NICHOLAS DI PAOLA

DESCOBERTA E CARACTERIZAÇÃO DE VÍRUS EMERGENTES E REERGENTES EM ÁREAS PERI-FLORESTAIS

Tese apresentado ao Departamento de Microbiologia do Instituto de Ciências Biomédicas da Universidadede São Paulo, para obtenção do Títulode Doutor em Ciências.

São Paulo
2018

NICHOLAS DI PAOLA

DISCOVERING AND CHARACTERIZING EMERGING AND RE-EMERGING VIRUSES IN COMMUNITIES ENCROACHING TROPICAL "HOTSPOTS"

Thesis presented to the Department of Microbiology of the Institute of Biomedical Sciences of the University of São Paulo, to obtain the degree of Doctor in Science.

São Paulo
2018

Nicholas Di Paola

Discovering and Characterizing Emerging and Re-Emerging Viruses in Communities Encroaching Tropical "Hotspots"

Thesis presented to the Department of Microbiology of the Institute of Biomedical Sciences of the University of São Paulo, to obtain the degree of Doctor in Science.

Concentration area: Microbiology

Advisor: Prof. Dr. Paolo Marinho de Andrade Zanotto

Original Version

São Paulo
2018

Nicholas Di Paola

Descoberta e Caracterização de Vírus Emergentes e Reergentes em Áreas Peri-florestais

Tese apresentado ao Departamento de Microbiologia do Instituto de Ciências Biomédicas da Universidade de São Paulo, para obtenção do Título de Doutor em Ciências.

Área de concentração: Microbiologia

Orientador: Prof. Dr. Paolo Marinho de Andrade Zanotto

Versão original

São Paulo
2018

CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
Serviço de Biblioteca e informação Biomédica
do Instituto de Ciências Biomédicas da Universidade de São Paulo

Ficha Catalográfica elaborada pelo(a) autor(a)

Paola, Nicholas Di
/ Nicholas Di Paola; orientadora Paolo Marinho de Andrade Zanotto. -- São Paulo, 2018.
132 p.

Tese (Doutorado)) -- Universidade de São Paulo, Instituto de Ciências Biomédicas.

1. Vírus emergente. 2. Vírus da febre amarela. 3. Parvovírus B19. 4. Uso do codão. 5. Evolução viral. I. Zanotto, Paolo Marinho de Andrade, orientador. II. Título.

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS

Candidato(a): Nicholas Di Paola

Título da Tese: **Descoberta e Caracterização de Vírus Emergentes e Reergentes em Áreas Peri-florestais**

Orientador(a): Prof. Dr. Paolo Marinho de Andrade Zanotto

A Comissão Julgadora dos trabalhos de Defesa da Tese de Doutorado, em sessão pública realizada a ...../...../....., considerou

( ) Aprovado(a) **( ) Reprovado(a)**

Examinador(a): Assinatura: .....
Nome: .....
Instituição: .....

Examinador(a): Assinatura: .....
Nome: .....
Instituição: .....

Examinador(a): Assinatura: .....
Nome: .....
Instituição: .....

Examinador(a): Assinatura: .....
Nome: .....
Instituição: .....

Presidente: Assinatura: .....
Nome: .....
Instituição: .....

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS

Cidade Universitária "Armando de Salles Oliveira"
Av. Prof. Lineu Prestes, 2415 – CEP. 05508-000 São Paulo, SP – Brasil
Telefone :(55) (11) 3091-7733 - telefax : (55) (11) 3091-8405
e-mail: cep@icb.usp.br

Comissão de Ética em Pesquisa

CERTIFICADO DE ISENÇÃO

Certificamos que o Protocolo CEP-ICB Nº **647/14** referente ao projeto intitulado: "*Discovering Novel Viruses in Communities Encroaching Tropical "Hotspots" via a Metagenomics Approach*" sob a responsabilidade de **Nicholas Di Paola,** foi analisado na presente data pela CEUA - COMISSÃO DE ÉTICA NO USO DE ANIMAIS e pela CEPSH- COMISSÃO DE ÉTICA EM PESQUISA COM SERES HUMANOS, tendo sido deliberado que o referido projeto não utilizará animais que estejam sob a égide da lei 11.794 de 8 de outubro de 2008, nem envolverá procedimentos regulados pela Resolução CONEP nº466 de 2012.

São Paulo, 26 de fevereiro de 2014.

PROF. DR. WOTHAN TAVARES DE LIMA
Coordenador da CEUA - ICB/USP

PROF. DR. PAOLO M.A ZANOTTO
Coordenador da CEPsh - ICB/USP

ICB USP

Cidade Universitária "Armando de Salles Oliveira", Butantã. São Paulo. SP - Av. Professor Lineu Prestes, 2415 - ICB III - 05508 000
Comissão de Ética em Pesquisa - Telefone (11) 3091-7733 - e-mail: cep@icb.usp.br

Decl. CEPSH.029.2017

DECLARAÇÃO

Em adendo ao Certificado de Isenção **CEP-ICB nº 647/2014**, de 26/02/2014, e por solicitação do Prof. Dr. *Paolo Marinho de Andrade Zanotto*, do departamento de Microbiologia, informo que o título do projeto do aluno *Nicholas Di Paola* foi alterado para ***"Discovering and Characterizing Emerging and Re-Emerging Viruses in Communities Encroaching Tropical 'Hotspots'."*** sem modificações de seu conteúdo, conforme sugerido pela banca de qualificação.

