Kedougou (SN). Conversely, the three NS5 recombinants did not cluster along phylogenetic trees (Figure S2C), although two of them were isolated in Kedougou from *A. daltzeili* mosquitoes in 2001 (ArD157995 and ArD158084) and the other (ArB1362) was isolated in Bouboui, Central African Republic, from *A. africanus* mosquitoes in 1968. The preferential distribution of recombinant strains along phylogenies was supported by significant *p*-values of AI and PS $\leq 2.00E-4$ (Dataset S2) and the adjacency patterns of E and NS5 recombinants were also confirmed by MC statistics (Dataset S2).

Glycosylation patterns in ZIKV envelope protein

Our analyses predicted several glycosylation sites in the E protein (Figure 3). We detected a probable motif (Asn-X-Thr) among E sequences from several ZIKV strains (Figure 3A), which suggests a N-linked glycosylation site in the residue Asn-154, in agreement with [17,26]. This residue is located on an α -helix in the E protein structure (yellow arrow in Figure 3A and yellow bead in Figure 3B). Our results also pointed several O-linked glycosylation sites in the E protein (red arrows in Figure 3A and red beads Figure 3B) but no C-mannosylated site. We found a probable mucin-type O-linked glycosylated site at residue Thr-170 in E protein from all ZIKV strains, and other mucin sites at residues Thr-245 and Thr-381 in some isolates (Figure 3A). In addition, we also uncovered probable O-GlcNAc attachment sites at residues Ser-142, Ser-227, Thr-231, Ser-304, Thr-366 and Thr-381 in E from some strains (Figure 3A).

Given the importance of the N-linked glycosylation site around position 154 of the E protein for infectivity and assembly of flaviviruses [72–74] and the fact that we observed polymorphisms in this motif (deletions and substitutions 156 Thr/Iso), we investigated the correlation between the conservation of this motif (Asn-X-Thr) and phylogenies for ZIKV strains. Our results suggested that the acquisition of this glycosylation site is a recurrent event in the history of ZIKV, given the observed changes from Isoleucine to Threonine and vice-versa more than once in the MCC tree (Figure S2D), supported by *p*-values for AI and PS $\leq 7.00E-4$ (Dataset S2). However, our conclusions are limited due to serial passages of the former ZIKV strains (Figure S2D) in mouse brain [26], which could result in the loss of this glycosylation site, as observed in West Nile virus [75].

Correlated evolutionary change along ZIKV phylogenies

Since it was demonstrated that the absence of an N-linked glycosylation site on the E protein enhances viral infectivity for C6/36 mosquito cells [72,73] and E protein of ZIKV strains from *A. daltzeili*, which was the unique vector source with significant MC–showed an absence of this glycosylation site, we investigated the correlation between this mosquito-source and N-linked glycosylation patterns of E protein along PST. Our results indicated the changes in glycosylation patterns (presence or absence) and vector (*A. daltzeili* or not) were correlated during ZIKV emergence, which was supported by BF for dependent model (BF ≈ 47.004) greater than for them to independent model. These findings could be related to the enzootic cycle of ZIKV in West Africa and the zoophilic behavior of *A. daltzeili* [71], whose females take blood meals from a broad range of vertebrates, which provides additional evidence for the absence of host preference (as described in Animal sources of ZIKV). Hence, further studies are necessary to understand the consequences of our results to ZIKV transmission cycle in nature.

Biological correlates of our findings

Our analyses indicated that ZIKV may have experienced several recombination events, which is uncommon among flaviviruses [66]. The recurrent loss and gain of the N-linked glycosylation site in the E protein could be related to mosquito-cell infectivity [73] and the correlated loss of this glycosylation site in ZIKV strains from *A. daltzeili* also provides indirect evidence for the enzootic cycle, since this vector has a zoophilic behavior [71] that may spread ZIKV among several hosts. Crucially, our results corroborated the notion that at least three distinct ZIKV clusters shared a common ancestor possibly with Ugandan lineages around 1920, followed by two events of East to West Africa spread (Figure 2): (i) one related to the MR766 cluster introduction to Côte d'Ivoire and posterior spread to Senegal and; (ii) other related to the Nigerian cluster introduction in Senegal and posterior dispersion to Côte d'Ivoire and Burkina Faso.

Supporting Information

Dataset S1 Spread of ZIKV strains in Africa and Asia. A kml file to picture the history of ZIKV movement into Africa and Asia during the time, it is executable in Google Earth program (<http://www.google.com/earth/>). (KML)

Dataset S2 Significance of the correlation among phylogenies and attributes of ZIKV lineages. (DOC)

Figure S1 Recombination analysis using Rec-HMM along ZIKV genomes. The dashed green lines indicate estimated breakpoints in the genomes. (TIF)

Figure S2 Maximum clade credibility (MCC) trees for concatenated sequences summarizing lineage states along a time-scaled tree, with posterior probability values shown near the nodes. (A) Most probable geographical location coded according to map (Figure 2): Uganda (UG), Central African Republic (CF), Dezidougou in Côte d'Ivoire (DE), Sokala-Sobara in Côte d'Ivoire (SS), Kedougou in Senegal (KE), Saboya

in Senegal (SA), Bandia in Senegal (BA), Dakar in Senegal (DA), Burkina Faso (BF), Nigeria (NG), Malaysia (MY) and Yap Island in the Federated States of Micronesia (FM); (B) most probable animal source; (C) recombination events per region; and (D) glycosylation polymorphisms.

(TIIF)

Table S1 Source, country and year of isolation from ZIKV strains used in this study.

(DOC)

Table S2 Detection of recombination events in ZIKV genomes.

(DOC)

References

- Hayes EB (2009) Zika virus outside Africa. *Emerg Infect Dis* 15: 1347–1350. doi:10.3201/eid1509.090442.
- Simpson DI (1964) Zika virus infection in man. *Trans R Soc Trop Med Hyg* 58: 335–338.
- Bearcroft WG (1956) Zika virus infection experimentally induced in a human volunteer. *Trans R Soc Trop Med Hyg* 50: 442–448.
- Foy BD, Kobylinski KC, Chilson Foy JL, Blitvich BJ, Travassos da Rosa A, et al. (2011) Probable non-vector-borne transmission of Zika virus, Colorado, USA. *Emerg Infect Dis* 17: 880–882. doi:10.3201/eid1705.101939.
- Dick GWA, Kitchen SF, Haddock AJ (1952) Zika virus. I. Isolations and serological specificity. *Trans R Soc Trop Med Hyg* 46: 509–520.
- Fagbamigbe AH (1979) Zika virus infections in Nigeria: virological and seroepidemiological investigations in Oyo State. *J Hyg (Lond)* 83: 213–219.
- Robin Y, Mouchet J (1978) Serological and entomological study on yellow fever in Sierra Leone. *Bull Soc Pathol Exot Filiales* 68: 249–258.
- Jan C, Languillat G, Renaudet J, Robin Y (1978) A serological survey of arboviruses in Gabon. *Bull Soc Pathol Exot Filiales* 71: 140–146.
- McCrae AW, Kirya BG (1982) Yellow fever and Zika virus epizootics and enzootics in Uganda. *Trans R Soc Trop Med Hyg* 76: 552–562.
- Saluzzo JF, Gonzalez JP, Hervé JP, Georges AJ (1981) Serological survey for the prevalence of certain arboviruses in the human population of the south-east area of Central African Republic. *Bull Soc Pathol Exot Filiales* 74: 490–499.
- Moulin E, Zeller H, Le Guenno B, Traoré-Lamizana M, Hervé JP, et al. (1993) Surveillance of the circulation of arbovirus of medical interest in the region of eastern Senegal. *Bull Soc Pathol Exot* 86: 21–28.
- Akoua-Koffi C, Diarra-Souba S, Béné VB, Ngichji JM, Bazoua T, et al. (2001) Investigation surrounding a fatal case of yellow fever in Côte d'Ivoire in 1999. *Bull Soc Pathol Exot* 94: 227–230.
- Darwish MA, Hoogstraal H, Roberts TJ, Ahmed IP, Omar F (1983) A sero-epidemiological survey for certain arboviruses (Togaviridae) in Pakistan. *Trans R Soc Trop Med Hyg* 77: 442–445.
- Marchette NJ, Garcia R, Rudnick A (1969) Isolation of Zika virus from *Aedes aegypti* mosquitoes in Malaysia. *Am J Trop Med Hyg* 18: 411–415.
- Olson JG, Ksiazek TG, Suhandiman, Trivibowo (1981) Zika virus, a cause of fever in Central Java, Indonesia. *Trans R Soc Trop Med Hyg* 75: 389–393.
- Duffy MR, Chen TH, Hancock WT, Powers AM, Kod JL, et al. (2009) Zika virus outbreak on Yap Island, Federated States of Micronesia. *N Engl J Med* 360: 2536–2543. doi:10.1056/NEJMoa0805715.
- Lancioti RS, Kosoy OL, Laven JJ, Velez JO, Lambert AJ, et al. (2008) Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. *Emerg Infect Dis* 14: 1232–1239.
- Heang V, Yasuda CY, Sovann L, Haddock AD, Travassos da Rosa AP, et al. (2012) Zika virus infection, Cambodia, 2010. *Emerg Infect Dis* 18: 349–351. doi:10.3201/eid1802.111224.
- Chambers TJ, Hahn CS, Galler R, Rice CM (1990) Flavivirus genome organization, expression, and replication. *Annu Rev Microbiol* 44: 649–688. doi:10.1146/annurev.mi.44.100190.003245.
- Kuno G, Chang G-J (2007) Full-length sequencing and genomic characterization of Bagaza, Kedougou, and Zika viruses. *Arch Virol* 152: 687–696. doi:10.1007/s00705-006-0903-z.
- Lindenbach BD, Rice CM (2003) Molecular biology of flaviviruses. *Adv Virus Res* 59: 23–61.
- Kuno G, Chang GJ, Tsuchiya KR, Karabatos N, Cropp CB (1998) Phylogeny of the genus *Flavivirus*. *J Virol* 72: 73–83.
- Gould EA, de Lamballerie X, Zanotto PM, Holmes EC (2003) Origins, evolution, and vector/host coadaptations within the genus *Flavivirus*. *Adv Virus Res* 59: 277–314.
- Gaunt MW, Sall AA, de Lamballerie X, Falconar AK, Dzhanian TI, et al. (2001) Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography. *J Gen Virol* 82: 1867–1876.
- Zanotto PM, Gould EA, Gao GF, Harvey PH, Holmes EC (1996) Population dynamics of flaviviruses revealed by molecular phylogenies. *Proc Natl Acad Sci U S A* 93: 548–553.
- Haddock AD, Schuh AJ, Yasuda CY, Kasper MR, Heang V, et al. (2012) Genetic Characterization of Zika Virus Strains: Geographic Expansion of the Asian Lineage. *PLoS Negl Trop Dis* 6: e1477. doi:10.1371/journal.pntd.0001477.
- Digoutte JP, Calvo-Wilson MA, Mondo M, Traoré-Lamizana M, Adam F (1992) Continuous cell lines and immune ascitic fluid pools in arbovirus detection. *Res Virol* 143: 417–422.
- Faye O, Faye O, Dupressoir A, Weidmann M, Ndiaye M, et al. (2008) One-step RT-PCR for detection of Zika virus. *J Clin Virol* 43: 96–101. doi:10.1016/j.jcv.2008.05.005.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792–1797. doi:10.1093/nar/gkh340.
- Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 27: 221–224. doi:10.1093/molbev/msp259.
- Martin DP, Lemey P, Lott M, Moulton V, Posada D, et al. (2010) RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26: 2462–2463. doi:10.1093/bioinformatics/btq467.
- Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16: 562–563.
- Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. *Virology* 265: 218–225. doi:10.1006/viro.1999.0056.
- Posada D, Randall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci U S A* 98: 13757–13762. doi:10.1073/pnas.241370698.
- Smith JM (1992) Analyzing the mosaic structure of genes. *J Mol Evol* 34: 126–129.
- Martin DP, Posada D, Grandall KA, Williamson C (2005) A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res Hum Retroviruses* 21: 98–102. doi:10.1089/aid.2005.21.98.
- Gibbs MJ, Armstrong JS, Gibbs AJ (2000) Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 16: 573–582.
- Boni MF, Posada D, Feldman MW (2006) An Exact Nonparametric Method for Inferring Mosaic Structure in Sequence Triplets. *Genetics* 176: 1035–1047. doi:10.1534/genetics.106.068874.
- Bland JM, Altman DG (1995) Multiple significance tests: the Bonferroni method. *BMJ* 310: 170.
- Westesson O, Holmes I (2009) Accurate detection of recombinant breakpoints in whole-genome alignments. *PLoS Comput Biol* 5: e1000318. doi:10.1371/journal.pcbi.1000318.
- Nagarajan N, Kingsford C (2011) GiRaF: robust, computational identification of influenza reassortments via graph mining. *Nucleic Acids Res* 39: e34. doi:10.1093/nar/gkq1232.
- Ronequis F, Teslenko M, van der Mark P, Ayres DL, Darling A, et al. (2012) MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst Biol* 61: 539–42. doi:10.1093/sysbio/syt029.
- Strimmer K, von Haeseler A (1997) Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc Natl Acad Sci U S A* 94: 6815–6819.
- Schmidt H, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18: 502–504. doi:10.1093/bioinformatics/18.3.502.
- Campbell V, Legendre P, Lapointe FJ (2011) The performance of the Congruence Among Distance Matrices (CADM) test in phylogenetic analysis. *BMC Evol Biol* 11: 64. doi:10.1186/1471-2148-11-64.
- Zwickl D (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. The University of Texas at Austin.
- Sukumaran J, Holder MT (2010) DendroPy: a Python library for phylogenetic computing. *Bioinformatics* 26: 1569–1571. doi:10.1093/bioinformatics/btq228.

Acknowledgments

We thank Dr. Scott Weaver for helpful comments and critical reading of the manuscript.

Author Contributions

Conceived and designed the experiments: AAS OumF. Performed the experiments: OumF JVCdO. Analyzed the data: PMAZ CCMF AAS OumF AI. Contributed reagents/materials/analysis tools: AAS OumF PMAZ CCMF OusF AI JVCdO. Wrote the paper: OumF CCMF AAS PMAZ OusF MD AI JVCdO.

48. Pond SLK, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21: 676–679. doi:10.1093/bioinformatics/bti079.
49. Julenius K (2007) NetCGlyc 1.0: prediction of mammalian C-mannosylation sites. *Glycobiology* 17: 868–876. doi:10.1093/glycob/cwm050.
50. Julenius K, Mølgaard A, Gupta R, Brunak S (2005) Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology* 15: 153–164. doi:10.1093/glycob/cwh151.
51. Gupta R, Brunak S (2002) Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput* 310–322.
52. Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* 4: 1633–1649. doi:10.1002/pmic.200300771.
53. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234: 779–815. doi:10.1006/jmbi.1993.1626.
54. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26: 283–291. doi:10.1107/S0021889902009944.
55. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7: 214. doi:10.1186/1471-2148-7-214.
56. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4: e88. doi:10.1371/journal.pbio.0040088.
57. Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian phylogeography finds its roots. *PLoS Comput Biol* 5: e1000520. doi:10.1371/journal.pcbi.1000520.
58. Bielejec F, Rambaut A, Suchard MA, Lemey P (2011) SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics* 27: 2910–2912. doi:10.1093/bioinformatics/btr481.
59. Parker J, Rambaut A, Pybus OG (2008) Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infect Genet Evol* 8: 239–246.
60. Barker D, Pagel M (2005) Predicting functional gene links from phylogenetic-statistical analyses of whole genomes. *PLoS Comput Biol* 1: e3. doi:10.1371/journal.pcbi.0010003.
61. Suchard MA, Weiss RE, Sinsheimer JS (2001) Bayesian Selection of Continuous-Time Markov Chain Evolutionary Models. *Mol Biol Evol* 18: 1001–1013. doi:10.1093/oxfordjournals.molbev.a003872.
62. Posada D, Crandall K A (2002) The effect of recombination on the accuracy of phylogeny estimation. *J Mol Evol* 54: 396–402. doi:10.1007/s00239-001-0034-9.
63. Schierup MH, Hein J (2000) Consequences of recombination on traditional phylogenetic analysis. *Genetics* 156: 879–891.
64. Wiens JJ (2003) Missing Data, Incomplete Taxa, and Phylogenetic Accuracy. *Syst Biol* 52: 528–538. doi:10.1080/10635150390218330.
65. Wiens JJ (2006) Missing data and the design of phylogenetic analyses. *J Biomed Inform* 39: 34–42. doi:10.1016/j.jbi.2005.04.001.
66. Simon-Lorière E, Holmes EC (2011) Why do RNA viruses recombine? *Nat Rev Microbiol* 9: 617–626. doi:10.1038/nrmicro2614.
67. Hanada K, Suzuki Y, Gojobori T (2004) A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. *Mol Biol Evol* 21: 1074–1080. doi:10.1093/molbev/msh109.
68. Holmes EC (2003) Patterns of intra- and interhost nonsynonymous variation reveal strong purifying selection in dengue virus. *J Virol* 77: 11296–11298.
69. Twiddy SS, Holmes EC, Rambaut A (2003) Inferring the Rate and Time-Scale of Dengue Virus Evolution. *Mol Biol Evol* 20: 122–129. doi:10.1093/molbev/mng010.
70. Kuno G, Chang G-JJ (2005) Biological transmission of arboviruses: reexamination of and new insights into components, mechanisms, and unique traits as well as their evolutionary trends. *Clin Microbiol Rev* 18: 608–637. doi:10.1128/CMR.18.4.608-637.2005.
71. Diallo M, Thonnin J, Traore-Lamizana M, Fontenille D (1999) Vectors of Chikungunya virus in Senegal: current data and transmission cycles. *Am J Trop Med Hyg* 60: 281–286.
72. Hanna SL, Person TC, Sanchez MD, Ahmed AA, Murtadha MM, et al. (2005) N-linked glycosylation of west nile virus envelope proteins influences particle assembly and infectivity. *J Virol* 79: 13262–13274. doi:10.1128/JVI.79.21.13262-13274.2005.
73. Lee E, Leang SK, Davidson A, Lobigs M (2010) Both E protein glycans adversely affect dengue virus infectivity but are beneficial for virion release. *J Virol* 84: 5171–5180. doi:10.1128/JVI.01900-09.
74. Mondotte JA, Lozach PY, Amara A, Gamarnik A V (2007) Essential role of dengue virus envelope protein N glycosylation at asparagine-67 during viral propagation. *J Virol* 81: 7136–7148. doi:10.1128/JVI.00116-07.
75. Chambers TJ, Hakey M, Nestorowicz A, Rice CM, Lustig S (1998) West Nile virus envelope proteins: nucleotide sequence analysis of strains differing in mouse neuroinvasiveness. *J Gen Virol* 79 (Pt 10): 2375–2380.