São Paulo, 09 de novembro de 2017.

Profa. Dra. **Camila Squarzoni Dale**
Coordenadora CEPSH-ICB/USP

Profa. Dra. **Luciane Valéria Sita**
Coordenadora CEUA-ICB/USP

To my grandfather Yura — your everlasting support pushed me through the toughest of times and kept me focused. You also showed me that, life could be beautiful…

PERSONAL ACKNOWLEDGMENTS

First, I would like to thank my mentor Professor Paolo Zanotto for his patience, guidance, wisdom, and support over the last 4+ years. We had many office meetings, many intense discussions (both academic and about life), and many car rides that have allowed me to grow as a scientist and as a person.

In addition, I would like to acknowledge the many colleagues that have shared the same lab room and guided me along the way: Atila I, Caio F, Carla B, Julian VB, Suzanna O, Marielton C, Shahab P, Rafaella B, Flavio M, Danielle O, Isabelle B, Anderson S and Renata Z. I would like to recognize my friends and colleagues abroad at the Institute Pasteur de Dakar for taking care of me and inspiring me to be better and work harder everyday: Moussa M, Henriette D, Gamou F, Oumar F, Oumar N, Moussa D, and especially Martin F and many others. I would also like to thank Kristina U from the University of Sterling whose intelligence and humor has led to many hours of scientific, philosophical, and just plain friendly discussions. I would also like to thank all of them for believing in me as a scientist, and in some select cases, as a soccer player.

I would like to acknowledge several professors who have supported me over the years and who were never shy to give me counsel about scientific and bureaucratic issues I came across: Prof. Durigon, Prof. Boccardo, Prof. Taborda, Prof. Menck, Prof. Sall, Prof. Souza, Dr. Passos, Prof. Romano and many others. I would also like to recognize Gisele Santana for her patience (with me) and all my bureaucratic frustrations.

My friends outside of the lab played a role in my happiness and perserverance throughout the tough times. I would like to thank all my friends at SPAC for making me feel welcome and forcing me to practice Portuguese. I would like to give a special thanks to Arthur, Janga, Stefano, Ton, Hugo, Flavia, Nadine, and several other friends who dealt with me at my best and worst, but were always ready to cheer me up and keep me going.

Lastly, I would like to thank my family and those who have given me love and support no matter where I was in the world: Yura, Galina, Mom, Tatiana, Marina, Kira, Isabelle, Elissa, Emily, Brandon and Monica. I would also like to thank Toothless for being the most amazing cat ever. They all made sure that I would always remember that love and happiness could take many forms and how it can be so simple.

FINANCIAL ACKNOWLEDGEMENTS

First and foremost, I would like to thank the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (personal project 2013/22136-1, and research group project 2014/17766-9). My research and advancement would not have been possible without their financial support. In particular, the approval of a 6-month internship at the Pasteur Institute of Dakar (Senegal) helped me grow as a researcher and as an individual.

I would like to express my gratitude to Dr. Amadou Sall and the Pasteur team for taking care of me and allowing me to participate in their amazing research.

I would like to especially thank the Pensabio/Illumina Grant Award team. Their technical and financial support was a key to the completion my PhD, along with the data generated that led to six academic papers.

Finally, appreciation also goes out to CAPES and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (process number 441105/2016-5) for their past and current support.

“Science knows no country, because knowledge belongs to humanity, and is the torch that illuminates the world.”

Louie Pasteur

RESUMO

Paola ND. Descoberta e Caracterização de Vírus Emergentes e Reergentes em Áreas Peri-florestais. [Tese (Doutorado em Microbiologia)]. São Paulo: Instituto de Ciências Biomédicas, Universidade de São Paulo; 2018.

A fragmentação e a invasão de florestas tropicais e a crescente concentração de assentamentos humanos aumentaram exponencialmente as chances de exposição a vírus emergentes e emergentes. Dado o grande potencial de espalhamento de patógenos em população humanas, a identificação e caracterização de agentes patogênicos circulantes podem melhorar a atenção primária e as capacidades de diagnóstico para um agente emergente futuro. As abordagens moleculares e metagenômicas que utilizam as tecnologias de sequenciação da próxima geração levaram a descoberta e caracterização de muitos vírus emergentes na última década. Além disso, as abordagens *in silico* também podem ajudar a identificar vírus emergentes usando apenas dados de sequenciamento publicamente disponíveis. Além disso, estimar a ascendência filogenética e até mesmo analisar as mudanças no uso de codons são ferramentas adicionais que podem melhorar a nossa compreensão de vírus emergentes ou reemergentes. Este projeto visou aplicar essas ferramentas em ambos os vírus que poderiam estar circulando no Brasil: Parvovírus B19 e vírus da Febre Amarela. Também exploramos as aplicações de modelos ocultos de Markov e índice de adaptação de codons usando dados publicamente disponíveis. Esperamos que este trabalho forneça uma prova de conceito para futuros projetos metagenômicos e demonstre a utilidade das várias técnicas moleculares e bioinformáticas no estudo de vírus emergentes.