Molecular phylogeography of tick-borne encephalitis virus in central Europe

Manfred Weidmann,¹ Stefan Frey,² Caio C. M. Freire,³ Sandra Essbauer,² Daniel Růžek,^{4,5} Boris Klempa,^{6,7} Dana Zubrikova,⁸ Maria Vögel,⁹ Martin Pfeffer,¹⁰ Frank T. Hufert,¹ Paolo M. Zlotto³ and Gerhard Dobler²

Correspondence
Manfred Weidmann
mweidma@gwdg.de

¹Department of Virology, University Medical Center, 37075 Göttingen, Germany

²Bundeswehr Institute of Microbiology, 80937 Munich, Germany

³Department of Microbiology, Biomedical Sciences Institute – ICB II University of São Paulo, 05508-000 São Paulo, Brazil

⁴Department of Virology, Veterinary Research Institute, Hudcova 70, CZ-62100 Brno, Czech Republic

⁵Institute of Parasitology, Biology Centre of the Czech Academy of Sciences, Branisovska 31, CZ-37005 Ceske Budejovice, Czech Republic

⁶Institute of Virology, Slovak Academy of Science, Dubravska cesta 9, 84505 Bratislava, Slovakia

⁷Institute of Virology Charité University Hospital, Berlin, Germany

⁸Institute of Parasitology, Slovak Academy of Science, Kosice, Slovakia

⁹Comparative Tropical Medicine and Parasitology, Faculty of Veterinary Medicine, Ludwig-Maximilians-University, Munich, Germany

¹⁰Institute of Animal Hygiene and Veterinary Public Health, University of Leipzig, Leipzig, Germany

In order to obtain a better understanding of tick-borne encephalitis virus (TBEV) strain movements in central Europe the E gene sequences of 102 TBEV strains collected from 1953 to 2011 at 38 sites in the Czech Republic, Slovakia, Austria and Germany were determined. Bayesian analysis suggests a 350-year history of evolution and spread in central Europe of two main lineages, A and B. In contrast to the east to west spread at the Eurasian continent level, local central European spreading patterns suggest historic west to east spread followed by more recent east to west spread. The phylogenetic and network analyses indicate TBEV ingressions from the Czech Republic and Slovakia into Germany via landscape features (Danube river system), biogenic factors (birds, red deer) and anthropogenic factors. The identification of endemic foci showing local genetic diversity is of paramount importance to the field as these will be a prerequisite for in-depth analysis of focal TBEV maintenance and long-distance TBEV spread.

Received 18 April 2013
Accepted 13 June 2013

INTRODUCTION

In the past decade Europe has experienced an increase in tick-borne encephalitis (TBE) incidence, with more than 3400 human TBE cases annually and more than 3600 human TBE cases in Russia in 2009 (Donoso Mantke *et al.*, 2011). Knowledge of the key factors involved in spread and spatiotemporal diversification of the causative viral agent, tick-borne encephalitis virus (TBEV), however, is limited. Field work on the focal activity of TBEV in ticks in central Europe (Labuda *et al.*, 2002) indicates dependence on microclimate, humidity and the 8 °C isotherm (Labuda

et al., 2002; Randolph *et al.*, 2000) and coincidence of tick and host population densities (Cagnacci *et al.*, 2012; Carpi *et al.*, 2008; Randolph *et al.*, 1999; Vor *et al.*, 2010). Several efforts are under way to analyse the tick and animal reservoirs involved in the complex three-host transmission cycle as the tick moults from larva to nymph to adult with a blood meal involved in each step (Achazi *et al.*, 2011; Cagnacci *et al.*, 2012; Klaus *et al.*, 2010; Knap *et al.*, 2012; Labuda & Nuttall, 2004).

As a member of the genus *Flavivirus* of the family *Flaviridae* the genome of TBEV consists of a single-stranded 11 kb (+) RNA coding for one large polyprotein yielding three structural proteins (C, prM, E) and seven non-structural proteins (NS1, 2a, 2b, 3, 4a, 4b, 5)

Three supplementary figures, one supplementary table and one supplementary file are available with the online version of this paper.

(Chambers *et al.*, 1990). It is commonly accepted that the envelope protein (E) plays a major role for the infection of mammalian and tick cells, and that it is the main determinant of virulence and a main target for neutralizing antibodies (Aberle *et al.*, 1999). A recent study looking at the evolutionary rates of two TBEV isolates from the same focus in Finland sampled 44 years apart concluded that the E gene in particular is subject to purifying selection. In spite of high immune pressure it is nevertheless the most constrained gene in the TBEV genome, thereby being a prime target for nucleotide-based phylogenetic analysis (Uzcátegui *et al.*, 2012).

In Germany, TBE infections are mainly reported from the southern federal states of Baden-Württemberg and Bavaria (SurvStat, 2012). Since 2005, detailed data on circulating viruses and corresponding sequences have been under intensive investigation in several foci in Bavaria (Klaus *et al.*, 2010; Kupča *et al.*, 2010; Weidmann *et al.*, 2011). A previously published analysis of 61 central European TBEV-E gene sequences generated from collected ticks unveiled a general east to west spread of TBEV within a small geographical range (Weidmann *et al.*, 2011; Zanotto *et al.*, 1995). On this small scale, a phylogenetic connection between one strain from Nova Rise in the Vysocina region of the Czech Republic (T-730) and one from Germany at Haselmühl (277 km apart) separated by the mountain range of the Bavarian forest (1400 m) implicated a non-continuous distribution pattern and possibly some anthropogenic influence in the TBEV spread. The genetic analysis of the new dataset of 102 TBEV sequences allowed in-depth analysis of possible spread patterns due to landscape features, biogenic and anthropogenic influences of TBEV in central Europe.

RESULTS

Dated tree analysis

In order to obtain a better understanding of TBEV strain spread patterns in central Europe the E gene sequences of an additional 41 TBE viruses collected from 1953 to 2011 from Slovakia (three sites), Austria (two sites) and Germany (14 sites) were determined. This included 26 new TBEV isolates from ticks collected in south-east Germany from 2010 to 2011 (Fig. 1a). In sum, the extended dataset allowed analysis of 102 TBEV-E genes from 38 sites (Table S1, available in JGV Online). The comparison of two alternative schemes for estimating the rate of change (μ) expressed in substitutions per site per year (s^{-1}/y) indicated that a low rate of $4.07 \times 10^{-5} s^{-1}/y$, within the 95% highest posterior probability density (HPD) interval from 4.05×10^{-5} to $4.08 \times 10^{-5} s^{-1}/y$ had a slightly better Bayes factor (BF) of 1.635 compared with a higher rate of $1.69 \times 10^{-4} s^{-1}/y$, with 95% HPD from 1.09×10^{-4} to 3.37×10^{-4} , suggestive of a TBEV history time frame spanning more than 300 years in central Europe (mean=314 and 95% HPD ranging from 261 to 365

years). The rates (μ) that we obtained are in line with previous estimates showing slow evolution of tick-borne viruses compared with mosquito-borne flavivirus (Twiddy *et al.*, 2003; Zanotto *et al.*, 1996). Moreover, our results indicated that the TBEV lineages are split into two main clades, designated clade A and B (Figs S1, S2, Fig. 1b) and experienced purifying selection in the E gene (Table 1).

TBEV spread patterns

Phylogeographic estimates revealed that discrete transitions explain TBEV spread much better than continuous diffusion, since discrete models had significantly higher likelihood (L) than continuous models. BF values did not indicate significant differences between the L from continuous models (Brownian diffusion and Cauchy distribution) and the L from discrete models (Table 2). Moreover, Mantel's test for correlation among matrices of patristic and geographical distances indicated that the relationship between them is significant ($P=9.9999 \times 10^{-6}$). The Mantel's correlogram lacked homogeneity across all geographical distances (Fig. 2) and significant positive correlation (solid squares above the zero line) was found only for distances less than 100 km. Correlation was significantly negative for distances between 100 and 200 km and positive again close to 300 km.

TBEV diversity

The study identified six sites with multiple ($n \geq 5$) E gene sequences. Their local diversity was analysed by bootstrapped maximum-likelihood trees (Fig. 3). TBEV-E nucleotide sequences at Zdar Kaplice ($n=7$, 1986–1987; Weidmann *et al.*, 2011), Borovany ($n=5$, 2000–2002), Burglengenfeld ($n=6$, 2010) and Heselbach ($n=5$, 2011) had low diversity with sequence identities as high as 99.6–99.9%, while those of sequences at Potepli ($n=7$, 1954–1967) and Haselmühl ($n=17$, 2009–2011) ranged from 94.8 to 99.8%. At Potepli and Haselmühl both lineage A and B strains were found.

Migratory routes

We analysed local migratory routes over time by estimating discrete state changes among isolation places for all ancestral nodes in time-scaled viral genealogies that were sampled during Markov Chain Monte Carlo (MCMC) convergence. Our results unveiled multiple migratory movements of TBEV from Germany to other places in the East during the past 300 years (Fig. 4, Table 3 and file S4). In the first half of the eighteenth century, multiple lineages of TBEV spread from Germany to Austria, Slovakia, North Bohemia and Central Bohemia. In the second half, the virus went from Germany to North Bohemia and from Central Bohemia to South Bohemia. From there, TBEV was reintroduced to Germany around 1930 and moved to North Moravia around 1970. Moreover, TBEV moved from North Bohemia to South Moravia around 1860.

Fig. 1. (a) Sampling sites: white circles, sites of previous study (Weidmann *et al.*, 2011); blue circles, extended set of sampling sites; light blue circles, new isolates at known sites. (b) Distribution of lineages A and B according to the maximum clade credibility (MCC) tree (Fig. S1). Yellow circles, lineage A; green circles, lineage B; circles with red rims, sites with TBEV genetic diversity (Fig. 3); mixed coloured circles, sites with lineage A and B strains. (c) Hypothesized migration routes of TBEV in the twentieth century. Country codes: D, Germany; A, Austria; CZ, Czech Republic; SK, Slovakia. NB, North Bohemia; CB, Central Bohemia; SB, South Bohemia.

Table 1. Codon sites under purifying selection detected in TBEV

Amino acid position*	ω †	P value
283	-4.80	7.09×10^{-3}
294	-4.08	1.24×10^{-2}
298	-5.00	4.12×10^{-3}
311	-3.00	3.70×10^{-2}
334	-4.00	1.23×10^{-2}
348	-3.00	3.70×10^{-2}
372	-4.71	7.09×10^{-3}
417	-9.13	3.65×10^{-2}
421	-3.00	3.70×10^{-2}
434	-3.53	2.27×10^{-2}
447	-4.00	1.23×10^{-2}
455	-3.00	3.70×10^{-2}
477	-3.00	3.70×10^{-2}
500	-4.71	6.44×10^{-3}
516	-5.88	1.82×10^{-3}
533	-3.53	2.27×10^{-2}
551	-3.00	3.70×10^{-2}
582	-5.00	4.12×10^{-3}
592	-4.71	6.44×10^{-3}
603	-3.53	2.27×10^{-2}
604	-10.59	1.20×10^{-5}
616	-3.33	2.82×10^{-2}
618	-3.53	2.44×10^{-2}
660	-3.53	2.27×10^{-2}
676	-3.00	3.71×10^{-2}
683	-3.53	2.27×10^{-2}
718	-10.70	1.20×10^{-5}
719	-4.04	1.23×10^{-2}
732	-3.03	3.70×10^{-2}
735	-3.82	3.39×10^{-2}
763	-3.03	3.70×10^{-2}

*The amino acid position was related to the first codon of polyprotein sequence of isolate Toro-2003 (GenBank ABD62793.3.).
†Difference between non-synonymous (d_N) and synonymous (d_S) rates per site.

Median joining (MJ) network analysis

Since linear inheritance patterns of viral sequences are amenable to MJ network analysis, we used this approach to

test if the TBEV strains might have migrated to south-east Germany by bypassing the Bavarian or Sumava mountains to the south. The resulting MJ network shows a discontinuous pattern of lineage A and B strains spread across central Europe and several introductions of TBEV strains into Germany (Fig. 5) and allowed spreading patterns to be deduced in great detail.

DISCUSSION

Phylogenetic analysis

The analysis of the extended dataset of TBEV-E gene sequences yielded a dated tree spanning more than 300 years and consisting of mainly two subclades designated lineages A and B. About half (48) of the strains analysed are from Germany (mainly Bavaria, Fig. 1a). Lineage A strains branching in a shallow forked subclade, suggest recent introduction at Haselmühl (Fig. 3, Fig. S1). Indeed in both clades, deeply forked subclades intermingle with shallow forked subclades, and could be explained by fairly recent evolutionary events, such as bottlenecks, pruning older lineages from both main clades, a phenomenon also described for mosquito-borne viruses (Zanotto *et al.*, 1996). Moreover, we identified several sites under purifying selection in the E gene (Table 1) and because of their potential bias towards underestimation of branch lengths on time-calibrated phylogenetics, we employed a substitution model with gamma correction (GTR + Γ) to better estimate hidden variation (Wertheim & Kosakovsky Pond, 2011). Nevertheless, patterns of molecular evolution from populations under strong purifying selection were shown to be very similar under simulated conditions to those of populations under neutral selection (Nicolaisen & Desai, 2012; Walczak *et al.*, 2012).

An additional analysis of patristic and geographical distances suggested that TBEV foci appear to cohere below a radius of 100 km, while continuous diffusion breaks apart for greater distances. This leads to suggest two distinct modes of dispersion of TBEV: (i) on a local scale TBEV disperses continuously with positive correlation between low genetic and geographical distances, and (ii) on a regional scale (distances greater than 100 km) TBEV spreads through migration events, since genetic and geographical distances

Table 2. BF values for comparisons among different models used during phylogeographic inferences

	lnP (model/data)	sem	BSSVS*	CTMC†	Brownian diffusion	Cauchy distribution
BSSVS	-5408.819	+/- 0.402	0	-1.062	11.679	11.478
CTMC	-5407.758	+/- 0.348	1.062	0	12.74	12.539
Brownian diffusion	-5420.498	+/- 0.999	-11.679	-12.74	0	-0.201
Cauchy distribution	-5420.297	+/- 0.879	-11.478	-12.539	0.201	0

†CTMC, Continuous time Markov chain.
*BSSVS, Bayesian stochastic search variable selection.

Fig. 2. Mantel's correlogram of patristic and geographical distances. Geographical distances (horizontal axis) plotted against Spearman's rank correlation coefficient (vertical axis). Zero line, no relationship between geographical and patristic distances. Points above the zero line, positive correlation; points below the zero line, negative correlation. Points with statistical significance ($P < 0.05$) are marked with solid (black) squares.

have a negative correlation. This observation is strengthened by the identification of six genetically diverse TBEV foci in total with at least two evolutionary rates and two foci with mixed lineages (Fig. 3).

The migration analysis indicates TBEV spread from Germany to other areas in Europe early in the eighteenth century and it was carried back to it in the twentieth century (Fig. 4). As recently confirmed, the main movement of TBEV appears to follow a clinal distribution oriented primarily from east to west (Heinze *et al.*, 2012; Zanotto *et al.*, 1995). Nevertheless, at a fine grain during short spans of space and time, rather complex spread patterns are expected to emerge. Considering the complex history of this central European region with trade, traffic, economic development and wars going on for centuries, these migrations may only be understandable in greater detail by using larger datasets, possibly comprising more comprehensive viral sequence information and a larger sample.

Landscape-dependent spread patterns

The combination of geographical and genetic information in the MJ network allowed for the analysis of detailed spreading patterns of TBEV. The introduction of one set of lineage A strains to Passau and onward to Amberg-Sulzbach directly linked to strains from Slovakia supports the idea of TBEV strains being introduced to Germany by circumventing the Bavarian mountains to the south (Fig. 1c) via the Danube river system.

Another set of lineage A strains appears to have migrated southward from the Heselbach and Burglengenfeld area in the northern part of south-east Germany. All sites are in proximity to tributaries to the Danube, or in the respective river plains or valleys [Heselbach (Naab), Burglengenfeld (Naab), Renholding (Geißa), Fürstenstein (Ilz), Rosenheim (Inn)].

As humidity and appropriate temperature are important for the life cycle of *Ixodes ricinus* (Randolph & Storey, 1999), it seems very likely that TBEV-infected ticks could easily establish foci along river valleys independently of the means of migration. *I. ricinus* and *Dermacentor reticulatus* have been shown to occur sympatrically (Silaghi *et al.*, 2012), and the latter is currently spreading from east to west along river valleys in Poland and from south to north along the corridor of the Dyje and Morava rivers in the Czech Republic (Široký *et al.*, 2011; Zygner *et al.*, 2009). The Morava flows to the south, directly passing TBEV isolation sites at Zahorska Ves and Malacky in Slovakia to meet the Danube. We therefore hypothesize that the Danube river system supports spread and establishment of new TBEV foci by providing favourable climatic conditions.

Biogenic spread patterns

The migration of lineage A strains towards the south of south-east Bavaria extends directly from strains described in investigations following an emergence of TBE cases in the Usti nad Labem (UNL) region (North Bohemia) in the 1990s (Figs 1b, 5; Weidmann *et al.*, 2011).

A recent comparative study on network calculation methods pointed out that parsimony approaches are less error-prone than MJ methods, with genes showing a high substitution rate as is the case for TBEV gene (Woolley *et al.*, 2008). We therefore also performed a parsimony analysis (Fig. S3). It confirmed the ingress patterns into Germany and particularly the ingress from North Bohemia observed in the MJ analysis.

Public health data in the Czech Republic indicate a general spread of TBE northward and westward (Daniel *et al.*, 2011; Kriz *et al.*, 2012). After eruption of TBE at the end of the 1990s in the UNL region, incidence increased first in the Kralovy Vary region (2001–2010) to the west and then even further west in south-east Bavaria (2001–2012; SurvStat, 2012). The regions Plzen and South Bohemia with the highest morbidity rates in the Czech Republic (10.9 and 23.2 per 10 000) cover the whole remaining expanse of the German–Czech border. Rodent and roe deer (*Capreolus capreolus*) ranges are too small to explain the spread of TBEV-infected ticks over long distances. Red deer (*Cervus elaphus*), however, have much larger resident ranges (<300 ha for females, <700 ha for males; Licoppe & Lievens, 2011) and stags that cannot acquire a herd of females migrate for long distances (<100 km) looking for a herd elsewhere. A population analysis of red deer populations in Sumava Park, on the eastern hillsides of

Fig. 3. TBEV-E gene diversity. Upper row: Starplot of TBEV-E gene sequences of isolates from Borovany, Burglengenfeld, Haselbach (one bar for all three sites 9.9×10^{-4} substitutions per site). Lower row: Haselmühl and Potepli (individual bars 1.0×10^{-2} substitutions per site). The bootstrap percentage values are derived from 1000 bootstrap trials. A and B indicate the two main lineages.

the Bavarian forest in Plzen and South Bohemia, provides some additional evidence for tick dispersal via red deer migration. During the cold war the red deer populations on either side of the Bavarian–Bohemian ecosystem were discreetly separated by border fortifications, and this still shows up in matrilineal mtDNA differentiation into predominantly west-European and Czech female red deer populations. Microsatellite-based population analyses, however, reveal that the populations have intermingled since the fortifications were removed two decades ago (Fickel *et al.*, 2012). The concomitant free migration of stags documented in these genetic signatures may have been a route of an accompanying ingress of TBEV-infested ticks into Germany.

In the MJ network, Central Bohemia itself acts as a sort of hub for introductions of TBEV lineage A strains to south-east Bavaria (Mühlendorf, Asbach) and in south-west Germany (Salem, Karlsruhe; Fig. 5). Prime candidate mechanisms for this type of spread could involve

migrations of passerine bird species, which have already been implicated in TBEV spread (Hoogstraal *et al.*, 1963; Waldenström *et al.*, 2007).

Anthropogenic spread patterns

Lineage B strains were introduced to Hauzenberg and Haselmühl. Introduction into Haselmühl had stood out at the end of our last phylogeographic analysis (Weidmann *et al.*, 2011) and the MJ network suggests a route from Brno via Nova Rise. This introduction was successful as evident from the many isolates from this site (Fig. 1). The most divergent strain at this site (HM666), however (Fig. 4), is directly connected to two strains from North Bohemia (T-754/II, T-740) indicating the additional arrival of a lineage A strain at this site. Lineage A strains appear also to have been introduced from South Bohemia (e.g. Zdar Kaplice to Müldorf and on to Asbach). Anthropogenic spread is a prime candidate for all of these introductions, but at this stage cannot be demonstrated as clear cut as described

Fig. 4. Fine-grained movement of TBEV in central Europe since the eighteenth century. Directional lines represent migratory events, with arrows connecting sources and destinations. The isolation places are shown by capital letters: Germany (G), Austria (A), Slovakia (K), North Bohemia (N), Central Bohemia (C), South Bohemia (S), South Moravia (B) and North Moravia (M). The last three places are in the Czech Republic. The lines are coloured in agreement with the median of the estimated dates of migration. Also see file S4 and Table 3 to access the confidence interval of estimated dates.

along the Siberian railway thoroughfare in Russia (Kovalev *et al.*, 2009). Prime candidates could be rodents on long-haul lorry transport.

Focal activity

The spread of tick-borne flaviviruses across the Eurasian continent has recently been reanalysed and it appears that the TBEV-Sib and TBEV-Fe types branched off the TBEV-Eur type to migrate towards the east about 3000 years ago in central Asia, whereas TBEV-Eu migrated towards the

west (Heinze *et al.*, 2012). The novelty of our data lies in the close examination of natural active TBEV foci, identified by searching for ticks in the area of the residency of confirmed TBE cases.

Focal activity of TBEV is subject to many factors and currently subject to investigation and discussion (Durmiši *et al.*, 2011; Fajs *et al.*, 2012; Lommano *et al.*, 2012). A reset of tick phenology due to conditions during diapause and warming conditions in spring allowing synchronous appearance and co-feeding of at least two of the three tick

Table 3. Estimated dates of migratory events of TBEV in central Europe

Origin	Destination	Median (years)	95 % HPD (years)
Germany	Central Bohemia	1731	1682–1779
Germany	Central Bohemia	1744	1669–1872
Germany	North Bohemia	1744	1669–1872
Germany	Austria	1748	1695–1803
Germany	Slovakia	1754	1701–1807
Germany	North Bohemia	1767	1743–1867
Germany	South Bohemia	1791	1734–1839
Germany	Austria	1809	1708–1867
Germany	Central Bohemia	1815	1755–1873
North Bohemia	South Moravia	1865	1755–1873
South Bohemia	Germany	1922	1894–1943
South Bohemia	North Moravia	1975	1958–1985

Fig. 5. Reduced MJ network. Plot of a 1228 character set of 102 TBEV-E gene sequences. Yellow circles, lineage A according to MCC tree in Fig. S1; green circles, lineage B. German TBEV-E sequences in the network are depicted in light blue. Central-, South- and North Bohemian clades as defined in Weidmann *et al.* (2011) are highlighted with areas shaded in light red. Foci with diversity (Fig. 3) are highlighted in bright red.

stages (larva, nymph, adult) is essential for saliva-assisted non-viraemic TBEV transmission (Labuda & Randolph, 1999; Randolph *et al.*, 1999). In contrast TBEV emergence at higher altitudes due to climate warming appears to depend mainly on vertical transmission among ticks (Daniel *et al.*, 2004; Danielová *et al.*, 2006).

Long-term studies on TBEV in Russia reveal that reproductive adult male voles with a suppressed immune system due to high testosterone levels are the best amplifiers of TBEV and that the virulence of TBEV increases in tick-host-tick passages (Moshkin *et al.*, 2009). Infection rates in host rodents in turn depend on abundance of tick larvae and nymphs feeding on the adult reproductive voles whose springtime activity coincides with

tick activity, thus establishing horizontal transmission of TBEV among the rodents (Moshkin *et al.*, 2009). More recently, new data on the role of rodents in maintenance of TBEV transmission in Slovenia (Knap *et al.*, 2012) and Scandinavia (Tonteri *et al.*, 2011) support the view that rodents also play a significant role in natural transmission cycles in Europe. Additionally, empirical data supported by mathematical modelling implicate roe deer population densities in sustaining tick populations and therefore TBEV transmission in Slovakia and Italy (Bolzoni *et al.*, 2012; Cagnacci *et al.*, 2012).

Our phylogenetic and network analysis of 102 TBEV-E gene sequences allowed us to infer the diverse recent evolutionary history of TBEV-E gene sequences in central

Europe. It identified local (focal) continuous and long-distance discontinuous spreading patterns of two lineages of TBEV in central Europe and TBEV foci with differing evolutionary rates.

Ecological assessment in respect of climate data, tick and host density data for these genetically diverse TBEV foci will be paramount to obtaining a detailed understanding of the maintenance and spread of TBEV.

The question of why lineages A and B differ in their spread pattern should also be analysed by virological means to characterize replication efficiencies in various cell tissues, dendritic cells and macrophages of relevant host species and in ticks. Since most virological data published to date were presumably generated with lineage B strains (e.g. TBEV strain Hypr71), a comparative analysis may yield important insights into these two lineages and ultimately explain if and how their spreading patterns differ.

METHODS

Virus strains. Viral sequences were obtained from TBEV strains from ticks (*Ixodes ricinus*, *I. hexagonus*), or mammalian hosts (strain 465: *Sciurus vulgaris*; strains V-352, V-361: *Apodemus sylvaticus* M2 A104: *Apodemus flavicollis* strains V-364, V-540, M5 CGI223: *Myodes glareolus*) in the Czech Republic (19 sites), Slovakia (two sites), Austria (three sites) and Germany (14 sites) over a period of more than 67 years. In the years 2010 and 2011 questing ticks were collected by flagging the low vegetation at seven sites in south Germany (Haselmühl, Burglengelfeld, Amberg-Sulzbach, Asbach, Neustadt a. d. W., Passau and Heselbach), in the vicinity of recent TBE cases reported by the local health authorities. Ticks were separated according to their sex and developmental stage, morphologically identified to the species level (Hillyard, 1996) and subsequently homogenized and processed for reverse transcription (RT)-PCR as described previously (Frey *et al.*, 2012). The year, host source and geographical location of the particular virus samples are summarized in Table S1.

RNA extraction PCR amplification and sequencing. All sequences generated since 2005 were obtained directly from RNA isolated from tick homogenates, without any isolation or attenuation in cell cultures or mouse brain.

Total nucleic acids were extracted from 200 µl sample using the MagNA Pure LC Total Nucleic Acid (NA) Isolation kit (Roche) and the MagNA Pure LC instrument (Roche). Isolated NA (50 µl) was stored at -80°C . Screening real-time RT-PCR, subsequent envelope (E) gene RT-PCR and sequencing of the screened positive tick samples were performed as described previously (Frey *et al.*, 2012; Weidmann *et al.*, 2011).

Phylogenetic analysis

Sequence alignment:

Nucleotide sequence alignments were performed using CLUSTAL W and bootstrapping (1000 repetitions) in DNASTAR LASERGENE MEGALIGN. Dendrograms were built using DENDROSCOPE (Huson *et al.*, 2007).

Bayesian phylodynamics analysis

For the analysis of the evolutionary history of the TBEV, a 1488 character-long alignment (including gaps) of 106 E genes, connected

to isolation year and sampling site (latitude and longitude), was used for a maximum-likelihood (ML) optimization in GARRI BUILD 0.951 (Zwickl, 2006) using the General Time Reversible (GTR) model of nucleotide substitution (Krejić, 1949) and optimizing topology, branch lengths, rate heterogeneity and model parameters. The date-stamped samples, allowed to infer a maximum clade credibility (MCC) tree with dated tips and internal nodes using a MCMC Bayesian approach and a GTR model with gamma-distributed rate variation (Γ) and a proportion of invariable sites (I) using a relaxed (uncorrelated lognormal) molecular clock in BEAST v1.5.4 (Drummond & Rambaut, 2007). The GTR+ Γ +I substitution model was used, since it was the best model obtained with MODELTEST (Posada, 2008). A relaxed molecular clock was used since flavivirus evolution generally approximates a molecular clock, but some minor rate differences may occur (Tiddy *et al.*, 2003). Four independent MCMC runs of four chains each were run for 100 million states. The convergence of parameters during MCMC runs was assessed by their effective sample size (ESS) reaching values above 200, as calculated with TRACER v1.5 (<http://tree.bio.ed.ac.uk/software/tracer/>). Since Bayesian inference is highly sensitive to prior selection and because new sequences were being analysed, we tested two different substitution rates to infer TBEV changes in time. Firstly we used an initial substitution rate (μ) of $8.0 \times 10^{-4} \text{ s}^{-1}/\text{y}$ in agreement with our previous estimate (Weidmann *et al.*, 2011) and we also used a prior based on a coarse rate estimate obtained with the Path-O-Gen v1.3 program (<http://tree.bio.ed.ac.uk/software/pathogen/>), which uses the best ML tree and sampling dates. To assess the best-fit rate to our data, we did a BF test (Suchard *et al.*, 2001) which compares the harmonic mean of the marginal likelihoods for each model. To investigate selection regimens on the E gene, we estimated the difference ($\omega = dN - dS$) between the non-synonymous (dN) and synonymous (dS) rates per codon sites, using the single likelihood ancestor counting (SLAC) algorithm with HyPhy v2.11 (Pond *et al.*, 2005), assuming a significance level of 5% ($\alpha = 0.05$). Thus, ω greater than zero suggests directional selection, while values below zero are indicative of purifying selection.

Phylogeography of TBEV

We tested two major hypotheses of TBEV spread in our area of study: dispersion following a continuous diffusion model (Lemey *et al.*, 2010), and spread by discrete events of migration (Lemey *et al.*, 2009). We also tested if the diffusion rates between localities were homogeneous (Brownian diffusion), or if the diffusion rates were independent following a Cauchy distribution. We fed the continuous models with geographical coordinates from the places of viral isolation. The second scenario was inferred using a continuous time Markov chain (CTMC) model (Lemey *et al.*, 2009) and under the special case of a Bayesian stochastic search variable selection (BSSVS) model. We assigned to each sequence a discrete label in agreement with its place of isolation (Central Bohemia, South Bohemia, North Bohemia, Germany, Slovakia and Austria). The time-scaled phylogenies to reconstruct spread patterns were obtained during the stationarity of four MCMC using BEAST. The MCMC was run for 2×10^8 generations for continuous models and 5×10^7 for discrete models to obtain model convergence. Our sample from the posterior included 10 000 trees that were used to construct the MCC tree. We used the BF test (Suchard *et al.*, 2001) to compare the marginal likelihood obtained from the MCMC for each model, after 1000 bootstrap replicates, to test the best-fit model to the data. The MCC tree from the best model with ancestral states in nodes was plotted in Google Earth software (<http://www.google.com/earth/index.html>), using SPREAD v1.0.5 (Bielejec *et al.*, 2011). In addition, to investigate the relationships between geographical and genetic distances we calculated, from the phylogenetic tree with the best likelihood obtained after 100 independent runs of the GARRI program, the patristic distances using PATRISTIC software (Fourment & Gibbs,

2006); the geographical distances between two points were estimated using the spherical law of cosines with a Perl script. The correlation between the matrices of geographical and patristic distances was inspected using the Mantel test implemented in the APE package v3.0.2 (Paradis *et al.*, 2004) from the R-project (<http://www.r-project.org/>). We investigated relationships between geographical and patristic distances with a Mantel correlogram (Carrel *et al.*, 2010), which was generated with the Bonferroni multiple correction and Spearman's rank correlation coefficient (ρ), using VEGAN package v2.05 (Oksanen *et al.*, 2012) from the R-project.

MJ network analysis

A network was constructed from a 102 TBEV-E gene 1222 character alignment set stripped of all homogeneous character using SPLITS TREE 4.0 with Epsilon 1 and 2000 spring embedded iterations. A parsimony splits analysis was also used.

ACKNOWLEDGEMENTS

Supporting grants: Czech Science Foundation, no. P502/11/2116; Academy of Sciences of the Czech Republic, grant Z60220518; the AdmireVet project, no. CZ.1.05/2.1.00/01.006 (ED006/01/01) of the Ministry of Education, Youth and Sports of the Czech Republic; German Federal Ministry of Education and Research, grant 01KJ0710. In addition, we acknowledge CAPES and CNPq-PQ for scholarships received by C. C. M. F. and P. M. A. Z. We are grateful for computational resources provided by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), projects 2010/19341-4 and 2011/17120-3. We thank F. X. Heinz for providing the TBEV Scharl case site.

REFERENCES

- Aberle, J. H., Aberle, S. W., Allison, S. L., Stiasny, K., Ecker, M., Mandl, C. W., Berger, R. & Heinz, F. X. (1999). A DNA immunization model study with constructs expressing the tick-borne encephalitis virus envelope protein E in different physical forms. *J Immunol* **163**, 6756–6761.
- Achazi, K., Růžek, D., Donoso-Mantke, O., Schlegel, M., Ali, H. S., Wenk, M., Schmidt-Chanasit, J., Ohlmeyer, L., Růžek, F. & other authors (2011). Rodents as sentinels for the prevalence of tick-borne encephalitis virus. *Borne Zoonotic Dis* **11**, 641–647.
- Bielejec, F., Rambaut, A., Suchard, M. A. & Lemey, P. (2011). SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics* **27**, 2910–2912.
- Bolzoni, L., Rosà, R., Cagnacci, F. & Rizzoli, A. (2012). Effect of deer density on tick infestation of rodents and the hazard of tick-borne encephalitis. II: population and infection models. *Int J Parasitol* **42**, 373–381.
- Cagnacci, F., Bolzoni, L., Rosà, R., Carpi, G., Hauffe, H. C., Valent, M., Tagliapietra, V., Kazimirova, M., Koci, J. & other authors (2012). Effects of deer density on tick infestation of rodents and the hazard of tick-borne encephalitis. I: empirical assessment. *Int J Parasitol* **42**, 365–372.
- Carpi, G., Cagnacci, F., Neteler, M. & Rizzoli, A. (2008). Tick infestation on roe deer in relation to geographic and remotely sensed climatic variables in a tick-borne encephalitis endemic area. *Epidemiol Infect* **136**, 1416–1424.
- Carrel, M. A., Emch, M., Jobe, R. T., Moody, A. & Wan, X. F. (2010). Spatiotemporal structure of molecular evolution of H5N1 highly pathogenic avian influenza viruses in Vietnam. *PLoS ONE* **5**, e8631.
- Chambers, T. J., Hahn, C. S., Galler, R. & Rice, C. M. (1990). Flavivirus genome organization, expression, and replication. *Annu Rev Microbiol* **44**, 649–688.
- Daniel, M., Danielová, V., Kriz, B. & Kott, I. (2004). An attempt to elucidate the increased incidence of tick-borne encephalitis and its spread to higher altitudes in the Czech Republic. *Int J Med Microbiol* **293** (Suppl 37), 55–62.
- Daniel, M., Benes, C., Danielová, V. & Kriz, B. (2011). Sixty years of research of tick-borne encephalitis a basis of the current knowledge of the epidemiological situation in Central Europe. *Epidemiol Mikrobiol Immunol* **60**, 135–155.
- Danielová, V., Rudenko, N., Daniel, M., Holubová, J., Materna, J., Golovchenko, M. & Schwarzová, L. (2006). Extension of *Ixodes ricinus* ticks and agents of tick-borne diseases to mountain areas in the Czech Republic. *Int J Med Microbiol* **296** (Suppl 40), 48–53.
- Donoso Mantke, O., Escadafal, C., Niedrig, M., Pfeffer, M. & Working Group For Tick-Borne Encephalitis Virus C. (2011). Tick-borne encephalitis in Europe, 2007 to 2009. *Euro Surveill* **16**, pii: 19976.
- Drummond, A. J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214.
- Durmiš, E., Knap, N., Saksida, A., Trilar, T., Duh, D. & Avšič-Županc, T. (2011). Prevalence and molecular characterization of tick-borne encephalitis virus in *Ixodes ricinus* ticks collected in Slovenia. *Vector Borne Zoonotic Dis* **11**, 659–664.
- Fajš, L., Durmiš, E., Knap, N., Strle, F. & Avšič-Županc, T. (2012). Phylogeographic characterization of tick-borne encephalitis virus from patients, rodents and ticks in Slovenia. *PLoS ONE* **7**, e48420.
- Felsenstein, J. (2004). *Inferring Phylogenies*. Sunderland: Sinauer Associates Inc.
- Fickel, J., Bubliy, O. A., Stache, A., Noventa, T., Jirsa, A. & Heurich, M. (2012). Crossing the border? Structure of the red deer (*Cervus elaphus*) population from the Bavarian-Bohemian forest ecosystem. *Mamm Biol* **77**, 211–220.
- Fourment, M. & Gibbs, M. J. (2006). PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. *BMC Evol Biol* **6**, 1.
- Frey, S., Mossbrugger, I., Altantuu, D., Battsetseg, J., Davaadorj, R., Tserennorov, D., Buyanjargal, T., Otgonbaatar, D., Zöller, L. & other authors (2012). Isolation, preliminary characterization, and full-genome analyses of tick-borne encephalitis virus from Mongolia. *Virus Genes* **45**, 413–425. (Epub ahead of print).
- Heinze, D. M., Gould, E. A. & Forrester, N. L. (2012). Revisiting the dinal concept of evolution and dispersal for the tick-borne flaviviruses by using phylogenetic and biogeographic analyses. *J Virol* **86**, 8663–8671.
- Hillyard, P. D. (1996). *Ticks of North-West Europe*. Dorset Press, Dorchester.
- Hoogstraal, H., Kaiser, M. N., Traylor, M. A., Guindy, E. & Gaber, S. (1963). Ticks (Ixodidae) on birds migrating from Europe and Asia to Africa 1959–61. *Bull World Health Organ* **28**, 235–262.
- Huson, D. H., Richter, D. C., Rausch, C., Dezulian, T., Franz, M. & Rupp, R. (2007). Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinformatics* **8**, 460.
- Klaus, C., Hoffmann, B., Hering, U., Mielke, B., Sachse, K., Beer, M. & Süß, J. (2010). Tick-borne encephalitis (TBE) virus prevalence and virus genome characterization in field-collected ticks (*Ixodes ricinus*) from risk, non-risk and former risk areas of TBE, and in ticks removed from humans in Germany. *Clin Microbiol Infect* **16**, 238–244.
- Knap, N., Korva, M., Dolinšek, V., Sekirnik, M., Trilar, T. & Avšič-Županc, T. (2012). Patterns of tick-borne encephalitis virus infection in rodents in Slovenia. *Vector Borne Zoonotic Dis* **12**, 236–242.
- Kovalev, S. Y., Chernykh, D. N., Kokorev, V. S., Snitkovskaya, T. E. & Romanenko, V. V. (2009). Origin and distribution of tick-borne