Palavras-chave: Vírus emergente. Vírus da febre amarela. Parvovírus B19. Uso do codão. Evolução viral. Filogenia. Modelos Ocultos de Markov.

ABSTRACT

Paolo ND. Discovering and Characterizing Emerging and Re-Emerging Viruses in Communities Encroaching Tropical “Hotspots”. [Ph. D. thesis (Microbiology)]. São Paulo: Instituto de Ciências Biomédicas, Universidade de São Paulo; 2018.

Fragmentation and encroachment of tropical rainforests and the growing concentration of human settlements have exponentially increased chances of exposure to re-emerging and emerging viruses. Given the large potential for pathogens to spillover and spread in a population, identifying and characterizing circulating human pathogens could improve the readiness and diagnostic capabilities for a future emergence. Molecular and metagenomic approaches using next-generation sequencing technologies have led to the discovery and characterization of many emerging viruses over the last decade. In complement, *in silico* approaches can also help identify emerging viruses using only publicly available sequencing data. Moreover, estimating the phylogenetic ancestry and even analyzing changes in codon usage are additional tools that can improve our understanding of an emerging or re-emerging virus. This project aimed to apply these tools to two viruses that could be circulating in Brazil: Parvovirus B19 and Yellow Fever virus. We also explored the applications of Hidden Markov models and codon adaptation index using publicly available data. We expect this work to provide a proof-of-concept for future metagenomic projects, and demonstrate the utility for several molecular and bioinformatics techniques in the study of emerging viruses.

Keywords: Emerging virus. Yellow Fever virus. Parvovirus B19. Codon usage. Viral Evolution. Phylogeny. Hidden Markov Models.

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF ABBREVIATIONS & SYMBOLS

B19V – Parvovirus B19
BEAST – Bayesian analysis of molecular sequences software
BLAST – Basic Local Alignment Search Tool (GenBank)
bp – base pairs
CAI – Codon adaptation index
CDC – Center for Disease Control
CDS – coding sequence
CHIKV – Chikungunya virus
DENV – Dengue virus
DNA – Deoxyribonucleic acid
EBOVZ – Zaire Ebola virus
ELISA - enzyme-linked immunosorbent assay
HCV – Hepatitis C virus
HIV – Human immunodeficiency virus
IFA – Immunofluorescent assay
IgG – Immunoglobulin G
IgM – Immunoglobulin M
LOESS - Locally weighted scatterplot smoothing
MCMC – Markov chain Monte Carlo
MERS-CoV – Middle East respiratory syndrome coronavirus
MOI – multiplicity of infection
Ne – effective population size
NGS – next generation sequencing
NS1 – non-structural protein 1
nt – nucleotides
ORF – open reading frame
PCR – Polymerase chain reaction
PFU – plaque-forming unit
qRT-PCR – quantitative Reverse-Transcription polymerase chain reaction
RdRp – RNA-dependent RNA polymerase
RNA – Ribonucleic acid
SARS – Severe acute respiratory syndrome
tMRCA - the most recent common ancestor
UTR – Untranslated region
VP1 – viral protein 1
Wm – Malthusian fitness
WNV – West Nile virus
YFV – Yellow Fever virus
ZIKV – Zika virus

TABLE OF CONTENTS

CHAPTER 1 – RESEARCH

1.1 Where there is no overlap, there is a gap

This text was published on December 9th, 2014 in the *RSBMT* (1)

Over the past 60 years, Brazil has seen extensive rural-urban migration, leading to human settlements encroaching on tropical ecosystems. Persistent human interaction with vectors from diverse tropical environments has led to the emergence of zoonotic pathogens varying in pathogenicity and infectivity. Over time, zoonotic viruses adapt to humans, eventually leading to secondary transmission in some cases. However, most zoonotic pathogens result in asymptomatic and/or dead-end cases, rarely leading to sustained human-human transmission. In very rare cases, a single transmission can explode into an outbreak with sustained networks of human-human transmission. This is the current situation in West African nations, initiated by a single emergence of the Zaire Ebola virus (EBOVZ) in Guéckédou in Guinea's eastern rainforest region (2). Although EBOVZ is a known virus, this was the first record of a case in this region, surprising unprepared health authorities with no experience of encountering Ebola viruses. Owing to ideal conditions, the 25th Ebola virus outbreak quickly spread to the neighboring countries, Sierra Leone and Liberia.