- encephalitis virus strains of the Siberian subtype in the Middle Urals, the north-west of Russia and the Baltic countries. *J Gen Virol* 90, 2884–2892.
- Krejčí, J. (1949). [Outbreak of encephalitis virus in the region of Vyskov.] *Lék Listy (Brno)* 4, 73–75, 112–116, 132–134.
- Kriz, B., Maly, M., Benes, C. & Daniel, M. (2012). Epidemiology of tick-borne encephalitis in the Czech Republic 1970–2008. *Vector Borne Zoonotic Dis* 12, 994–999.
- Kupča, A. M., Essbauer, S., Zoeller, G., de Mendonça, P. G., Brey, R., Rinder, M., Pfister, K., Spiegel, M., Doerrbecker, B. & other authors (2010). Isolation and molecular characterization of a tick-borne encephalitis virus strain from a new tick-borne encephalitis focus with severe cases in Bavaria, Germany. *Ticks Tick-borne Dis* 1, 44–51.
- Labuda, M. & Nuttall, P. A. (2004). Tick-borne viruses. *Parasitology* 129 (Suppl S1), S221–S245.
- Labuda, M. & Randolph, S. E. (1999). Survival strategy of tick-borne encephalitis virus: cellular basis and environmental determinants. *Zentralbl Bakteriol* 289, 513–524.
- Labuda, M., Elecková, E., Licková, M. & Sabó, A. (2002). Tick-borne encephalitis virus foci in Slovakia. *Int J Med Microbiol* 291 (Suppl 33), 43–47.
- Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. (2009). Bayesian phylogeography finds its roots. *PLOS Comput Biol* 5, e1000520.
- Lemey, P., Rambaut, A., Welch, J. J. & Suchard, M. A. (2010). Phylogeography takes a relaxed random walk in continuous space and time. *Mol Biol Evol* 27, 1877–1885.
- Licoppe, A. & Lievens, J. (2011). Seasonal home ranges of red deer (*Cervus elaphus*) and its impact on the scale of culling management. In *Abstracts XXXth International Union of Game Biologists Congress, 5–9 September 2011, Barcelona, Spain*, p. 319. Available at http://www.alospaincongress.com/contenido/Llibre_Abrats%20retocat%20iugb.pdf
- Lommano, E., Burri, C., Maeder, G., Guerne, M., Bastic, V., Patalas, E. & Gern, L. (2012). Prevalence and genotyping of tick-borne encephalitis virus in questing *Ixodes ricinus* ticks in a new endemic area in western Switzerland. *J Med Entomol* 49, 156–164.
- Moshkin, M. P., Novikov, E. A., Tkachev, S. E. & Vlasov, V. V. (2009). Epidemiology of a tick-borne viral infection: theoretical insights and practical implications for public health. *Bioessays* 31, 620–628.
- Nicolaisen, L. E. & Desai, M. M. (2012). Distortions in genealogies due to purifying selection. *Mol Biol Evol* 29, 3589–3600.
- Oksanen, J., Blanchet, F., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B. & Simpson, G. L. (2012). *vegan*: Community Ecology Package. R package version 2.0–5. <http://cran.r-project.org/package=vegan/index.html>
- Paradis, E., Claude, J. & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290.
- Pond, S. L. K., Frost, S. D. W. & Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21, 676–679.
- Posada, D. (2008). jModelTest: phylogenetic model averaging. *Mol Biol Evol* 25, 1253–1256.
- Randolph, S. E. & Storey, K. (1999). Impact of microclimate on immature tick-rodent host interactions (Acari: Ixodidae): implications for parasite transmission. *J Med Entomol* 36, 741–748.
- Randolph, S. E., Miklisová, D., Lysy, J., Rogers, D. J. & Labuda, M. (1999). Incidence from coincidence: patterns of tick infestations on rodents facilitate transmission of tick-borne encephalitis virus. *Parasitology* 118, 177–186.
- Randolph, S. E., Green, R. M., Peacey, M. F. & Rogers, D. J. (2000). Seasonal synchrony: the key to tick-borne encephalitis foci identified by satellite data. *Parasitology* 121, 15–23.
- Silaghi, C., Woll, D., Mahling, M., Pfister, K. & Pfeffer, M. (2012). *Candidatus Neohhrlichia mikurensis* in rodents in an area with sympatric existence of the hard ticks *Ixodes ricinus* and *Dermacentor reticulatus*, Germany. *Parasit Vectors* 5, 285.
- Široký, P., Kubelová, M., Bednář, M., Modrý, D., Hubálek, Z. & Tkadlec, E. (2011). The distribution and spreading pattern of *Dermacentor reticulatus* over its threshold area in the Czech Republic –how much is range of this vector expanding? *Vet Parasitol* 183, 130–135.
- Suchard, M. A., Weiss, R. E. & Sinsheimer, J. S. (2001). Bayesian selection of continuous-time Markov chain evolutionary models. *Mol Biol Evol* 18, 1001–1013.
- SurvStat (2012). See <http://www3.rki.de/SurvStat/QueryForm.aspx> Robert Koch Institute.
- Tonteri, E., Jääskeläinen, A. E., Tikkakoski, T., Voutilainen, L., Niemimaa, J., Henttonen, H., Vaheri, A. & Vapalahti, O. (2011). Tick-borne encephalitis virus in wild rodents in winter, Finland, 2008–2009. *Emerg Infect Dis* 17, 72–75.
- Twiddy, S. S., Holmes, E. C. & Rambaut, A. (2003). Inferring the rate and time-scale of dengue virus evolution. *Mol Biol Evol* 20, 122–129.
- Uzcátegui, N. Y., Sironen, T., Golovljova, I., Jääskeläinen, A. E., Välimaa, H., Lundkvist, A., Plyusnin, A., Vaheri, A. & Vapalahti, O. (2012). Rate of evolution and molecular epidemiology of tick-borne encephalitis virus in Europe, including two isolations from the same focus 44 years apart. *J Gen Virol* 93, 786–796.
- Vor, T., Kiffner, C., Hagedorn, P., Niedrig, M. & Rühle, F. (2010). Tick burden on European roe deer (*Capreolus capreolus*). *Exp Appl Acarol* 51, 405–417.
- Walczak, A. M., Nicolaisen, L. E., Plotkin, J. B. & Desai, M. M. (2012). The structure of genealogies in the presence of purifying selection: a fitness-class coalescent. *Genetics* 190, 753–779.
- Waldenström, J., Lundkvist, A., Falk, K. I., Garpmo, U., Bergström, S., Lindegren, G., Sjöstedt, A., Mejlon, H., Fransson, T. & other authors (2007). Migrating birds and tickborne encephalitis virus. *Emerg Infect Dis* 13, 1215–1218.
- Weidmann, M., Ruzek, D., Krivanec, K., Zöller, G., Essbauer, S., Pfeffer, M., Zanotto, P. M., Hufert, F. T. & Dobler, G. (2011). Relation of genetic phylogeny and geographical distance of tick-borne encephalitis virus in central Europe. *J Gen Virol* 92, 1906–1916.
- Wertheim, J. O. & Kosakovsky Pond, S. L. (2011). Purifying selection can obscure the ancient age of viral lineages. *Mol Biol Evol* 28, 3355–3365.
- Woolley, S. M., Posada, D. & Crandall, K. A. (2008). A comparison of phylogenetic network methods using computer simulation. *PLoS ONE* 3, e1913.
- Zanotto, P. M., de A., Gao, G. F., Gritsun, T., Marin, M. S., Jiang, W. R., Venugopal, K., Reid, H. W. & Gould, E. A. (1995). An arbovirus cline across the northern hemisphere. *Virology* 210, 152–159.
- Zanotto, P. M., Gould, E. A., Gao, G. F., Harvey, P. H. & Holmes, E. C. (1996). Population dynamics of flaviviruses revealed by molecular phylogenies. *Proc Natl Acad Sci U S A* 93, 548–553.
- Zwickl, D. J. (2006). *Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets under the Maximum Likelihood Criterion*. Austin: University of Texas.
- Zygner, W., Górski, P. & Wedrychowicz, H. (2009). New localities of *Dermacentor reticulatus* tick (vector of *Babesia canis canis*) in central and eastern Poland. *Pol J Vet Sci* 12, 549–555.

Phylogeography of Rift Valley Fever Virus in Africa Reveals Multiple Introductions in Senegal and Mauritania

P. O. Ly Soumaré^{1*}, Caio C. M. Freire^{2*}, Ousmane Faye¹, Mawlouth Diallo¹, Juliana Velasco C. de Oliveira², Paolo M. A. Zanotto², Amadou Alpha Sall¹

1 Institut Pasteur de Dakar, Dakar, Senegal, **2** Laboratory of Molecular Evolution and Bioinformatics, Department of Microbiology, Biomedical Sciences Institute, University of Sao Paulo, Sao Paulo, Brazil

Abstract

Rift Valley Fever (RVF) virus (Family Bunyaviridae) is an arthropod-borne RNA virus that infects primarily domestic ruminants and occasionally humans. RVF epizootics are characterized by numerous abortions and mortality among young animals. In humans, the illness is usually characterized by a mild self-limited febrile illness, which could progress to more serious complications. RVF virus is widespread and endemic in many regions of Africa. In Western Africa, several outbreaks have been reported since 1987 when the first major one occurred at the frontier of Senegal and Mauritania. Aiming to evaluate the spreading and molecular epidemiology in these countries, RVFV isolates from 1944 to 2008 obtained from 18 localities in Senegal and Mauritania and 15 other countries were investigated. Our results suggest that a more intense viral activity possibly took place during the last century compared to the recent past and that at least 5 introductions of RVFV took place in Senegal and Mauritania from distant African regions. Moreover, Barkedji in Senegal was possibly a hub associated with the three distinct entries of RVFV in West Africa.

Citation: Soumaré POL, Freire CCM, Faye O, Diallo M, Oliveira JVCde, et al. (2012) Phylogeography of Rift Valley Fever Virus in Africa Reveals Multiple Introductions in Senegal and Mauritania. PLoS ONE 7(4): e35216. doi:10.1371/journal.pone.0035216

Editor: Tetsuro Ikegami, The University of Texas Medical Branch, United States of America

Received: December 16, 2011; **Accepted:** March 13, 2012; **Published:** April 23, 2012

Copyright: © 2012 Soumaré et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was funded by Institut Pasteur de Dakar in Senegal and FAPESP (Fundação de Amparo a Pesquisa do Estado de São Paulo, Brazil) project #00/04205-6 (VGDN program). CCMF has CAPES scholarship and PMAZ holds a CNPq-PQ scholarship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: freire@ime.usp.br

† These authors contributed equally to this work.

Introduction

Rift Valley fever (RVF) is a mosquito borne anthroponozoonosis affecting livestock and also humans throughout Africa and the Arabian Peninsula. Animal infections lead to high mortality among young ruminants and abortions of pregnant females while human infections are associated with a wide array of syndromes ranging from influenza-like illness to severe symptoms including hemorrhages, encephalitis, hepatitis, ocular complications and fatal outcomes [1]. Animal and human mortality and morbidity caused by RVF often result in high public health burden and substantial economic losses in endemic countries [2]. RVF is caused by the RVF virus (RVFV), a phlebovirus (Family Bunyaviridae) with three single-stranded RNA segments genome: L (large), M (medium) and S (small). The L segment encodes the RNA-dependent RNA polymerase [3], the M segment envelope glycoproteins G1 and G2 and two non-structural proteins of 14 and 78 kDa molecular weight, respectively [4]. The S segment has an ambisense coding strategy, since it encodes for the nucleocapsid protein (N) on the 5'-3', positive sense and a non-structural protein (NSs) in the 5'-3', negative sense [5]. The latter protein plays a major role in innate immunity by blocking interferon gene expression [6].

Although RVF was discovered in 1930 [7,8], the disease remained a veterinary concern until a major outbreak occurred in Egypt in 1977 [1,9], where RVF caused about one million human infections, 600 deaths and severe widespread epizootics [10]. Subsequently, more studies on RVFV transmission characterized additional major outbreaks derived from endemic sub-Saharan African countries [11–16]. The exchange from the enzootic cycle (*i.e.*, among wild animals) to peridomestic transmission (*i.e.*, among animals such as cattle, lambs and goats) is thought to cause the occasional spillover outbreaks into humans [17]. The most recent epidemics occurred in Kenya, Somalia, Tanzania in 2007 [18], South Africa in 2008 [19] and again in 2010 [20], Sudan in 2008 where more than 200 deaths were reported [21] and Mauritania in 2010 (Faye, unpublished data). A second introduction occurred when RVF, coming from East Africa, moved into Saudi Arabia and Yemen in the Arabian Peninsula [22]. It went for the first time outside of Africa –where it had been confined so far– becoming a threat to the Middle East. In the light of its natural history, RVFV appears to be a great example of an emerging pathogen with great potential of dispersion and impact on human and animal health and, a good model for an integrated approach to human and animal health (*i.e.*, the ‘one health’ concept). Several authors using molecular phylogeny methods attempted to unravel mechanisms underlying its distribution and dispersal in Africa [16,23–26].

These studies revealed that: (i) RVFV genomes carry low genetic diversity ranging from 1 to 5% sequence divergence, (ii) change at a mean evolutionary rates ranging from 2.8 to 3.9E-4 substitution/site/year [27], (iii) can be grouped into several lineages [23,25], (iv) can be grouped by geographic origin [16], (v) undergo reassortment in nature [28] and, (vi) the extant genetic diversity coalesces about 120–130 years ago [23]. However, except for Kenya [27], limited information is available on the evolution of RVFV in a country or regional scale. This lack of information poses the necessity to investigate the determinants of RVFV circulation in West Africa. In order to fill this gap, we sequenced 48 RVFV isolates from Senegal and Mauritania and along with sequences from Guinea and Burkina Faso, to allow inferences on the dispersal patterns of RVFV in West Africa. We also added to the study samples from East Africa to establish migratory patterns in Africa at a coarse grain.

Results and Discussion

Sequence analyses

Forty-eight RVFV isolates collected over a period of 20 years (1983 to 2003) from different areas of Senegal and Mauritania were included in the study (Table S1). For each strain NSs, G2, and polymerase genes were partially sequenced. Overall levels of sequence diversity and polymorphisms were within ranges previously observed and no deletions or insertions were found [24,29]. Our analyses of recombination, using several different methods, showed the absence of intra gene recombination on the regions we studied and low levels of genetic diversity, in agreement with what was previously observed by [16,23–27]. There are many plausible explanations for the low rates of genetic change estimated for RVFV, which could entail factors involved in genetic variability reduction, ranging from speed of replication, high rate of vertical transmission to reduced viremia. Moreover, the alternation between vertebrates and invertebrate hosts during the RVFV life cycle could hinder viral variation by imposing strong selective advantage in keeping highly adapted phenotypes [30]. Likewise, selection analyzes of the three genomic segments uncovered several sites under strong negative selection, indicated by $\omega < 0$, (Table 1) shows purging of deleterious polymorphisms in functionally important genes. On the other hand, a lack of positively selected sites indicated by $\omega > 0$, shows a lack directional change, typical of highly adapted genotypes. These findings also agree with the fact that the natural cycle of RVFV imposes several transmission bottlenecks due to alternation between a diverse set of arthropods vectors and mammal hosts that keep selective pressure towards genomic stability, such as the preservation of the S segment *in vivo* [30]. Moreover, the genomic stability we observed is consistent with the enzootic status of RVFV in several places in Africa.

Phylogenetics of RVFV

We first investigated the phylogenetic signal content in our data by reconstructing 50 thousand quartets for each gene segment using the likelihood mapping method (see methods section). Our results (Table S2) indicated that the M segment had the higher phylogenetic signal content given its lower percentage of unresolved quartets, followed by the L and S segments. Maximum clade credibility (MCC) trees obtained during Bayesian inference for the S (Figure 1), M (Figure 2) and L (Figure 3) segments and trees obtained by maximum likelihood (data not shown), grouped samples from Senegal and Mauritania as sister taxa, as indicated by the preferential clustering of yellow and green dots in the tips of distinct clades in the trees. The occurrence of green and yellow

Table 1. Negatively selected codon sites detected in RVFV.

Genomic segment	Amino acid position*	ω^*	p-value
Large	559	−5.0648	0.0085
Large	860	−5.0648	0.0085
Large	956	−5.0648	0.0085
Large	1091	−6.3310	0.0030
Large	1332	−5.0648	0.0085
Large	1870	−5.0648	0.0085
Medium	64	−5.1841	0.0093
Medium	172	−54.8731	0.0072
Medium	225	−6.9101	0.0021
Medium	282	−4.5937	0.0071
Medium	439	−5.0000	0.0041
Medium	971	−5.5281	0.0071
Medium	1115	−11.6919	0.0047
Small	116	−5.0000	0.0047
Small	124	−4.5667	0.0073
Small	127	−8.0000	0.0002
Small	148	−4.5667	0.0073
Small	151	−5.9157	0.0048

*The amino acid position was related to first codon of sequence from Mauritanian OS8 strain. The accession numbers of Large, Medium and Small segments are, respectively, DQ375395.1, DQ380185.1 and DQ380177.1.

*Negatively selected sites ($\omega < 0$) identified by the significant difference between non-synonymous (dN) rates and synonymous (dS) rates per site using the Single Likelihood Ancestor Counting (SLAC) method with HyPhy. doi:10.1371/journal.pone.0035216.t001

dots in five distinct groups of taxa in Figure 1, 2 and 3 is evidence for independent shared introductions in these countries. These findings were well supported by significant posterior probabilities and by high bootstrap values for maximum likelihood trees, inferred with GARLI (data not shown). Furthermore, the phylogenetic adjacency of lineages, indicated by clustering of green and yellow dots in Figure 1, 2 and 3, indicated a recurrent exchange between the two neighbor countries during the last century. We then did phylogeographic reconstructions based on the S segment alone, since it had the highest posterior probabilities during state reconstructions. This allowed a coarse-grained view of the spatio-temporal patterns of RVFV spread in the African continent in general and, the pattern of movement into Mauritania and Senegal in particular. Importantly, the time to the most recent common ancestor (TMRCA) parameter values that we estimated from sequences from each genetic segment were different. This could be explained to some extent by the large and distinct number of negatively selected sites among segments, which could lead to an underestimation of the age of nodes near to the root of tree [31]. In addition, albeit obtaining high posterior values in MCC trees, we used different-sized sequences during our reconstructions of viral genealogies, which is acceptable within the likelihood framework, but could nonetheless, lead to underestimation of branch lengths for taxa with shorter sequences [32]. On the other hand, segmented genome allows reassortment events [28] that even at low frequency [23], this could cause differences in TMRCA estimates for different segments.