As of December 2, 2014, the CDC and WHO had conjunctly reported 16,933 cases and 6,002 deaths. Critically, the virus itself is not necessarily the main factor explaining the current high EBOVZ case count, as it is behaving similarly to the ones isolated during the first fully recorded outbreak of EBOVZ in 1976. This time, the virus is not experiencing restrictions with respect to the availability of susceptible hosts, as it shows high transmission in densely populated metropolitan settings. Accordingly, the basic reproductive rate (*R*) has been estimated to be around 2.56 in areas with the highest transmission rates, which is comparable to fomite-transmitted viruses such as small pox and influenza (3). Indeed, as we cannot predict the success of a pathogen or the time or location of emergence, we should focus on what we can control, *i.e.*, setting up efficient health systems to report and isolate cases and contacts and maximizing

surveillance efforts and resources to prevent future outbreaks. The arrival of all dengue virus serotypes (DENV-1, -2, -3, and -4), West Nile virus, and more recently, the Chikungunya virus (CHIKV) in the Americas is a reminder that even with a strong healthcare foundation, a novel emerging or re-emerging virus can cause morbidity and mortality in major metropolitan settings if it is not identified and reported quickly.

A recent review in *Revista da Sociedade Brasileira de Medicina Tropical*/Journal of the Brazilian Society of Tropical Medicine discussed the introduction of the Mayaro virus (MAYV) and more recently, CHIKV in Brazil (4). Both have the potential to cause increased morbidity and mortality in Brazil because of arbovirus infections, and therefore, require our attention.

Originally isolated in the 1950s, CHIKV has migrated through the Caribbean and Americas since then, with imported cases reported in Brazil since 2010. Autochthonous transmission was first recorded in September 2014 in the state of Amapá. CHIKV has already experienced sustained transmission in the states of Bahia and Minas Gerais and may have spread widely already. Both *Aedes aegypti* and *Aedes albopictus* are competent mosquito vectors of CHIKV in an urban cycle, and they are widespread across Brazil. Moreover, the three genotypes of CHIKV can be transmitted efficiently by all 35 strains of *Aedes aegypti* and *Aedes albopictus*, justifying the suggestion that outbreaks in other states are inevitable (4).

MAYV is another pathogen responsible for sporadic cases in Brazil, mainly in the Amazonian region. With MAYV residing in primate reservoirs, *Haemagogus* and *Aedes aegypti* mosquitos initiate transmission via a pseudo-sylvatic cycle, similar to yellow fever (5). Although there have been no large outbreaks, health authorities should preemptively prepare for a re-emergence in unexpected localities.

Symptoms of CHIKV and MAYV are similar to those of other common infectious agents, such as the DENV, and less common agents, such as Rocio, Oropouche, Saint Louis encephalitis, and Yellow Fever viruses (6). Importantly, the availability of rapid and efficient assays precisely identifying viral agents is crucial to diagnose patients and analyze epidemiological data. In addition, for medical personnel to follow proper treatment protocols, diagnostics should be readily available, accurate, sensitive, and economical to differentiate between these agents without a “coincidental” assumption

(*i.e.*, during a dengue outbreak, patients with similar symptoms may be infected by other arboviruses but not subjected to formal serological assay). Interestingly, this mirrors initial diagnostic scenarios in Guinea. According to the CDC, in the early stages of EBOVZ infection, patients could present with fever, severe headache, fatigue, muscle pain, vomiting, diarrhea, abdominal pain, or unexplained hemorrhaging. In West Africa, this symptom collection could indicate a number of differential diagnoses including malaria, typhoid fever, leptospirosis, rickettsiosis, African trypanosomiasis, Lassa fever, cholera, Marburg virus, and Crimean-Congo hemorrhagic fever (7). Therefore, in the early stages of an outbreak, lack of positive confirmatory serological tests could delay health authorities' responses, depending on the disease severity. Logically, health structure, resources, and sociological qualities differ between West African nations and the Americas. However, we must continuously remind ourselves that a similar scenario can occur in Brazil. Are Brazil's diagnostic and response capabilities prepared for the emergence of a novel pathogen?

It has previously been argued that many arbovirus infections have been neglected in Brazil, with few epidemiological studies conducted to understand their distribution. For instance, dengue fever is difficult to distinguish from other acute febrile illnesses, which emphasizes the importance of a differential diagnosis (6,8). Indeed, most diagnoses in the country are made on clinical and epidemiological grounds. In 2014 (as of November 2), 547,612 clinical cases of dengue fever were reported, but 8,423 (1.54%) were laboratory confirmed (9,10). These numbers do not represent the countless mildly symptomatic or nonspecific febrile cases. Therefore, the true rates of infection and transmission are underestimated, and the possibility of infection from other pathogens is ignored. With respect to the CDC Ebola case definition, DENV, CHIKV, and MAYV could cause overlapping symptoms with the above diagnoses, supporting the need for broad-spectrum, sensitive, and economic diagnostic assays to identify infectious agents without concerns regarding cross-reaction of closely related agents. This does not imply that an Ebola outbreak is a given in Brazil but exemplifies a complication that could arise in such a scenario.

The emergence of a novel pathogen should always be preconsidered. Moreover, if it has the potential for establishing itself within a human transmission network, having

diagnostic tools to confirm its presence and prompt notification of national health authorities should be priorities. Depending on the pathogen's transmissibility, contact tracing and quarantine should be established rapidly. Although EBOV is not novel in the sense of being unknown, it was novel in the locality of its zoonotic emergence in West Africa. Once a country undergoes a zoonotic-related emergence event leading to morbidity and mortality, the aftermath leads to preemptive measures to ensure that the scale of the next outbreak is significantly diminished. Continuous sentinel and active surveillance are costly and time consuming from a resource perspective. However, considering resource limitations, early diagnostic efforts and surveillance of emerging pathogens are the best ways to reduce mortality, morbidity, and the enormous health care costs associated with epidemics (11).

Brazil has thousands of communities surrounded by tropical ecosystems teeming with potentially pathogenic agents. Are there countries with limited resources whose healthcare system has successful prevented large outbreaks efficiently and strategically, which who we can learn from? Uganda could be a good model; similar in respect to climate and ecological diversity, also limited in resources disposed towards surveillance and epidemic management. Uganda presented a model system of emerging infectious disease control, demonstrating from previous EBOV outbreaks that continuous surveillance, that dedicated teams, devoted to contact tracing, case isolation, and cooperating with local communities and bordering countries, successfully limits case counts and spread (12). Brazil and other nations could learn from Uganda's experience, applying the approach to CHIKV, MAYV, DENV, and other pathogens. Although our ability to diagnose, treat, trace, and contain pathogen outbreaks has improved over the last decade, it has not been sufficient to prevent large outbreaks of new serotypes of dengue, and there is no indication that CHIKV will be halted. Therefore, one has to conceive that what is ongoing in West Africa could happen in Brazil. *Where there is no overlap, there is a gap;* therefore, we should establish much-needed constructive interactions between governments, academia, and the private sector to continue searching for novel methods of diagnostics, improve surveillance, strategize, and maximize resources to prepare for any future emergence. The question is not one of whether an emergence will occur but when and whether we will be ready.

1.2 Objectives

The overall research goal is to identify and characterize emerging and re-emerging viruses through molecular and bioinformatics (*in silico*) approaches in São Paulo, Brazil. The specific objectives are:

(i) To identify highly divergent viruses in metagenomic sequencing data using a Hidden Markov Modeling approach

(ii) Monitoring changes in a virus' codon usage over time in relation to a host's codon usage using codon adaptation index

(iii) Using next generation sequencing to identify the etiological agents in clinically suspected dengue fever patient's that were negative for the Dengue virus and antibody

(iv) To sequence and characterize Yellow Fever virus isolated from a convalescent patient's urine and semen

1.3 Materials and Methods

1.3.1 Profile HMM construction and HMMER3

Complete amino acid (aa) sequences from human pathogenic viral genera and families, 70 in total, were downloaded from GenBank. Sequences were then aligned and manually fragmented into smaller alignments ranging from 70-200 aa. Cuts were based on the level of conservation between sequences: highly conserved regions were made the central “foundation” of a fragmented alignment and then were extended to adjacent regions that were less conserved regions and highly varied between species. Next, alignments were built into profile hidden Markov models (pHMM) using *hmmbuild* from the HMMER3 package (13).

To use pHMMs against sequencing databases, DNA and RNA reads were first downloaded from NCBI’s SRA database. Next, reads were trimmed using Trimmomatic 2.0, converted to fasta format using *fastx*, and then translated to amino acids in all 6 reading frames using an in-house script. Using *hmmsearch*, pHMMs were quarried against the translated dataset. To prevent false-positive matches, we set the search e-value to 10^{-5}. Hits were identified and individually extracted into Geneious 9.0.5 (www.geneious.com) for downstream assembly and analysis.

1.3.2 Codon Adaptation Index Sequence Databases.

For each virus (Yellow Fever virus, Dengue 2 virus, Tobacco Mosaic virus, West Nile virus), all available complete genome sequences with collection date and country of isolation information were downloaded from Genbank. Open reading frames of sequences were exported aligned using MAFFT 7.222 (14) and then manually curated using the Geneious 9.0.5 software (www.geneious.com).

1.3.3 Codon Adaptation Indices of Viral Coding Genes.

To investigate if our selected viruses have evidence for codon adaptation in humans, we calculated the codon adaptation indices (CAI) for each coding region of each respective virus. For Yellow Fever virus, Dengue 2 virus, and West Nile virus, the complete polyprotein was used to calculate CAI. For Tobacco Mosaic virus, coding regions were concatenated into frame and then used to generate CAI values.

To calculate normalized CAI, we first used the CAIcal program (15) to obtain a "raw" CAI (rCAI). Next, an "expected neutral CAI" (eCAI) value was calculated by generating 1000 random sequences using similar length, codon composition and GC-content. Normalized CAI values were then compared among different time points and viral lineages using a non-parametric rank test because central tendencies trend varied throughout time more than each time point variance. To obtain our normalized CAI threshold, rCAI/eCAI values were calculated. Values greater than 1 were taken as evidence for codon adaptation to the reference set of codon preferences (15). Values lower than 1 was taken as evidence that mutational bias are driving codon selection.