Based on the S segment of the RVFV that better resolved phylogeographic patterns, we observed that our trees had temporal coherence, since the age of key nodes of the S MCC

tree agreed with known events, such as the first report of RVFV in Kenya in 1930 [7, 8] and the estimated arrival in this country from Zimbabwe. A root for the tree was dated between 1909 and 1920 and placed somewhere near the Southern tip of the Rift Valley. We assume that variation in genetic diversity in time (measured as effective population size times the generation time, $N_e g$) estimated with the Monte Carlo Markov Chain (MCMC) method, correlates with RVFV demography and is proportional to the number of infections in time. The Bayesian skyride plots (BSP) for all genomic segments of RVFV had an increase in $N_e g$ in the first half of last century followed by a continuous decrease in the last half (Figure 4). Moreover, the BSPs agreed with a recent origin of RVFV and intense viral activity in the first half of 20th century followed by dispersal in the African Continent. While the reduction in viral population size or infection events ($N_e g$) are consistent with enzootic activity near the present. This finding agrees with the RVFV spread among large populations of susceptible hosts following to the introduction of livestock in East Africa namely Kenya in the first half of the 20th century to improve husbandry [27, 33]. In essence, the higher levels of $N_e g$ in the past (Figure 4) agree with the fact that RVFV has been active

Phylogeographic reconstructions summarized in Figure S1 unveiled a complex pattern of viral movement across long distances in the African continent that we represented in maps (Figures 5 and 6). The dates in the map arrows in Figures 5 and 6 indicate the reconstructed TMRCAs for a lineage found at the locality shown by the arrowhead and is assumed as the oldest possible year of introduction of that lineage at a given locality. We can observe a stepwise spread from a locality to the next, by focusing on East Africa, were denser sampling allows greater detail (Figure 6). Our inferences unveiled five distinct introductions in Senegal and Mauritania during the past century (Figure 5 and KML file in Dataset S1). The first arrival in West Africa (blue line in Figure 5) was related to strains from Zimbabwe, South Africa and Uganda possibly from around 80 years ago, spreading to several locations in Senegal (SN) and Mauritania (MR). These inferences are consistent with serological surveys done by the Pasteur Institute in Dakar that suggest that the coastal zone of West Africa may have had RVFV during the first half of the last

Figure 2. Maximum clade credibility (MCC) tree for the medium (M) segment. Relationship among 128 strains of RVFV isolated from different localities and countries. Samples from Mauritania and Senegal are shown by green and yellow dots respectively. The recurrent independent clustering of green and yellow dots suggests multiple introductions of RVFV into these countries throughout the 20th century. Posterior probability values greater than 90% are shown near tree nodes.
doi:10.1371/journal.pone.0035216.g002

century (data not published). The second introduction to West Africa (black line in Figure 5 and Figure 6) was associated with strains from South Africa or Zimbabwe sharing a common ancestor around 72 years ago. This introduction was associated with samples from Diawara (SN) and Hodh El Garbi (MR), and an outbreak in 1998 in Diawara [15]. The third introduction in West Africa (shown as a green line in Figure 5 and Figure 6) consisted of strains that originated in Zimbabwe in the beginning of last century, then moved to Central African Republic around 70 years ago and, from there reached Senegal and Mauritania around 47 years ago and Guinea 38 years ago. Samples from this introduction were isolated in Barkedji (SN) in 1993 and Rosso (MR) in 1987 and in Guinea in 1981 and 1984. The fourth introduction (shown as solid black line in Figure 5 and Figure 6) was evidenced by a single strain isolated in Kedougou in 1983 associated with an old South African lineage that moved to Egypt samples around 47 years ago and Madagascar around 40 years ago. The most recent introduction in west Africa (fifth introduction, red line, in Figure 5 and Figure 6) was due to a wave of

RVFV that spread to distant places in Africa caused by a lineage from the Southern part of Africa that moved to Zimbabwe and Kenya 47 years ago and then reached Madagascar, Somalia, Tanzania, Angola, Senegal, Mauritania and Saudi Arabia during the last 28 years. The most interesting feature of this RVFV wave was that it recurrently reemerged in Kenya from where it was broadcasted to other localities.

Phylogeography of RVFV in West Africa

The history of introductions, exchanges and spread of RVFV among Senegalese and Mauritanian territories is shown in greater detail in Figure 6. The map in Figure 6 summarizes independent phylogeographic reconstructions, one for each introduction. The MCC trees (available from authors upon request) agreed well with global MCC tree Figure S1. According to our reconstructions, the first introduction possibly originating from South Africa in the 30's appears to have taken place in Barkedji (SN). Subsequently, it moved to Rosso (MR) around 1970 and Kolda (SN) around 1980. The Rosso lineage was then broadcasted to the provinces of

Figure 3. Maximum clade credibility (MCC) tree for the long (L) segment. Relationship among 126 strains of RVFV isolated from different localities and countries. Samples from Mauritania and Senegal are shown by green and yellow dots respectively. The recurrent independent clustering of green and yellow dots suggests multiple introductions of RVFV into these countries throughout the 20th century. Posterior probability values greater than 90% are shown near tree nodes.
doi:10.1371/journal.pone.0035216.g003

Garack, Nkic and Keur Macene in Mauritania around 1980. Later in 1985, a lineage from Nkic moved into Terg in the North of Mauritania. On the other hand, a lineage from Kolda moved into east Mauritania to Ayoun El Atrouss around the end of the 80's, and from there, to Guimi around 2000. The second introduction, possibly from central Africa, reached Diawara in Senegal around 1940. From there it moved to Hodh el Garbi in Mauritania around 1970. As for the first and third introductions, the fifth (most recent introduction, indicated by a red line) also appears to have entered Senegal in Barkedji and from there, reached the Mauritanian territories of Lazaret and Hseytine. They then spread from Hseytine to Kiffa (MR) and from Lazaret to Matar Lajam and Dar Naim and from the last to Tijiga during the last decade.

Biologic correlates of complex RVFV spread patterns

By contrasting different scales of RVFV spread, fine-grain in West Africa and coarse-grained elsewhere in the continent, we unveiled an intricate dispersal pattern that shows: (i) viral dispersal

across long distances, some times at fast rates, as indicated by the large distances among places of isolation of closely-related strains [26], (ii) persistence in different places for more than 70 years, like that observed for old lineages from South Africa, Zimbabwe, Central African Republic, Kenya and Senegal and, (iii) a gradual spread at nearby localities that we show in greater detail in Senegal and Mauritania. The capacity of RVFV to cross large geographical extensions, could be assisted by the movement of infected mammals and mosquitoes [34]. Likewise, animal herds, acting as reservoirs, could warrant the maintenance of RVFV enzootic circulation at a regional level [35]. The spread of RVF from endemic regions to areas without disease through animal migration routes has been postulated for previous outbreaks. For example, RVFV has been exported to Egypt in 1977, probably by lambs from Sudan [36] and from Eastern Africa to Madagascar in 1991 [16] and then to the Arabian Peninsula in 2000 [37]. Nevertheless, some evidence for enzootic activity has also been put forward. For example, it has been proposed that there is persistent enzootic activity of RVFV in Senegal based on seroprevalence data, which could help explain outbreaks in Mauritania and Egypt

Figure 4. Bayesian skyride plot (BSP) for the three genomic segments of RVFV. The effective population size times the generation time ($N_e g$) parameter approximates the number of infections in time. The plots overlay indicated complex oscillations with higher viral activity starting at around 1930, culminating sometime between the 40's and 60's, followed by a steady decrease from the 70's to the present. The stabilization of near the present is consistent with enzooticism, since the virus is not kept in human populations.
doi:10.1371/journal.pone.0035216.g004

Figure 5. Large-scale geographic spread of RVFV in Africa and Saudi Arabia based on the S segment. The directed lines connect the sources and target localities (shown by arrows) of viral lineages. The distinct introductions into Senegal and Mauritania were represented by different colors. The estimated time to the most recent common ancestor of strains from different countries are shown within rectangles and could be interpreted as the oldest possible year of introduction of that lineage at that locality.
doi:10.1371/journal.pone.0035216.g005

Figure 6. Geographic spread of RVFV in Mauritania and Senegal. The directed lines connect the sources and target localities (shown by arrows) of viral lineages. The distinct introductions into Senegal and Mauritania were represented by different colors. The estimated time to the most recent common ancestor of strains from different countries are shown within rectangles and could be interpreted as the oldest possible year of introduction of that lineage at that locality.
doi:10.1371/journal.pone.0035216.g006

without heavy rain falls [13]. Moreover, an enzootic cycle has been described in Barkedji, which could be maintained by vertical transmission among *Aedes* mosquitoes [27].

Crucially, our results also support the notion that Barkedji functions as a hub, broadcasting RVFV to other localities in West Africa. At a coarse grain, we observed that the Zimbabwe spread viruses during the first half of last century, while Kenya experienced a more intense activity as a hub during the second half. The notion that Barkedji may be an important gateway to RVFV in Senegal and Mauritania was previously suggested by serologic and entomologic surveys [13,14]. An important role of Barkedji appears to be independently supported by the fact that it is a known crossroad of migration movements of herds between the southern and northern regions of Senegal and, to a larger extent, to southern Mauritania. Moreover, the distinct introductions in Barkedji may help explain the maintenance of the enzootic cycle at the regional scale, by feeding the zoonotic pool required for its persistence [35]. Because mosquitoes cannot fly more than a few hundred meters during their lifetime, their role in long-range dissemination may be limited [38]. This important role however,

could be fulfilled by infected members of roaming herds that could help the introduction and dispersion of RVFV strains across Africa [39]. Perhaps the lack of well-characterized reservoirs in both sides of the Senegalese Valley could be explained by a constant replenishment of potential hosts in the area. Our data agrees with the notion that infected hosts may regularly introduce and reintroduce viruses through places such as Barkedji, possibly through migratory routes. Nevertheless, any additional understanding of the putative role of Barkedji as a RVFV gateway to Senegal and Mauritania requires investigating a multitude of factors and processes that ultimately would help maintain and replenish the enzootic pool.

Materials and Methods

Ethics Statements and Clinical samples

The study was submitted and approved by the Ethics Committee of Institute Pasteur of Dakar in Senegal. Samples from Senegal and Mauritania are part of the Institute Pasteur in Dakar CRORA collection (Centre collaborateur OMS de

référence et de recherche pour les arbovirus et virus de fièvres hémorragiques). Animal and human samples were sampled from 1983 to 2003 at outbreak sites. Human samples used to isolate the virus were sent by physicians as part of routine diagnostics procedures and were verbally authorized (when possible) by patients that were kept anonymous. On the other hand, samples obtained from vertebrates and insects, were provided by veterinarians and entomologists from national health authorities during RVFV outbreaks. No ethical statement for animal experiments was demanded, since no animal experiments were conducted, other than sampling near human cases. Nevertheless, all animal blood sampling was performed while minimizing suffering. In addition, all samples of the study were collected in the course of routine surveillance by national health authorities.

Virus Isolates

The strains used were obtained from mosquitoes, humans and animals in Senegal and Mauritania. Date of isolation of strains and animal source are listed in Table S1, geographical origins are shown in Figure S2. RVFV samples were obtained from the virus collection of the WHO reference center for arboviruses research in the Institute Pasteur of Dakar. They have been propagated in Vero E6 or AP61 cells cultured in Leibovitz medium L15 supplemented with 5% fetal bovine serum, antibiotics, Fungizone and Tryptose (for AP61 cells). Viruses were harvested after infected cells presented a cytopathic effect. Additional reference sequences from other localities in Africa were obtained from GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>) and they were listed in Dataset S2.

Viral RNA extraction and amplification

Viral RNA was extracted from cells culture supernatant by using the QIAgen Viral RNA mini kit (QIAgen, Valencia, CA) according to the manufacturer's instructions and collected in 50 μ l of elution buffer. Partial regions of the G2 (718 bp), NSs (601 bp) and L (129 bp) genes were amplified using the sets of primers MRV1a-MRV2g, NS3a-NS2g, and Wag-Xg respectively (Table S3). Reverse transcription-PCR reactions for the cDNA synthesis were performed for NSs and G2 using the AMV Reverse Transcriptase (M5101). Taq DNA polymerase (M1865) of Promega Corporation (Wisconsin, USA) was used for the PCR amplification of both regions. Primers sequences and protocols were previously described [29]. Whereas for the L region the RT-PCR was amplified with the superscript II (SII) (Invitrogen, USA) and Takara taq DNA polymerase (Takara, USA) for the synthesis of the cDNA and polymerase amplification respectively. First strand cDNA synthesis was performed in a final volume of 20 μ l by incubating first 10 μ l of eluted RNA, 1 μ l dNTP (10 mM-Amersham), 1 μ l of reverse primer (100 μ g/ μ l) and 1 μ l H₂O at 65°C for 5 min and then rapidly chilled on ice. The following mixture: 4 μ l 5xbuffer, 1 μ l DTT, 1 μ l RNasin Ribonuclease Inhibitor 2500 U (Promega) and 1 μ l SII transcriptase (*18080-044*) was added before an incubation at 50°C for 60 min followed by a heat inactivation at 70°C for 15 min. The amplification reaction was initiated with 5 μ l of the resulting cDNA as template. Thermocycler program was: one cycle of 95°C for 2 min 30 sec, 45 cycles each one with 95°C for 45 sec, 45°C for 30 sec and 72°C for 45 sec, and one final extension cycle at 72°C for 5 min. RT-PCR amplification from field sample isolates was done in the Institute Pasteur in Dakar, Senegal.

Sequencing

RT-PCR products were separated by electrophoresis on a 1% agarose gel. Bands of the appropriate molecular size were excised and DNA was recovered using the QiaQuick Gel Extraction Kit (QIAgen) as specified by the manufacturer. Both strands sequencing was performed using the same reverse and forward primers as for the amplification. Sequencing reactions were purified by precipitation: (Hi-Di from Applied Biosystems Foster City, CA) and finally run on an ABI Prism 3100 Genetic Analyzer (Applied Biosystems). DNA sequencing was done at the Microbiology Department of the Biomedical Sciences Institute at the University of São Paulo, Brazil. GenBank accession numbers for S segment sequences are from JN995251 to JN995298 and for M sequences are from JN995299 to JN9954345, while sequences from the L segment portion of the polymerase gene are available from the authors upon request because GenBank requires sequences longer than 200 nucleotides.

Sequence analysis

Sequences were inspected and assembled with Geneious Pro3.5.4 program (<http://www.geneious.com/>) and aligned using the multiple sequence alignment algorithm Clustal W [40] and Se-Al v 2.0 (<http://tree.bio.ed.ac.uk/software/seal/>). To prevent potential biases during phylogenetic inference due to recombination, we first analyzed all sequences with RDP4beta 4.8 program that incorporates RDP, GENECONV, Chimaera, Maxchi, Bootscan, SiScan and 3Seq [41] to uncover evidence for recombination events. Only events with p -values ≤ 0.05 that were detected by three or more methods were considered using the Bonferroni correction to prevent false positive results. In addition, intending to infer the selection pressures acting on each genomic segment of RVFV, we estimated the difference between the non-synonymous (dN) and synonymous (dS) rates per codon site using the Single Likelihood Ancestor Counting (SLAC) algorithm available in HyPhy v0.99 [42], assuming a significance level of 1% ($\alpha = 0.01$). In the HyPhy output, values of ω are expressed as $\omega = dN/dS$. Therefore, ω greater than zero are indicative of directional, positive selection ($\omega > 0$), while values below zero ($\omega < 0$) indicate purifying, negative selection.

Phylogenetic Analyses

Prior to the analyses, the phylogenetic signal content of the sequence datasets to phylogenetic reconstruction was investigated by Likelihood mapping [43] with TREE-PUZZLE [44]. Phylogenetic trees were generated by Maximum Likelihood (ML) criterion using GARLI v0.96 [45] that uses a stochastic genetic algorithm to estimate simultaneously the best topology, branch lengths and substitution model parameters that maximize the log-Likelihood (lnL). The confidence of ML trees was assessed by the convergence of lnL scores from ten independent replicates. Likewise previous RVFV reports [23], GTR model with Gamma-distributed rate variation (γ) and a proportion of invariable sites (I) substitution model was used. Since we had dates of isolation, we estimated substitution rates per site per year (μ) with R8s v1.71 [46] using the Penalized Likelihood method [47] that employs a semi-parametric approach, using different substitution rates on every branch with a nonparametric roughness penalty, which impose costs according to the model once rates change too quickly from branch to branch. In addition, Maximum Clade Credibility (MCC) trees were inferred using a Markov Chain Monte Carlo (MCMC) Bayesian approach under GTR+ γ +I and a relaxed (uncorrelated lognormal) molecular clock [48] with the μ previously estimated on the program BEAST v1.6.1 [49]. Moreover, we investigated the variation in effective population size times the generation time (Ne.g) to infer viral demography by Bayesian Skyride Plots [50]. MCMC convergence

was obtained during four independent runs with 50 million of generations, which were sufficient to obtain a proper sample from the posterior at MCMC stationarity. The stationarity of parameters was also assessed by allowing the effective sample sizes (ESS) to reach values above 200 as inspected using Tracer v1.5 (<http://tree.bio.ed.ac.uk/software/tracer/>). Furthermore, to infer the history of geographical dispersion of RVFV strains, we used a discrete model attributing state characters representing isolation locality of each of the strains with the Bayesian Stochastic Search Variable (BSSVS) algorithm [51] implemented in BEAST v1.6.1. This method estimates the most probable state at each node in the MCC trees, allowing us to reconstruct ancestral positions for ancestral viral lineages along the tree. For phylogeographic reconstructions, each locality was coded as a discrete trait. For Senegal and Mauritania we coded for each locality, down to the city level, while for other African countries we coded localities at national level (coarse grain). We argue that this approach was justified since it allow to focus on the movement within and between Senegal and Mauritania at a finer grain, while also showing its relationship with other African countries at a coarse grain. Likewise, to map routes of spread we resolved cities in Senegal and Mauritania and pinned incoming and outgoing routes at the center of the other African countries.

Supporting Information

Figure S1 Maximum clade credibility (MCC) tree summarizing geographical states reconstructions along a time-scaled tree. Ancestral states reconstructions indicating the most probable location of a lineage in time along colored branches and dates of nodes are shown. Countries are ISO coded as follows: Burkina Faso (BF), Central African Republic (CF), Egypt (EG), Guinea (GN), Kenya (KE), Madagascar (MG), Mauritania (MR), Namibia (NA), Saudi Arabia (SA), Senegal (SN), South Africa (ZA), Tanzania (TZ), Uganda (UG), Zimbabwe (ZW), Angola (AO), Zambia (ZM), Somalia (SO). (TIF)

Figure S2 Map from Senegalese and Mauritanian territories. Circles represent the locations from RVFV isolates. (TIF)

References

- Laughlin LW, Meegan JM, Strausbaugh LJ, Morens DM, Watten RH (1979) Epidemic Rift Valley fever in Egypt: observations of the spectrum of human illness. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 73: 630–633.
- Rich KM, Wanyike F (2010) An assessment of the regional and national socio-economic impacts of the 2007 Rift Valley fever outbreak in Kenya. *The American journal of tropical medicine and hygiene* 83: 52–57. doi:10.4269/ajtmh.2010.09-0291.
- Müller R, Poch O, Delaune M, Bishop DH, Bouloy M (1994) Rift Valley fever virus L segment: correction of the sequence and possible functional role of newly identified regions conserved in RNA-dependent polymerases. *The Journal of general virology* 75(Pt 6): 1345–1352.
- Collett MS, Purchio AF, Keegan K, Frazier S, Hays W, et al. (1985) Complete nucleotide sequence of the mRNA segment of Rift Valley fever virus. *Virology* 144: 228–245.
- Giorgi C, Accardi L, Nicoletti L, Gro MC, Takehara K, et al. (1991) Sequences and coding strategies of the S RNAs of Toscana and Rift Valley fever viruses compared to those of Punta Toro, Sicilian Sandfly fever, and Unkinimbi viruses. *Virology* 180: 738–753.
- Billecoq A, Spiegel M, Vialat P, Kohl A, Weber F, et al. (2004) NSs protein of Rift Valley fever virus blocks interferon production by inhibiting host gene transcription. *Journal of virology* 78: 9798–9806. doi:10.1128/JVI.78.18.9798-9806.2004.
- Findlay GM, Daubney R (1931) The Virus of Rift Valley Fever or Enzootic Hepatitis. *The Lancet* 24: 1330–1331.
- Daubney R, Hudson JR (1932) Rift Valley Fever. *The Lancet* 6: 611–612. doi:10.1017/S0031182000003036.
- Meegan JM (1979) The Rift Valley fever epizootic in Egypt 1977–78. 1. Description of the epizootic and virological studies. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 73: 618–623.
- Martin V, Chevalier V, Ceccato P, Anyamba A, De Simone L, et al. (2008) The impact of climate change on the epidemiology and control of Rift Valley fever. *Revue Scientifique Et Technique (International Office of Epizootics)* 27: 413–426.
- Chevalier V, Thiongan Y, Lancelot R (2009) Endemic transmission of Rift Valley Fever in Senegal. *Transboundary and emerging diseases* 56: 372–374. doi:10.1111/j.1865-1682.2009.01083.x.
- Faye O, Diallo M, Diop D, Bezeid OE, Bâ H, et al. (2007) Rift Valley fever outbreak with East-Central African virus lineage in Mauritania, 2003. *Emerging infectious diseases* 13: 1016–1023.
- Zeller HG, Fontenille D, Traoré-Lamizana M, Thiongan Y, Digoutte JP (1997) Enzootic activity of Rift Valley fever virus in Senegal. *The American Journal of Tropical Medicine and Hygiene* 56: 265–272.
- Traoré-Lamizana M, Fontenille D, Diallo M, Bâ Y, Zeller HG, et al. (2001) Arbovirus surveillance from 1990 to 1995 in the Barkedji area (Ferkol) of Senegal, a possible natural focus of Rift Valley fever virus. *Journal of Medical Entomology* 38: 480–492.
- Marrama L, Spiegel A, Ndiaye K, Sall AA, Gomes E, et al. (2003) Domestic transmission of Rift Valley Fever virus in Diawara (Senegal) in 1998. *The Southeast Asian Journal of Tropical Medicine and Public Health* 36: 1487–1495.
- Sall AA, Zanotto PM, Vialat P, Sène OK, Bouloy M (1998) Molecular epidemiology and emergence of Rift Valley fever. *Memórias do Instituto Oswaldo Cruz* 93: 609–614.