1.3.4 Codon Usage Tables.

The codon usage tables for *Culex pipiens*, *Aedes aegypti*, *Macaca mulata*, *Columba livia* and *Rhipicephalus microplus* were downloaded directly from the publically available Codon Usage Database (www.kazusa.or.jp/codon).

We used 3 in-house codon usage tables that aimed to represent human house-keeping genes, highly expressed human antiviral immune genes, and *Nicotiana tabacum* (tobacco) house-keeping genes. The human codon usage table available on the Kazusa Codon Usage Database was last updated in 2007. Since highly or constitutively expressed genes will manifest greater codon preference biases (16–18), we first identified which genes would be potential "targets" for viral codon adaptation. The 3804 identified human house-keeping genes by Eisenberg and Levanon in 2013 (19) were selected for the human house-keeping gene codon usage table.

To identify and select highly expressed human antiviral genes, we looked for studies that quantified innate immunity gene expression during infection, specifically differences between mice and human immune responses (20,21), various protein

functionalities of the innate immune system during viral infection (22–25), and immune responses during flavivirus infection (26,27). Ultimately, we selected 25 genes that we believed to represent highly expressed antiviral immune genes. To address the concern that we may be incorrectly comparing a population's (virus) versus an individual's genetic components, we crossed-referenced all human coding sequences used with haplotype date from the 1000 Genomes Project to look for single-nucleotide polymorphisms in coding regions of interest (28). Variants in house-keeping and immune genes were very rare and did not alter the overall codon frequencies of our generated codon usage tables. Codon usage tables are made available at Github (https://github.com/CaioFreire/CUB).

1.3.5 WNV Phylogenetic inference and Malthusian fitness estimate.

First, 790 West Nile Virus complete polyprotein sequences from Genbank were downloaded. Sequences were aligned and curated by coding sequences using Muscle and Geneious respectively. Using FastTree.v2 (29), a maximum-likelihood tree was inferred, and allowed us to partition sequences by previously described lineages (30,31). Using 61 lineage 2 sequences, the rate of each substitution type under the general reversible model (GTR) substitution model, the proportion of invariant sites (I), and shape parameter of a gamma distribution with four rate categories (Γ_4) was estimated from the data. Tip times corresponding to the year of virus sampling, a lognormal uncorrelated relaxed clock using continuous quantile parameterization (32), and a GMRF Bayesian Skyride coalescence model (33) was used and run using Beast 1.8.3 (34). The MCMC analysis was run for 100 million chains to ensure convergence (achieved by a single run, with 10% of the run, discarded as representing "burn in"). A MCC was produced and annotated by the use of TreeAnnotator in the BEAST package. To infer the recent demographic history of WNV in Europe, we employed the Bayesian skyride method (33) and previously described methods to estimate the temporal dynamics of effective population size (Ne.g) of WNV, which approximates the number of infections in time. To reveal the dynamics of viral population size growth, we calculated

the Malthusian fitness (*Wm*) for the polyprotein, which was approximated by the ratio of the population size in sequential time points (*Wm* =Ne.gt/Ne.gt−1) (35).

1.3.6 Sample Collection of Guarujá

Samples were collected from November 10th, 2013, to February 25th of 2014. We collected 182 serum samples from patients who showed signs of acute-febrile illness, but tested negative for all dengue-specific ELISA and PCR assays. Patients of any age with symptoms and clinical signs of dengue disease were examined in Primary Health or Emergency Care Units. Patients were included who presented with the following symptoms and complications: fever, joint-pain (arthralgia), headache, rash, retro-orbital pain, muscle pain (myalgia), difficulty swallowing (odynophagia), vomiting, diarrhea, and thrombocytopenia. These patients reside in the coastal city of Guarujá (23°59′37″S 46°15′23″W), which is situated in the southeastern state of São Paulo, Brazil.

Patient records were retrospectively collected with the kind aid of the Guarujá Secretary of Health. From the 182 total serum samples collected, only 111 records that included basic patient information were available. Only 70 contained diagnostic information and 56 contained a physician's medical records from all Guarujá public healthcare clinics. Only 25 had usable hematology data. All adult subjects included in this study provided an informed written consent, and a parent or guardian of any child participant provided the written informed consent on their behalf.

1.3.7 Illumina Sequencing, Trimming & Contig Assembly

Viral RNA was extracted from the nasopharyngeal aspirates using the QIAamp Viral RNA Mini Kit (Qiagen; Valencia, CA), and purified with DNAse I and concentrated using the RNA Clean & ConcentratorTM-5 Kit (Zymo research; Irvine, CA). The paired-end RNA libraries were constructed and validated using the TruSeq Stranded Total RNA HT Sample Prep Kit (Illumina; San Diego, CA) was used to construct and validate paired-end RNA libraries. Sequencing was done at the Core Facility for Scientific

Research – University of São Paulo (CEFAP-USP/GENIAL) using the Illumina NextSeq platform. Each sample was barcoded individually, which allowed separation of reads for each patient. Short, unpaired reads/bases, and low-quality reads were removed using Trimmomatic v0.36 (36). Paired-end reads (Qphred>33) were assembled *de novo* with SPAdes 3.10 using default parameters (37). Contigs over 500 nts were searched against the refseq Genbank repository using BLAST.