Table S1 Source, geographical origin and date of isolations of RVFV strains used in this study. *Patient died. (DOC)

Table S2 Likelihood mapping of the three RVFV genomic segments. *The percentage of the unresolved quartets is an indicator of phylogenetic suitability from data under analysis. If the percentage is higher, the suitability is lower. Our results suggested that Medium segment is the best for phylogenetic inference due to the lower associated uncertainty. (DOC)

Table S3 Description of PCR primers used to amplify Senegal and Mauritania samples. *Relative position to strain MP12 whose GenBank accession numbers from Small, Medium and Large segments are respectively X53771, M11157 and X56464. (DOC)

Dataset S1 Spread of RVFV strains in Africa. A kmz file to picture the history of RVFV migration in Africa along the time, executable in Google earth (<http://earth.google.com>). (KML)

Dataset S2 Additional Reference Sequences of RVFV. An xls file showing the sequences recovered from GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>). (XLS)

Acknowledgments

We would like to thank the Editor and two anonymous referees for their thorough contribution on all relevant aspects of the manuscript.

Author Contributions

Conceived and designed the experiments: POLS OF MD PMAZ AAS. Performed the experiments: POLS OF MD JVCO. Analyzed the data: POLS CCMF AAS PMAZ. Contributed reagents/materials/analysis tools: AAS PMAZ. Wrote the paper: POLS CCMF PMAZ AAS. First co-authors: POLS CCMF.

17. Weaver SC (2005) Host range, amplification and arboviral disease emergence. *Archives of virology Supplementum*, pp 33–44.
18. World Health Organization (2007) WHO | Rift Valley Fever in Kenya, Somalia and the United Republic of Tanzania. Available: http://www.who.int/csr/don/2007_05_09/en/. Accessed 11 August 2011.
19. Archer BN, Weyer J, Paweska J, Nkosi D, Leman P, et al. (2011) Outbreak of Rift Valley fever affecting veterinarians and farmers in South Africa, 2008. *South African Medical Journal = Suid-Afrikaanse Tydskrif Vir Geneeskunde* 101: 263–266.
20. World Health Organization (2010) WHO | Rift Valley fever in South Africa. Available: http://www.who.int/csr/don/2010_03_30/en/. Accessed 10 August 2011.
21. World Health Organization (2008) WHO | Rift Valley Fever in Sudan - update 5. Available: http://www.who.int/csr/don/2008_01_22/en/index.html. Accessed 11 August 2011.
22. World Health Organization (2000) WHO | Saudi Arabia and Yemen: First cases of Rift Valley fever reported outside Africa, 2000. Available: <http://www.who.int/csr/outbreaknetwork/saudi-arabia/en/>. Accessed 11 August 2011.
23. Bird BH, Khristova ML, Rollin PE, Ksiazek TG, Nichol ST (2007) Complete genome analysis of 33 ecologically and biologically diverse Rift Valley fever virus strains reveals widespread virus movement and low genetic diversity due to recent common ancestry. *Journal of virology* 81: 2805–2816. doi:10.1128/JVI.02095-06.
24. Bird BH, Githinji JWK, Macharia JM, Kasiti JL, Muriithi RM, et al. (2008) Multiple virus lineages sharing recent common ancestry were associated with a Large Rift Valley fever outbreak among livestock in Kenya during 2006–2007. *Journal of virology* 82: 11152–11166. doi:10.1128/JVI.01519-08.
25. Grobbelaar AA, Weyer J, Leman P, Kemp A, Paweska JT, et al. (2011) Molecular epidemiology of Rift Valley fever virus. *Emerging infectious diseases* 17: 2270–2276. doi:10.3201/eid1712.111035.
26. Sall AA, de A Zanotto PM, Vialat P, Sene OK, Bouloy M (1998) Origin of 1997–98 Rift Valley fever outbreak in East Africa. *Lancet* 352: 1596–1597.
27. Pepin M, Bouloy M, Bird BH, Kemp A, Paweska J (2010) Rift Valley fever virus (Bunyaviridae: Phlebovirus): an update on pathogenesis, molecular epidemiology, vectors, diagnostics and prevention. *Veterinary research* 41: 61.
28. Sall AA, Zanotto PM, Sene OK, Zeller HG, Digoutte JP, et al. (1999) Genetic reassortment of Rift Valley fever virus in nature. *Journal of virology* 73: 8196–8200.
29. Sall AA, de A Zanotto PM, Zeller HG, Digoutte JP, Thiongang Y, et al. (1997) Variability of the NS(S) protein among Rift Valley fever virus isolates. *The Journal of general virology* 78(Pt 11): 2853–2858.
30. Moutailler S, Roche B, Thiberge J-M, Caro V, Rougeon F, et al. (2011) Host alteration is necessary to maintain the genome stability of rift valley fever virus. *PLoS neglected tropical diseases* 5: e1156. doi:10.1371/journal.pntd.0001156.
31. Wertheim JO, Kosakovsky Pond SL (2011) Purifying selection can obscure the ancient age of viral lineages. *Molecular biology and evolution* doi:10.1093/molbev/msr170.
32. Wiens JJ (2003) Missing data, incomplete taxa, and phylogenetic accuracy. *Systematic Biology* 52: 528–538.
33. Johnson K (1993) Emerging Viruses in Context: An Overview of Viral Hemorrhagic Fevers. In: Morse S, ed. *Emerging Viruses*. New York: Oxford University Press. 317 p.
34. Pfeiffer M, Dobler G (2010) Emergence of zoonotic arboviruses by animal trade and migration. *Parasites & Vectors* 3: 35. doi:10.1186/1756-3305-3-35.
35. Favier C, Chalvet-Monfray K, Sabatier P, Lancelot R, Fontenille D, et al. (2006) Rift Valley fever in West Africa: the role of space in endemicity. *Tropical Medicine & International Health: TM & IH* 11: 1878–1888. doi:10.1111/j.1365-3156.2006.01746.x.
36. Gad AM, Feinsod FM, Allam IH, Eisa M, Hassan AN, et al. (1986) A possible route for the introduction of Rift Valley fever virus into Egypt during 1977. *The Journal of Tropical Medicine and Hygiene* 89: 233–236.
37. Müller BR, Godsey MS, Crabtree MB, Savage HM, Al-Mazrao Y, et al. (2002) Isolation and genetic characterization of Rift Valley fever virus from *Aedes vexans* arabiensis, Kingdom of Saudi Arabia. *Emerging Infectious Diseases* 8: 1492–1494.
38. Yamar BA, Diallo D, Kebe CMF, Dia I, Diallo M (2005) Aspects of bioecology of two Rift Valley Fever Virus vectors in Senegal (West Africa): *Aedes vexans* and *Culex poicilipes* (Diptera: Culicidae). *Journal of Medical Entomology* 42: 739–750.
39. LaBeaud AD, Cross PC, Getz WM, Glinka A, King CH (2011) Rift Valley fever virus infection in African buffalo (*Syncerus caffer*) herds in rural South Africa: evidence of interepidemic transmission. *The American Journal of Tropical Medicine and Hygiene* 84: 641–646. doi:10.4269/ajtmh.2011.10-0187.
40. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* 22: 4673–4680.
41. Martin DP, Lemey P, Lott M, Moulton V, Posada D, et al. (2010) RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics (Oxford, England)* 26: 2462–2463. doi:10.1093/bioinformatics/btq467.
42. Pond SLK, Frost SDW, Muse SV (2003) HyPhy: hypothesis testing using phylogenies. *Bioinformatics (Oxford, England)* 21: 676–679. doi:10.1093/bioinformatics/btg079.
43. Strimmer K, von Haeseler A (1997) Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proceedings of the National Academy of Sciences of the United States of America* 94: 6815–6819.
44. Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics (Oxford, England)* 18: 502–504.
45. Zwickl D (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion.
46. Sanderson MJ (2003) r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics (Oxford, England)* 19: 301–302.
47. Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular biology and evolution* 19: 101–109.
48. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS biology* 4: e88. doi:10.1371/journal.pbio.0040088.
49. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7: 214. doi:10.1186/1471-2148-7-214.
50. Minin VN, Bloomquist EW, Suchard MA (2008) Smooth skyline through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular biology and evolution* 25: 1459–1471. doi:10.1093/molbev/msn090.
51. Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian phylogeography finds its roots. *PLoS computational biology* 5: e1000520. doi:10.1371/journal.pcbi.1000520.

Reemergence of Rift Valley Fever, Mauritania, 2010

Ousmane Faye, Hampathé Ba, Yamar Ba, Caio C.M. Freire, Oumar Faye, Oumar Ndiaye, Isselmou O. Elgady, Paolo M.A. Zanotto, Mawlouth Diallo, and Amadou A. Sall

A Rift Valley fever (RVF) outbreak in humans and animals occurred in Mauritania in 2010. Thirty cases of RVF in humans and 3 deaths were identified. RVFV isolates were recovered from humans, camels, sheep, goats, and *Culex antennatus* mosquitoes. Phylogenetic analysis of isolates indicated a virus origin from western Africa.

Rift Valley fever (RVF) is an acute febrile viral disease that affects domestic ruminants and humans. The disease in animals is characterized by abortions among pregnant females and high mortality rates for offspring (1). The causative pathogen, RVF virus (RVFV), is a member of the family *Bunyaviridae*, genus *Phlebovirus*, and its genome consists of 3 RNA segments: large (L), medium (M), and small (S) (2). In human infections, symptoms are generally mild but may evolve to severe symptoms, such as hemorrhaging, meningoencephalitis, and retinopathy, with fatal outcomes (3). RVF is endemic to sub-Saharan Africa, Egypt, Saudi Arabia, and Yemen (4). In western Africa, a major RVF outbreak occurred in Mauritania and Senegal in 1987 and resulted in 220 human deaths (5). This outbreak was followed by epizootics/epidemics in 1998 and 2003 in different provinces in Mauritania (Hodh El Gharbi, Assaba, Brakna, Trarza, and Gorgol Provinces) (6,7).

In October 2010, the health services of Adrar Province in Mauritania were informed of the deaths of 2 girls with hemorrhagic syndromes and 2 persons with acute fever, arthralgia, and headaches in Amogjar Province. Blood samples collected from a sick patient were sent to Institut Pasteur in Dakar, Senegal where RVFV IgM and RVFV genome were detected. During October 28–November 11, 2010, suspected RVF cases were identified in several cities Adrar Province (Atar, Aoujeft, Chinguetti), Inchiri Province (Akjoujt), and Hodh el Gharbi Province (Tintane, Kobeni)

(Figure 1), and high mortality rates among camels, sheep, and goats were reported in Adrar and Inchiri Provinces (8). Clinical samples from humans and animals with suspected RVF were positive for this virus.

These cases led to identification of 2 additional confirmed cases of acute RVF (positive for IgM against RVFV or RVFV RNA) in Adrar and Inchiri Provinces. Multidisciplinary investigations involving epidemiologists, virologists, entomologists, and veterinary teams in areas where the cases occurred were conducted in December 2010 as part of other investigations (8), describe the extent of the outbreak, and investigate possible RVF reemergence factors in Mauritania. We report results of laboratory and field investigations among humans, animals, and mosquitoes during the 2010 outbreak.

The Study

During October–December 2010, surveillance and investigation teams collected 80 clinical samples from 36 humans with suspected RVF and 44 of their contacts in Atar, Aoujeft, Akjoujt, Chinguetti, Kobeni, and Tintane Moughatas (Figure 1). A human with a suspected case of RVF was defined as any patient with fever associated with or without hemorrhage, jaundice, or neurologic symptoms during October–December. A human contact was defined as a healthy person living within the immediate environment (i.e., same household or neighborhood) of a human or animal with a confirmed case of RVF. Serum samples from suspected case-patients were tested by using real time reverse transcription PCR, ELISA (for detection of IgM), and virus isolation.

Samples were obtained from animals living near suspected or confirmed case-patients, and a questionnaire about the species, sex, age and reproductive status of female animals was completed by their owners. Sampling of mosquitoes was conducted in December 2010 by using animal-baited traps placed in and around houses in which suspected and/or confirmed case-patients were identified. RVFV isolation and identification was performed on mosquitoes and human serum samples as described (9). Coding regions for RVFV nonstructural protein and glycoprotein 2 were amplified by using reverse transcription PCR and described sets of primers (10,11) and then purified after electrophoresis on agarose gels, and sequenced for phylogenetic analyses.

IgM against RVFV or viral RNA genome was detected in 30 (37.5%) of 80 patients, and 6 RVFV isolates were recovered (Table 1). The dates of symptom onset for confirmed cases of RVF in humans indicated that the outbreak probably started at the end of October, peaked in November, and decreased the second week of December. Among 30 patients (8 contacts and 22 suspected case-patients) with acute RVFV infections, the male: female ratio

Author affiliations: Institut Pasteur, Dakar, Senegal (O. Faye, Y. Ba, O. Faye, M. Diallo, A.A. Sall); Institut National de Recherche en Santé Publique, Nouakchott, Mauritania (H. Ba, I. O. Elgady); and University of São Paulo, São Paulo, Brazil (C.C.M. Freire, P.M.A. Zanotto)

DOI: <http://dx.doi.org/10.3201/eid2002.130996>

Figure 1. Location of study sites tested for patients with confirmed Rift Valley fever, Mauritania, 2010.

was 5:1 (25:5), and the median age was 35 years (range 17–70 years); 3 deaths were reported. Except for 1 man (a teacher), all infected men were herdsmen and all infected women were housewives.

For the animal investigation, serum samples were obtained and processed from 83 small ruminants (70 goats and 13 sheep) with a median age 4 years (range 1–10 years) from all localities tested (Table 2, Appendix, wwwnc.cdc.gov/EID/article/20/2/13-0996-T2.htm). IgM against RVFV was detected in 23 (27.7%) of 83 animals. In addition, samples from 5 sick camels were collected in neighborhoods of persons with suspected and confirmed cases of RVF, and RVFV RNA was detected in 3 of these samples. Information for populations living in Adrar indicated that the first probable case of the disease among animals was in a sick camel in Aoujeft during the last week of October (8).

A total of 2,741 mosquitoes belonging to 5 genera and 11 species were collected for entomologic investigation. *Culex antennatus* was the most abundant mosquito species (83%), followed by *Anopheles gambiae* (9%) and *An. rufigipes* (2%) mosquitoes (Table 3). Three RVFV strains were isolated from *Cx. antennatus* mosquitoes collected in Safia (Adrar Province).

Phylogenetic analysis based on RVFV partial small and medium RNA sequences showed a low level of variation between isolates from humans, animals, and mosquitoes. Isolates belonged to the West Africa lineage and were closely related to strains from Diawara (1998) and Hodh El

Garbi (1998) in Mauritania and from Kenya (1965), Zimbabwe (1974), and Zambia (1975) (Figure 2).

Conclusions

After a 4-fold increase in rainfall in Mauritania from 2009 through 2010 (59 mm vs. 196 mm; Aoujeft Weather Station), RVF clinical cases were confirmed in Adrar and Inchiri Provinces. Although RVF was reported in some provinces of Mauritania in 1987, 1998, and 2003 (5–7), cases in humans from Adrar and Inchiri Provinces have not been previously reported. However, in 2010 a total of 30 laboratory-confirmed human case-patients were recorded in these 2 provinces; 3 of the case-patients died. Both areas had a drought in recent years or their average rainfall was not >60 mm before 2010. The study showed animal abortion rates of 26.13% (23/88) and 30.68% (27/88) in 2 groups of small ruminants living near patients with recent RVFV infections. Three camels were infected with RVFV, and the cases were probably linked to human infections. The first human death from RVF was reported on November 11 in the village of Tawaz, in which a high mortality rates were observed in camel populations (11).

Serologic evidence of RVF in camels from northern Sudan and southern Egypt in 1977 (12) led to camels being suspected as key contributors in RVFV transmission (13). The involvement of camels in human infections could explain the high male:female ratio (25:5) for infections in humans because men are more involved in animal

Table 1. Seroepidemiologic and virologic results for humans, Rift Valley Fever outbreak, Mauritania, 2010*

City	No. tested	No. positive for IgM against RVFV	No. positive for RVFV by RT-PCR	No. with isolated virus
Atar	54	12	10	3
Aoujeft	12	1	7	3
Akjoujt	10	0	4	0
Chinguetti	2	0	1	0
Kobeni	1	0	0	0
Tintane	1	0	0	0
Total	80	0	22	6

*RVFV, Rift Valley fever virus; RT-PCR, reverse transcription PCR.

Table 3. Virologic results for mosquitoes tested during a Rift Valley fever outbreak, Mauritania, 2010*

Mosquito species	No. tested	No. mosquito pools	No. mosquitoes positive for RVFV by RT-PCR	No. with isolated virus
<i>Aedes vexans</i>	13	3	0	0
<i>Anopheles gambiae</i>	259	10	0	0
<i>Anopheles pharoensis</i>	51	7	0	0
<i>Anopheles rufipes</i>	47	7	0	0
<i>Anopheles pretoriensis</i>	3	1	0	0
<i>Anopheles gambiae</i>	36	2	0	0
<i>Anopheles pharoensis</i>	2,276	46	4	4
<i>Anopheles rufipes</i>	16	4	0	0
<i>Anopheles pretoriensis</i>	30	5	0	0
Total	2,736	85	4	3

*RVFV, Rift Valley fever virus; RT-PCR, reverse transcription PCR.

husbandry. In 2006, a similar sex difference for infection was observed in Kenya, where men had a prevalence rate of IgG against RVFV that was 3 times higher than that in women (14).

Entomologic investigations in Mauritania showed that only *Cx. antennatus* mosquitoes were infected with RVFV. Similar findings were reported for previous RVF outbreaks in Mauritania. Furthermore, *Cx. antennatus* mosquitoes were found to be infected with RVFV in Nigeria in 1967 and in Kenya during 1981–1984, and vector competence was demonstrated in laboratory studies (15).