1.3.8 Phylogenetic Reconstruction: Maximum-likelihood estimation

To infer the phylogenetic relationships of our sequences with other partial or complete genome sequences, we downloaded all available sequences from Genbank that had country and date of the samples isolation. Sequences were aligned using Clustal Omega version 1.2.1 (http://www.clustal.org/omega) or Muscle v3.8.31 (38) and manually curated by using JalView version 1.18-β8 (http://www.jalview.org/download) or Se-Al v2 (39). A maximum-likelihood tree was estimated in the nucleotide general time-reversible with gamma-distribution rate variation and invariant sites model. Support for the tree was accessed after 10,000 nonparametric boostrap replicates with FastTree version 2.1.8 (29).

1.3.9 Molecular Testing for Parvovirus B19

The detection of the B19V IgM and IgG antibodies were done using the Serion ELISA classic IgM and IgG test kits. Serum samples were extracted on the NUCLISENS easyMag platform (bioMerieux). The eluate, containing RNA and total DNA, was stored in a freezer at -70 °C until the time of its use. PCR, followed by a Semi-NESTED reaction was performed targeting the VP1/VP2 region using primers designed by Durigon *et al*., 1996 and Erdman *et al*., 1993 (40,41). The two primers used in the PCR were P12f (nt 4127-4148; 5'- CAGCCATACCACCACTGGGACA-3') and P17r (nt 4824-4803; 5'- TTACGCATCCTGGCTGAGGGCA-3'), with a fragment of approximately 697bp; semi-NESTED primers were P13f (nt 4214-4237; 5'-

GACAAAGAGTATCAGCAAGGAGTG-3') and P17r (nt 4824-4803; 5'-TTACGCATCCTGGCTGAGGGCA-3'), with final product of approximately 610 bp. The amplification was performed in a thermocycler with the following conditions: a cycle of 94 °C for 5 min for the activation of Hf, followed by 40 cycles of denaturation at 94 °C for 30 seconds, Hybridization at 55 ºC for 30 seconds and extension at 72 ºC for 45 seconds and finally a final extension cycle at 72 ºC for 7 minutes. Semi-NESTED reaction was performed using the same PCR conditions. As control, a well-known sample was used, and DNA/RNAse free water was used as the negative control. Detection of the PCR/semi-NESTED amplified product was performed by 2% agarose gel electrophoresis (Life Technologies) in TAE buffer [0.5x], stained with 0.5µg/ml ethidium bromide. In 3.0µL of bromophenol blue, 10.0µL of the sample was mixed and submitted to electrophoresis in a horizontal vessel for 60 minutes at 100V.

1.3.10 Sanger sequencing of Parvovirus B19

Samples that were found positive for B19V using a semi-NESTED PCR were selected for downstream sequencing. Purification of the PCR product was performed with the ExoSap kit (exonuclease I - Amersham Pharmacia Biotech), according to the manufacturer's instructions. After purification of the Semi-NESTED product, the amplified VP1/VP2 region using P13f/P17r primers was sequenced using the "ABI PRISM Dye™ Terminator Cycle Sequencing Ready Reaction kit" ("Big Dye" - Applied Biosystems, Inc., USA), using 3.0µL of BigDye® Terminator v1.1/3.1 Sequencing Buffer (5X), 1.0µL of Terminator Ready Reaction Mix and 1.0µL of the primer. The enzymatic extension was performed on GeneAmp PCR System 9700 thermocycler for 25 cycles of 96 °C for 10 seconds for denaturation of template DNA, 50 °C for 10 seconds for primer annealing and 60 °C for 4 minutes for extension. The obtained product was further purified with the X-terminator kit (Applied Biosystems Inc., USA) according to the manufacturer's instructions. The purified product was subjected to POP7 polymer electrophoresis using ABI model 3100 automatic sequencer (Applied Biosystems, Inc., USA).

CHAPTER 2 – LITERATURE REVIEW

2.1 A Continuously Growing Threat

The devastating 1918 H1N1 influenza pandemic, the recent Middle Eastern Respiratory Syndrome (MERS) and the Severe Acute Respiratory Syndrome (SARS), are examples of significant viral outbreaks that highlight the importance of monitoring potential emerging viruses. Viral evolution, globalization and climate change are factors that further amplify the growing danger of novel viral epidemics (42–45). Infectious diseases are the leading cause of human mortality and morbidity globally (46). Of these, zoonotic pathogens cause approximately 75% of emerging infectious diseases in humans (47). The most notorious viral examples include influenza, measles, HIV, dengue, Nipah, and SARS (48) among others. This is partially the result of 10,000 years of expansion of agriculture, domestication of animals and the increasing contact with wildlife reservoirs containing large concentrations of potential zoonotic pathogens. Today, humans are still continuously and increasingly exposed to vectors and animal infected by potentially harmful agents (49–51).