Phylogenetic analysis suggests reemergence of the local RVFV focus and emphasizes the need to strengthen RVF surveillance and other control measure in Mauritania after heavy rainfall.

Acknowledgments

We thank Magueye Ndiaye, Oumar Ndiaye, Carlos Fortez, Moussa Dia, Mireille Mondo, and Roughyetou Sylla providing for excellent technical assistance with laboratory and field investigations; and the authorities and the field agents of the Ministry of Health of Mauritania for facilitating investigation of this outbreak.

Figure 2. Maximum-likelihood trees for the RNA A) small and B) medium segments of Rift Valley fever virus. Trees show relationships among strains isolated from different localities and countries. Samples from Mauritania 2010 are indicated in **boldface** and designated SH211856_RIM_2010, SH211857_RIM_2010, and SH211977_RIM_2010. GenBank accession nos. are KF717588, KF717589, and KF717590 for the medium segment and KF717591, KF717592, and KF717593 for the small segment. Values along the branches indicate bootstrap values. Scale bars indicate nucleotide substitutions per site.

This study was supported by grants from the Institut Pasteur de Dakar, Senegal.

Dr Ousmane Faye is a virologist and acarologist at the Arbovirus and Viral Hemorrhagic Fever Unit, Institut Pasteur, Dakar, Senegal. His primary research interests include the virology and vectorial transmission of arboviruses and viral hemorrhagic fevers.

References

1. Daubney R, Hudson JR, Garnham PC. Enzootic hepatitis or Rift Valley fever: an undescribed virus disease of sheep, cattle, and man from east Africa. *Journal of Pathology and Bacteriology*. 1931;34:545–79. <http://dx.doi.org/10.1002/path.1700340418>
2. Murphy FA, Fauquet CM, Bishop DH, Ghabrial SA, Jarvis AW, Martelli GP, et al. Family Bunyaviridae. In: *Virus taxonomy: Classification and nomenclature of viruses*. Sixth report of the International Committee on Taxonomy of Viruses. New York: Springer-Verlag; 1995. p. 300–15.
3. Laughlin LW, Meegan JM, Strausbaugh LJ, Morens DM, Watten RH. Epidemic Rift Valley fever in Egypt: observations of the spectrum of human illness. *Trans R Soc Trop Med Hyg*. 1979;73:630–3. [http://dx.doi.org/10.1016/0035-9203\(79\)90006-3](http://dx.doi.org/10.1016/0035-9203(79)90006-3)
4. Balkhy HH, Memish ZA. Rift Valley fever: an uninvited zoonosis in the Arabian Peninsula. *Int J Antimicrob Agents*. 2003;21:153–7. [http://dx.doi.org/10.1016/S0924-8579\(02\)00295-9](http://dx.doi.org/10.1016/S0924-8579(02)00295-9)
5. Digoutte JP, Peters CJ. General aspects of the 1987 Rift Valley fever epidemic in Mauritania. *Res Virol*. 1989;140:27–30. [http://dx.doi.org/10.1016/S0923-2516\(89\)80081-0](http://dx.doi.org/10.1016/S0923-2516(89)80081-0)
6. Nabeth P, Kane Y, Abdalahi MO, Diallo M, Ndiaye K, Ba K, et al. Rift Valley fever outbreak in Mauritania in 1998: seroepidemiological, virological, entomological and zoological investigation. *Emerg Infect Dis*. 2001;7:1052–4. <http://dx.doi.org/10.3201/eid0706.010627>
7. Faye O, Diallo M, Diop D, Bezeid OE, Bâ H, Niang M, et al. Rift Valley fever outbreak with East-Central African virus lineage in Mauritania, 2003. *Emerg Infect Dis*. 2007;13:1016–23. <http://dx.doi.org/10.3201/eid1307.061487>
8. El Mamy AB, Baba MO, Barry Y, Isselmou K, Dia ML, El Kory MO, et al. Unexpected Rift Valley fever outbreak, northern Mauritania. *Emerg Infect Dis*. 2011;17:1894–6. <http://dx.doi.org/10.3201/eid1710.110397>
9. Digoutte JP, Calvo-Wilson MA, Mondo M, Traore-Lamizana M, Adam F. Continuous cell lines and immune ascitic fluid pools in arbovirus detection. *Res Virol*. 1992;143:417–22. [http://dx.doi.org/10.1016/S0923-2516\(06\)80135-4](http://dx.doi.org/10.1016/S0923-2516(06)80135-4)
10. Muller R, Saluzzo JF, Lopez N, Dreier T, Turell M, Smith J, et al. Characterization of clone 13, a naturally attenuated avirulent isolate of Rift Valley fever virus, which is altered in the small segment. *Am J Trop Med Hyg*. 1995;53:405–11.
11. Takehara K, Min MK, Battles JK, Sugiyama K, Emery VC, Dalrymple JM, et al. Identification of mutations in the M RNA of a candidate vaccine strain of Rift Valley fever virus. *Virology*. 1989;169:452–7. [http://dx.doi.org/10.1016/0042-6822\(89\)90171-2](http://dx.doi.org/10.1016/0042-6822(89)90171-2)
12. Davies FG, Koros J, Mbugua H. Rift Valley fever in Kenya: the presence of antibodies to the virus in camels (*Camelus dromedarius*). *J Hyg (Lond)*. 1985;94:241–4. <http://dx.doi.org/10.1017/S0022172400061441>
13. Eisa M, Obeid HMA, El Sawi AS. Rift Valley fever in the Sudan. I. Results of field investigations of the first epizootic in Kosti District, 1973. *Bulletin of Animal Health and Production in Africa*. 1977;24:343–7.
14. LaBeaud AD, Muchiri EM, Ndlovu M, Mwanje MT, Muiruri S, Peters CJ, et al. Inter-epidemic Rift Valley fever virus seropositivity, northeastern Kenya. *Emerg Infect Dis*. 2008;14:1240–6. <http://dx.doi.org/10.3201/eid1408.080082>
15. Linthicum KJ, Davies FG, Kairo A, Bailey CL. Rift Valley fever virus (family *Bunyaviridae*, genus *Phlebovirus*). Isolations from *Diptera* collected during an inter-epizootic period in Kenya. *J Hyg (Lond)*. 1985;95:197–209. <http://dx.doi.org/10.1017/S0022172400062434>

Address for correspondence: Amadou A. Sall, Unité des Arbovirus et Virus de Fièvre Hémorragique, Institut Pasteur, 36 Ave Pasteur, Boîte Postale 220, Dakar, Senegal; email: asall@pasteur.sn

EID Online Podcast
For Emerging Infectious Diseases content on the go,
see podcast link at www.cdc.gov/eid

Anexo V

PLOS Neglected Tropical Diseases

Reassortment and distinct evolutionary dynamics of Rift Valley Fever virus genomic segments

--Manuscript Draft--

Manuscript Number:	
Full Title:	Reassortment and distinct evolutionary dynamics of Rift Valley Fever virus genomic segments
Short Title:	RVFV reassortment and different evolutionary dynamics
Article Type:	Research Article
Keywords:	Rift Valley Fever virus; Reassortment; Evolutionary dynamics; Genomic segments
Corresponding Author:	Paolo M. A. Zanotto University of São Paulo São Paulo, São paulo BRAZIL
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of São Paulo
Corresponding Author's Secondary Institution:	
First Author:	Caio C. M. Freire
First Author Secondary Information:	
Order of Authors:	Caio C. M. Freire
	Atila Iamarino
	Peinda O. Ly Soumare
	Ousmane Faye
	Amadou A. Sall
	Paolo M. A. Zanotto
Order of Authors Secondary Information:	

Abstract:	<p>Rift Valley Fever virus (RVFV) is a member of Bunyaviridae family that causes a febrile disease affecting mainly ruminants and occasionally humans in Africa, with symptoms that range from mild to severe. RVFV has a tri-segmented ssRNA genome that permits reassortment and could generate new more virulent strains. In this study, we reveal the importance of reassortment for RVFV evolution using viral gene genealogy inference and phylodynamics. We uncovered seven events of reassortment that originated RVFV lineages with discordant origins among segments. Moreover, we also found that despite similar selection regimens, the three segments have distinct evolutionary dynamics; the longer segment L evolves at a significant lower rate. Episodes of discordance between population size estimates per segment also coincided with reassortment dating. Our results show that RVFV segments are decoupled enough to have distinct demographic histories and to evolve under different molecular rates.</p>
Suggested Reviewers:	<p>Edward Holmes University of Sydney edward.holmes@sydney.edu.au</p> <p>Scott. C Weaver University of Texas Medical Branch at Galveston sweaver@utmb.edu</p> <p>Michèle Bouloy Institut Pasteur</p>
	<p>michele.bouloy@pasteur.fr</p> <p>Giovanna Carpi Yale University giovanna.carpi@yale.edu</p>
Opposed Reviewers:	
Additional Information:	
Question	Response

<p>Data Availability</p> <p>PLOS journals require authors to make all data underlying the findings described in their manuscript fully available, without restriction and from the time of publication, with only rare exceptions to address legal and ethical concerns (see the PLOS Data Policy and FAQ for further details). When submitting a manuscript, authors must provide a Data Availability Statement that describes where the data underlying their manuscript can be found.</p> <p>Your answers to the following constitute your statement about data availability and will be included with the article in the event of publication. Please note that simply stating 'data available on request from the author' is not acceptable. If, however, your data are only available upon request from the author(s), you must answer "No" to the first question below, and explain your exceptional situation in the text box provided.</p> <p>Do the authors confirm that all data underlying the findings described in their manuscript are fully available without restriction?</p>	<p>Yes - all data are fully available without restriction</p>
<p>Please describe where your data may be found, writing in full sentences. Your answers should be entered into the box below and will be published in the form you provide them, if your manuscript is accepted. If you are copying our sample text below, please ensure you replace any instances of XXX with the appropriate details.</p> <p>If your data are all contained within the paper and/or Supporting Information files, please state this in your answer below. For example, "All relevant data are within the paper and its Supporting Information files."</p> <p>If your data are held or will be held in a public repository, include URLs, accession numbers or DOIs. For example, "All XXX files are available from the XXX</p>	<p>All relevant data are within the paper and its Supporting Information files.</p>

database (accession number(s) XXX, XXX)." If this information will only be available after acceptance, please indicate this by ticking the box below. If neither of these applies but you are able to provide details of access elsewhere, with or without limitations, please do so in the box below. For example:

"Data are available from the XXX Institutional Data Access / Ethics Committee for researchers who meet the criteria for access to confidential data."

"Data are from the XXX study whose authors may be contacted at XXX."

* typeset

Additional data availability information:

INSTITUTO DE CIÊNCIAS BIOMÉDICAS-USP
Excelência em Ensino e Pesquisa

Homepage: www.icb.usp.br

Cidade Universitária – Ed. Biomédicas II

Av. Prof. Lineu Prestes, 1374 – Cep. 05508-900 São Paulo, SP - Brasil

Telefone : (55) (11) 3091-7259 – telefax: (55) (11) 3091 7354

Homepage: www.icb.usp.br/bmm

e-mail: microsec@icb.usp.br

São Paulo, October 15, 2014

Dear *PLoS Neglected Tropical diseases* editors,

We are submitting the article entitled “Reassortment and distinct evolutionary dynamics of Rift Valley Fever virus genomic segments” by Freire *et al.* We believe this manuscript is interesting to *PLoS Genetics* readers, given that we show evolutionary adaptive mechanisms by which the virus, now with mortality rates above 30%, may further increase its potential as an emergent disease.

The tri-segmented genome of this virus is known to allow recombination through reassortment, but not to the extent that we now report. Using all data currently available for RVFV and using a reassortment analysis framework developed for influenza virus, we identified RVFV reassortants. Reassortment was shown to allow for ‘shifts’ in the capabilities of several other important pathogens, but our report raises concerns also because it questions previous phylogenetic studies on RVFV which, not only disregarded systematic errors due to the mixing of recombinant genes during phylogenetic inference, but also did not account for different mutation rates and ancestral relations among segments.

These reassortment events could have broader implications, since we show that despite similar purifying selection regimens the genome segments evolve under different mutation rates and show distinct demographic histories. Reassortment has been poorly described for RVFV but is known to contribute for increased pathogenicity for other members of the Bunyaviridae family.

Yours faithfully

Prof. Paolo Zanotto, *D. PHIL.*

Reassortment and distinct evolutionary dynamics of Rift Valley Fever virus genomic segments

Caio C. M. Freire¹, Atila Iamarino¹, Peinda O. Ly Soumaré², Ousmane Faye², Amadou A. Sall², Paolo M. A. Zanotto^{1*}

1. Laboratory of Molecular Evolution and Bioinformatics, Department of Microbiology, Biomedical Sciences Institute, University of Sao Paulo, Sao Paulo, Brazil.
2. Institut Pasteur de Dakar, Dakar, Senegal.

*Corresponding author: Paolo M. A. Zanotto. Laboratory of Molecular Evolution and Bioinformatics, Department of Microbiology, Biomedical Sciences Institute – ICBII, University of São Paulo, Av. Prof. Lineu Prestes, 1734 São Paulo - SP - Brazil – 05508-000. E-mail: pzanotto@usp.br. Phone: +55 1130918453. Fax: +55 1130917354.

Abstract

Rift Valley Fever virus (RVFV) is a member of Bunyaviridae family that causes a febrile disease affecting mainly ruminants and occasionally humans in Africa, with symptoms that range from mild to severe. RVFV has a tri-segmented ssRNA genome that permits reassortment and could generate new more virulent strains. In this study, we reveal the importance of reassortment for RVFV evolution using viral gene genealogy inference and phylodynamics. We uncovered seven events of reassortment that originated RVFV lineages with discordant origins among segments. Moreover, we also found that despite similar selection regimens, the three segments have distinct evolutionary dynamics; the longer segment L evolves at a significant lower rate. Episodes of discordance between population size estimates per segment also coincided with reassortment dating. Our results show that RVFV segments are decoupled enough to have distinct demographic histories and to evolve under different molecular rates.

Author summary

Rift Valley Fever virus (RVFV) is an emerging pathogen that often infects ruminants through Africa and occasionally humans with mortality rates that could reach 30% of the patients. The potential to spread for other places beyond Africa was demonstrated by the recent outbreaks in Arabian Peninsula, where the virus is now endemic. The tri-segmented genome of RVFV is known to allow recombination through reassortment, *i.e.*, by the exchange of genomic segments. Using all data currently available for RVFV and using a reassortment analysis framework developed for influenza virus, we identified RVFV reassortants. Reassortment was shown to allow for 'shifts' in the capabilities of several other important pathogens, but our report raises concerns also because it questions previous phylogenetic studies on RVFV which, not only disregarded systematic errors due to the mixing of recombinant genes during phylogenetic inference, but also did not account for different mutation rates and ancestral relations among segments. These reassortment events could have broader

implications, since we show that despite similar purifying selection regimens the segments evolve under different mutation rates and show distinct demographic histories. Reassortment has been poorly described for RVFV but is known to contribute for increased pathogenicity for other Bunyavirus.

Introduction

Rift Valley Fever (RVF) is an emergent arthropod-borne illness that primarily affects ruminants in eastern and sub-Saharan Africa [1,2], causing severe socioeconomic impact in both animals and humans [3]. The causative agent of this disease is a Phlebovirus of the Bunyaviridae family, the RVF virus (RVFV). Since RVFV may be successfully vectored and widely spread by more than 30 distinct mosquito species [4,5], it has become a major public health concern not only in Africa [6], as demonstrated by the recent outbreaks in Saudi Arabia and Yemen [7]. The virus spread can be scattered and erratic, but fast and efficient [8] and can cause high mortality in humans and animals, with rates around 30%, as was the case during the Kenyan outbreak in 2007 [9].

Like all Bunyaviruses, RVFV has a three-segmented RNA genome organized in three segments: L (large), M (medium) and S (small). The L and M segments encode in the complementary sense, while the S segment has an ambisense arrangement. The L segment encodes the RNA-dependent RNA polymerase [10], the M segment codes for the envelope glycoproteins G1 and G2 and two non-structural proteins with 14 and 78 kDa, [11]. The S segment encodes the nucleocapsid protein (N) on the complementary sense, and a non-structural protein (NSs) in the other [12] that plays a major role in innate immunity and interacts with interferon signaling pathways [13].

This segmented genome permits natural reassortment [14–16], an event that requires that distinct viral strains co-infect a single cell at the same time, and may increase viral genetic diversity [17]. RVFV is able to infect an extensive plethora of vertebrate hosts (lambs, goats, cattle, rodents and humans) and arthropod-vectors [6], which may help to maintain the virus at a given place in space and facilitates co-infection in time. This begs the question of the importance of reassortment during the recent dissemination of the virus. A question that can be investigated, given that RVFV genome organization permits unraveling distinct evolutionary histories among segments reconstructed using the coalescent theory [18]. Although already observed in nature, estimates of

reassortment in RVFV are made difficult by the scarcity of complete genomic sequences from several localities, such as West Africa. Moreover, analyses done on concatenated or combined datasets, using more than one segment in the same genealogy, did not account for reassortment as a potential source of systematic error during phylogenetic inference [19]. We studied reassortment in RVFV using a framework developed initially for influenza virus [20]. We also investigated whether dynamics of viral genomic segments could have incongruous (*i.e.*, unlinked) evolutionary histories due to reassortment [21].

Results and Discussion

Adjacency patterns in phylogenies show reassortment events. We analyzed the posterior set of trees (PST) generated during independent phylodynamic Monte Carlo Markov Chain (MCMC) runs for all segments and identified two RVFV lineages with discordant phylogenetic histories for M and for L and S segments. Our results confirmed previous findings that the Kenyan bovine isolate 2007000608 was formed via reassortment of M segments [22] (Table 1 and green taxon labels in Figure 1). Furthermore, we observed a new reassortment event in the Tanzanian lineage TANTan00107 (Table 1 and red taxon labels in Figure 1), sampled from a human source. Since TANTan00107 was used in concatenated datasets [19], this finding demonstrates the importance of studying segments separately, to prevent systematic errors during phylogenetic inferences. Moreover, we investigated the occurrence of reassortment between pairs of segments (L and M, L and S, M and S), using the largest datasets (Figure 1 and Table 1). We found no additional reassortants between the L and the other two segments. However, we were able to detect five additional new reassortants between M and S segments (Figure 1B and Table 1). Three of these reassortant lineages (76370, SA75 and 3574), isolated in Zimbabwe and South Africa in human, cattle and sheep (Table 1), form a highly supported reassortant cluster that reappears in independent phylogenies for both the M and S segments (orange, purple and blue taxon labels in Figure 1B). The other two reassortants were from Mauritania (211HMMRRO1987 and 11ANMMRHG1998, pink and cyan taxon labels in Figure 1B), and were isolated from human and sheep, respectively (Table 1). The parental relationships can be accessed in Figure S1.

The genetic diversity inferred by patristic distances indicates distinct time scales for each segment. Since a segmented genome allows separate evolutionary histories per

chromosome [18,21], we investigated the amount of divergence for each segment by estimating patristic distances across viral gene genealogies, which may vary more than genetic distances and thus better capture discordances [23]. Pairwise distance distributions for the three datasets (for segments L, M and S with 120 sequences each) deviated from normality and were significantly different ($\chi^2 = 136.21$ d.f. = 2 and p -value $< 2.2E-16$ in Kruskal-Wallis test) after an exhaustive comparison among segments. Crucially, we obtained the ‘difference values’ of 689.73, 513.84 and 1203.57 for comparisons between L and M, L and S, and M and S, respectively. All values were greater than the critical value of 247.76, for the significance level (α) = 0.05. These comparisons revealed that the segments have different amount of genetic changes, indicating distinct time-scales for the segments.

Similar selection regimens for each segment support reassortment as the main driver of individual segment evolution. Differences in selection regimen among segments could explain the observed differences among patristic distances. Therefore, we investigated selection acting at the codon level on each segment. We only found sites with statistical significance under purifying selection for all analyzed genes ($w < 0$ and p -value < 0.05 , in Table S1). We compared with Fisher’s exact test the amount of negatively selected codons with statistical significance and found 98 codons out of 2066, 58 out of 1197, and 20 out of 510 for L, M and S respectively. However, there was no significant difference among the proportions of codons under selection for each segment (p -value = 0.74). Moreover, we compared the intensity of selection on each segment with w , using the Kruskal-Wallis test and found no significant differences among w for each segment ($\chi^2 = 3.92$, d.f. = 2 and p -value = 0.14). Therefore, our results did not support apparent differences in the selection regime acting on codons of each segment. Additionally, we estimated the Tajima’s D for each segment and found that they were significantly negative and similar for all segments (L: D= -3.19, p = 0.0013; M: D= -3.18, p = 0.0014; S: D= -3.32, p =0.0009). These similar negative Tajima’s D values did not change after excluding reassortant taxa out of the analysis, and agree with our findings on purifying selection. Negative Tajima’s Ds also point to possible past bottlenecks, also corroborated by the drop in effective population size (N_e) after 1970 for all segments (Figure 2C). Nevertheless, since the selective regimen is similar among segments and, given the significant differences in patristic distance distributions, these results strongly suggest that RVFV segments are evolving

under different time-scales, which constitutes an independent source of evidence for reassortment.

RVFV segments have different evolutionary dynamics that relates to reassortment. In principle reassortment events could decouple evolutionary dynamics for different genomic segments [18,21]. Therefore, we estimated the evolutionary rates (m) for each segment independently and compared the likelihoods using the Akaike's information criterion through Markov chain Monte Carlo (AICM) and path-sampling tests to choose the m that better fit our data (Table S2). The best m for the L segment was 2.31E-4 substitutions per site per year (s/s/y) ranging from 1.88E-4 to 2.72E-4 (within 95% highest posterior densities, HPD) (Figure 2A). For the M segment, the mean m of 3.80E-4 s/s/y (HPD = 2.53E-4 to 5.05E-4 s/s/y) had a better fit for the data. Finally, the S segment had a m of 3.36E-4 s/s/y (HPD = 2.42E-4 to 4.26E-4 s/s/y). In sum, the mean rate for M segment was at least 1.65 times higher than m for L and 1.45 times that for S. These differences implied an older root estimated for the L segment (Figure 2B) and overlapping roots for S and M. Importantly, the mean effective population size (N_e) of each segment experienced detectable periods of anti-correlation (Figure 2D), which had significant negative Kendall coefficients ($t < 0$ and $p\text{-value} \leq 0.05$ in the Figure 2D) in two time intervals: (i) between 1930 to 1950 and (ii) 1990 and 2010. Interestingly, these two periods of anti-correlation superimpose the estimated ages for the coalescence of reassortant lineages (Figure S2 and vertical intervals in Figure 2C). Although the nature of the association between the anti-correlation in the BSP with the estimated times for the coalescence nodes of reassortant lineages remains unclear, it could also suggest decoupling of the evolutionary dynamics among RVFV segments. Nevertheless, our findings support the notion that reassortment events in both the S and M segments took place during the recent evolutionary past of RVFV (Figures 1B and S1B). This could imply some form of linkage between these 2 genes, which is consistent with the fact that S and M interact during viral morphogenesis [24]. Remarkably, we found reassortant lineages amongst human samples (Table 1), a concerning result given that reassortment events in Bunyaviruses are known to have originated pathogenic new lineages from milder parental ones [25].

Material and Methods

Sequence datasets. We used our previously generated sequence [8] and all other RVFV sequences available in GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>). Only

sequences with available date of isolation were added in our dataset (Table S3). Sequences were aligned using the multiple alignment program Clustal Omega [26] and curated with SeaView v4.4.0 [27]. To prevent potential biases during phylogenetic inference due to recombination, we first analyzed all sequences with RDP4 program that incorporates RDP, GENECONV, Chimaera, Maxchi, Bootscan, SiScan and 3Seq methods [28]. Only recombination events with p -values ≤ 0.05 that were detected by three or more methods were considered employing the Bonferroni correction to avoid false positive results. We then selected the RVFV lineages with available sequences for the three segments (L, M and S), which resulted in three datasets with 120 sequences each. We also selected the lineages with pairs of segments available in all combinations (L/M, L/S, and M/S), resulting in datasets with 127, 128 and 162 sequences, respectively.

Reassortment detection analyses. To generate a posterior set of trees (PST) we used a Markov Chain Monte Carlo (MCMC) approach with MrBayes v3.2.1 [29], employing a general time reversible substitution model (GTR) [30] with a gamma distributed rate variation (G) [31] and a proportion of invariable sites (I). MCMC stationarity of two independent runs of 4 chains each (3 heated and 1 cold) was obtained after 20 millions generations, and trees were sampled every 2000 generations. Reassortment events were inferred using GiRaF v1.02 program, which compares a distribution of phylogenetic trees of different segments searching for groups of discordant splits of the trees using a biclique algorithm [20]. We used the PST for GiRaF searches, discarding the first 500 trees, culling splits with frequencies less than 5%, and set a confidence threshold of 70% to report a reassortment event.

Patristic distances analyses. Firstly, 100 maximum likelihood (ML) phylogenetic trees were independently inferred using GARLI v2.0 [32], that uses a stochastic algorithm to estimate simultaneously the best topology, branch lengths and substitution model parameters that maximize the log Likelihood (lnL). Assuming that simpler models would be satisfactorily accounted for when estimating transition probabilities with the GTR+G+I model, we selected the tree with best lnL estimated with it. The patristic distances were calculated from the best tree for each segment with the R-package adephylo 1.16 [33]. The differences among patristic distances were accessed with Kruskal-Wallis rank sum test from R-program (<http://www.r-project.org/>). The multiple comparisons among patristic distances of each segment was done with the multiple comparison test after Kruskal-Wallis from the R-package pgirmess v1.5.9 [34].

Selection analyses. To evaluate selection patterns on the coding sequences for each segment: polymerase in the L segment, glycoprotein in M, nonstructural and nucleocapsid in S; we estimated the difference ($\omega = dN-dS$) between the non-synonymous (dN) and synonymous (dS) rates per codon sites, using the single likelihood ancestor counting (SLAC) algorithm with HyPhy v2.11[35], assuming $\alpha = 0.05$. In this way, ω greater than zero suggests directional selection, while values below zero indicate purifying selection. We used the Fisher's exact test to investigate the differences in the total number of sites under selection among the segments and Kruskal-Wallis rank sum test to access the differences among ω for each segment. Both tests were done with the R program. Moreover, to evaluate neutral patterns of evolution, Tajima's D were estimated for L, M and S datasets; using the R-package pegas v0.5-1 [36].

Phylogenetics. Since we had dates of isolation for each RVFV sequence, we estimated coarse substitution rates per site per year (μ) using Path-O-Gen v1.4 (<http://tree.bio.ed.ac.uk/software/pathogen/>) with the best Maximum Likelihood (ML) trees for each segment, after 100 independent replications of Garli v2.0.1019 [32] using GTR+G+I. In addition, Maximum Clade Credibility (MCC) trees were inferred using a MCMC approach under GTR+G+I and a relaxed (uncorrelated lognormal) molecular clock [37] using program BEAST v1.8.0 [38], with the previously estimated μ values as priors. Moreover, we used the variation in effective population size ($N_e.g$) to infer the viral demography with Bayesian Skyride Plots (BSP) [39]. MCMC convergence was obtained for four independent runs with 100 million steps, which were sufficient to obtain a convergence, as inspected by effective sample sizes (ESS) values above 200 for all parameters. Since we obtained different rates for each segment, we re-ran BEAST again changing the mean rate among segments (Table S2). To test the model that best-fit the data we employed a Akaike's information criterion through Markov chain Monte Carlo (AICM) test [40], implemented in Tracer v1.6 (<http://tree.bio.ed.ac.uk/software/tracer/>). Also, the suitability of the m prior to the data was tested again with path-sampling (PS) and stepping-stone (SS) algorithms [40]. Because our Bayesian Skyride were evidently not stationary at the time span of the observations, we evaluated the correlation between pairs of $N_e.g$ for segments using Kendall t correlation coefficient, to test for the correlation in intervals of 20 years between 1930 and 2010. To access the significance of correlations, we adopted $\alpha = 0.05$.

References

1. Bird BH, Ksiazek TG, Nichol ST, Maclachlan NJ (2009) Rift Valley fever virus. *J Am Vet Med Assoc* 234: 883–893. doi:10.2460/javma.234.7.883.
2. Faye O, Ba H, Ba Y, Freire CCM, Faye O, et al. (2014) Reemergence of rift valley Fever, mauritania, 2010. *Emerg Infect Dis* 20: 300–303. doi:10.3201/eid2002.130996.
3. Rich KM, Wanyoike F (2010) An assessment of the regional and national socio-economic impacts of the 2007 Rift Valley fever outbreak in Kenya. *Am J Trop Med Hyg* 83: 52–57. doi:10.4269/ajtmh.2010.09-0291.
4. Weaver SC, Reisen WK (2010) Present and future arboviral threats. *Antiviral Res* 85: 328– 345. doi:10.1016/j.antiviral.2009.10.008.
5. Diallo M, Nabeth P, Ba K, Sall a a, Ba Y, et al. (2005) Mosquito vectors of the 1998-1999 outbreak of Rift Valley Fever and other arboviruses (Bagaza, Sanar, Wesselsbron and West Nile) in Mauritania and Senegal. *Med Vet Entomol* 19: 119–126. doi:10.1111/j.0269283X.2005.00564.x.
6. Pepin M, Bouloy M, Bird BH, Kemp A, Paweska J (2010) Rift Valley fever virus(Bunyaviridae: Phlebovirus): an update on pathogenesis, molecular epidemiology, vectors, diagnostics and prevention. *Vet Res* 41: 61.
7. World Health Organization (2000) WHO | Saudi Arabia and Yemen: First cases of Rift Valley fever reported outside Africa, 2000. Available: <http://www.who.int/csr/outbreaknetwork/saudiarabia/en/>. Accessed 11 August 2011.
8. Soumaré POL, Freire CCM, Faye O, Diallo M, de Oliveira JVC, et al. (2012) Phylogeography of Rift Valley Fever virus in Africa reveals multiple introductions in Senegal and Mauritania. *PLoS One* 7: e35216. doi:10.1371/journal.pone.0035216.
9. CDC (2007) Rift Valley fever outbreak--Kenya, November 2006-January 2007. *MMWR Morb Mortal Wkly Rep* 56: 73–76.
10. Müller R, Poch O, Delarue M, Bishop DH, Bouloy M (1994) Rift Valley fever virus L segment: correction of the sequence and possible functional role of newly identified regions conserved in RNA-dependent polymerases. *J Gen Virol* 75 (Pt 6): 1345–1352.
11. Collett MS, Purchio AF, Keegan K, Frazier S, Hays W, et al. (1985) Complete nucleotide sequence of the M RNA segment of Rift Valley fever virus. *Virology* 144: 228–245.

12. Giorgi C, Accardi L, Nicoletti L, Gro MC, Takehara K, et al. (1991) Sequences and coding strategies of the S RNAs of Toscana and Rift Valley fever viruses compared to those of Punta Toro, Sicilian Sandfly fever, and Uukuniemi viruses. *Virology* 180: 738–753.
13. Perrone LA, Narayanan K, Worthy M, Peters CJ (2007) The S segment of Punta Toro virus (Bunyaviridae, Phlebovirus) is a major determinant of lethality in the Syrian hamster and codes for a type I interferon antagonist. *J Virol* 81: 884–892. doi:10.1128/JVI.01074-06.
14. Sall AA, Zannotto PM, Sene OK, Zeller HG, Digoutte JP, et al. (1999) Genetic reassortment of Rift Valley fever virus in nature. *J Virol* 73: 8196–8200.
15. Grobbelaar A a, Weyer J, Leman P a, Kemp A, Paweska JT, et al. (2011) Molecular epidemiology of Rift Valley fever virus. *Emerg Infect Dis* 17: 2270–2276. doi:10.3201/eid1712.111035.
16. Bird BH, Khristova ML, Rollin PE, Ksiazek TG, Nichol ST (2007) Complete genome analysis of 33 ecologically and biologically diverse Rift Valley fever virus strains reveals widespread virus movement and low genetic diversity due to recent common ancestry. *J Virol* 81: 2805–2816. doi:10.1128/JVI.02095-06.
17. Simon-Lorieri E, Holmes EC (2011) Why do RNA viruses recombine? *Nat Rev Microbiol* 9: 617–626. doi:10.1038/nrmicro2614.
18. Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, et al. (2008) The genomic and epidemiological dynamics of human influenza A virus. *Nature* 453: 615–619. doi:10.1038/nature06945.
19. Nderitu L, Lee JS, Omolo J, Omulo S, O’Guinn ML, et al. (2011) Sequential Rift Valley fever outbreaks in eastern Africa caused by multiple lineages of the virus. *J Infect Dis* 203: 655–665. doi:10.1093/infdis/jiq004.
20. Nagarajan N, Kingsford C (2011) GiRaF: robust, computational identification of influenza reassortments via graph mining. *Nucleic Acids Res* 39: e34. doi:10.1093/nar/gkq1232.
21. Carpi G, Holmes EC, Kitchen A (2010) The evolutionary dynamics of bluetongue virus. *J Mol Evol* 70: 583–592. doi:10.1007/s00239-010-9354-y.
22. Bird BH, Githinji JWK, Macharia JM, Kasiiti JL, Muriithi RM, et al. (2008) Multiple virus lineages sharing recent common ancestry were associated with a Large Rift Valley fever outbreak among livestock in Kenya during 2006-2007. *J Virol* 82: 11152–11166. doi:10.1128/JVI.01519-08.
23. Fourment M, Gibbs MJ (2006) PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. *BMC Evol Biol* 6: 1. doi:10.1186/1471-2148-6-1.

24. Terasaki K, Murakami S, Lokugamage KG, Makino S (2011) Mechanism of tripartite RNA genome packaging in Rift Valley fever virus. *Proc Natl Acad Sci U S A* 108: 804–809. doi:10.1073/pnas.1013155108.
25. Bowen MD, Trappier SG, Sanchez a J, Meyer RF, Goldsmith CS, et al. (2001) A reassortant bunyavirus isolated from acute hemorrhagic fever cases in Kenya and Somalia. *Virology* 291: 185–190. doi:10.1006/viro.2001.1201.
26. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7: 539. doi:10.1038/msb.2011.75.
27. Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 27: 221–224. doi:10.1093/molbev/msp259.
28. Martin DP, Lemey P, Lott M, Moulton V, Posada D, et al. (2010) RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26: 2462–2463. doi:10.1093/bioinformatics/btq467.
29. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, et al. (2012) MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst Biol*. doi:10.1093/sysbio/sys029.
30. Tavaré S (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect Math Life Sci* 17.
31. Yang Z (1995) A space-time process model for the evolution of DNA sequences. *Genetics* 139: 993–1005.
32. Zwickl D (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. The University of Texas at Austin.
33. Jombart T, Balloux F, Dray S (2010) adephylo: new tools for investigating the phylogenetic signal in biological traits. *Bioinformatics* 26: 1907–1909. doi:10.1093/bioinformatics/btq292.
34. Giraudoux P (2013) pgirmess: Data analysis in ecology.
35. Pond SLK, Frost SDW, Muse S V (2005) HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21: 676–679. doi:10.1093/bioinformatics/bti079.
36. Paradis E (2010) pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 26: 419–420. doi:10.1093/bioinformatics/btp696.
37. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4: e88. doi:10.1371/journal.pbio.0040088.

38. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol*. doi:10.1093/molbev/mss075.
39. Minin VN, Bloomquist EW, Suchard MA (2008) Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol* 25: 1459–1471. doi:10.1093/molbev/msn090.
40. Baele G, Lemey P, Bedford T, Rambaut A, Suchard M a, et al. (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol* 29: 2157–2167. doi:10.1093/molbev/mss084.
41. Kortekaas J, Oreshkova N, Cobos-Jiménez V, Vloet RPM, Potgieter C a, et al. (2011) Creation of a nonspreading Rift Valley fever virus. *J Virol* 85: 12622–12630. doi:10.1128/JVI.00841-11.

Figure legends

Figure 1. Reassortments revealed by the incongruities among phylogenies for RVFV segments. A- Reassortments between L and M segments. B- Reassortments between M and S segments. Pairs of reassortant taxa are shown in colors specific to the reassortant pair. The order of taxa is the same between each pair of trees. Parental strains could be observed in Figure S1. Only reassortant taxon names are shown for clarity. Reconciled, parallel trees, showing reassortants, including all taxon names are shown in Figure S1.

Figure 2. Phylodynamic estimates for the best-fit evolutionary rate for each segment. A- Distribution of evolutionary rates (m) estimated during MCMC convergence. B- Distribution of estimated the root (TMRCA) for each segment. C- BSP showing the effective population size ($N_{e.g}$) estimated for L, M and S segments. The vertical colored intervals indicate the estimates of the node age of detected reassortant lineages (red for 2007000608 and TANTan00107; gray for SA75, 3574 and 76370). The node ages are shown in Figure S2. D - Correlation between pairs of $N_{e.g}$ for 20 years intervals. Kendall correlation coefficients (τ) < 0 indicated anti-correlation in the intervals and p -values < 0.05 point to statistical significance of the correlations. All estimates are based on datasets with 120 sequences.

Supporting information legends

Figure S1. Paired phylogenetic trees for RVFV segments showing phylogenetic discrepancies among reassortant lineages. A – Phylogenetic trees for L and M. B – Phylogenetic trees for M and S. Posterior probability support values are shown close to the nodes. Pairs of reassortants are colored accordingly.

Figure S2. Time-scaled MCC trees, showing times of reassortment. A – MCC tree for L segment. B - MCC tree for M segment. C- MCC tree for S segment. Labels of reassortants are colored in red. TMRCA's are shown close to the nodes.

Table S1. Significant detected codon sites under purifying selection in RVFV.

Table S2. Akaike's information estimated by Markov chain Monte Carlo (AICM) and pathsampling tests for different prior propositions to explain evolution of RVFV strains.

Table S3. RVFV strains used in this study.

Tables

Table 1. RVFV reassortant lineages.

Strain	Country	Year	Host	L accession number	M accession number	S accession number	Re
2007000608	Kenya	2007	Cattle	EU574023	EU574049	EU574079	
211HMMRRO1987	Mauritania	1987	Human	Unavailable	JN995301	JN995253	
11ANMMRHG1998	Mauritania	1998	Sheep	Unavailable	JN995312	JN995264	
SA75	South Africa	1975	Human	DQ375428	DQ380189	DQ380175	
3574	South Africa	1974	Sheep	JF784386	JF784387	JF784388	
TANTan00107	Tanzania	2007	Human	HM586959	HM586970	HM586981	
76370	Zimbabwe	1970	Cattle	DQ375426	DQ380188	DQ380165	

[re to download high resolution image](#)

[re to download high resolution image](#)