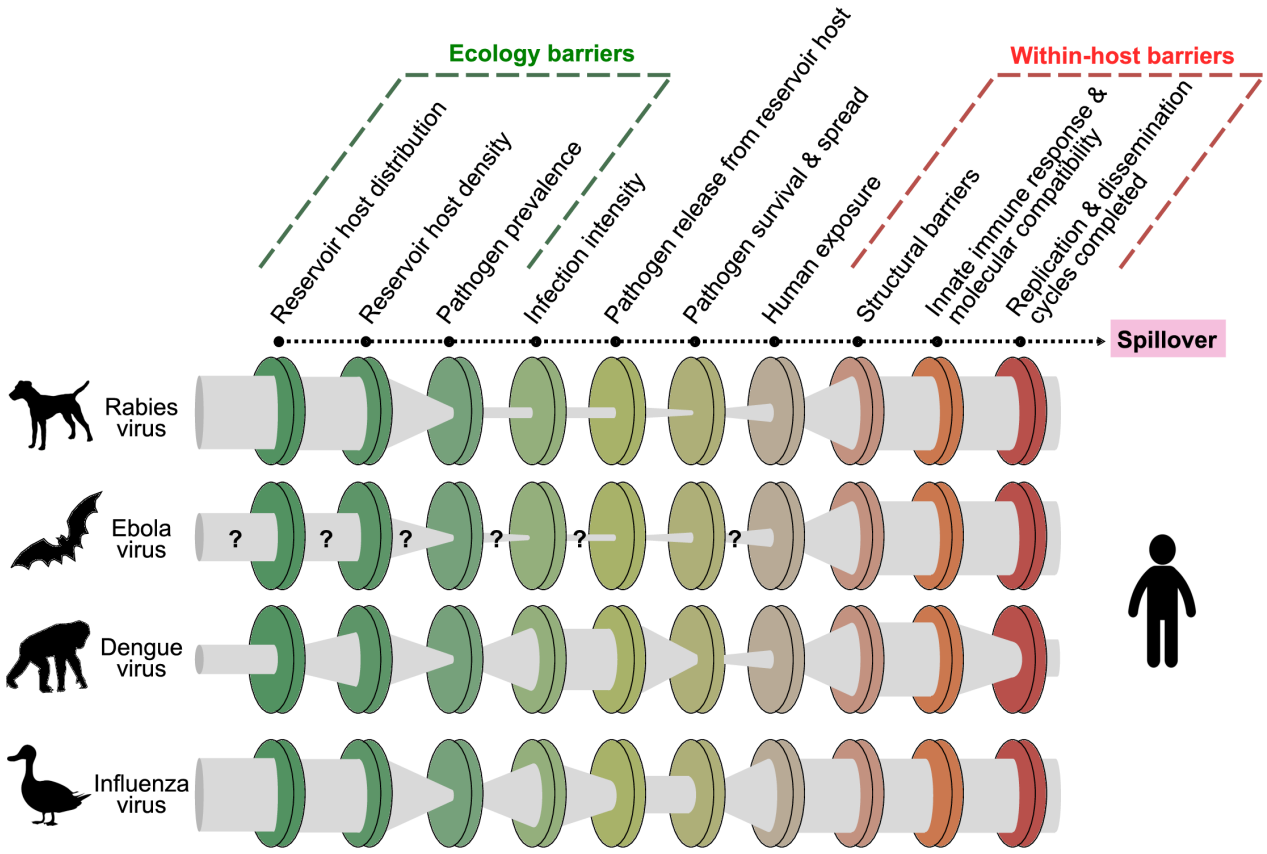
2.2 The Zoonotic Barrier

Pathogens from animal vectors usually do not cause human endemics instantly (51). Many infectious agents fail to transition and sustain themselves in human populations directly from the zoonotic host. The Marburg virus and Tick-borne encephalitis virus are two examples of viruses that have not completed the transition to sustained human-to-human transmission. For a virus to spillover from one host species to another *i.e.* mosquitoes to humans, several barriers or bottlenecks must be surpassed (Figure 1) (52).

The evolution of diseases was partially molded by domestication and increased contact with animals. Domestication has played a role in the development of zoonotic diseases infecting humans for the recent past (50). Highly infectious epidemic diseases such as measles and tuberculosis from cattle, and influenza arising from ducks and pigs

have arisen as a result of animal domestication. Nevertheless, we still do not know the origins of many notable diseases arising from domesticated animals such as smallpox.

Figure 1 – Bottlenecks of viruses on the path to spillover



Viruses need to overcome a range of barriers to jump from one species to another. The widths of the gaps in the illustrated barriers represent the ease or difficulty depending on the specific virus. Questions marks (?) represent points where information is lacking or poorly understood. For the Rabies virus, domestic dogs are widespread and infection in human hosts is generally fatal. Various aspects of Ebola virus dynamics remain poorly understood, but human-to-human transmission barriers are minimal. Non-human primates are believed to be the reservoir for the Dengue virus, but mosquito vectors facilitate transmission to humans. The Influenza virus is found is many different species, and can transmit easily between them. Figure was adapted and inspired from (52).

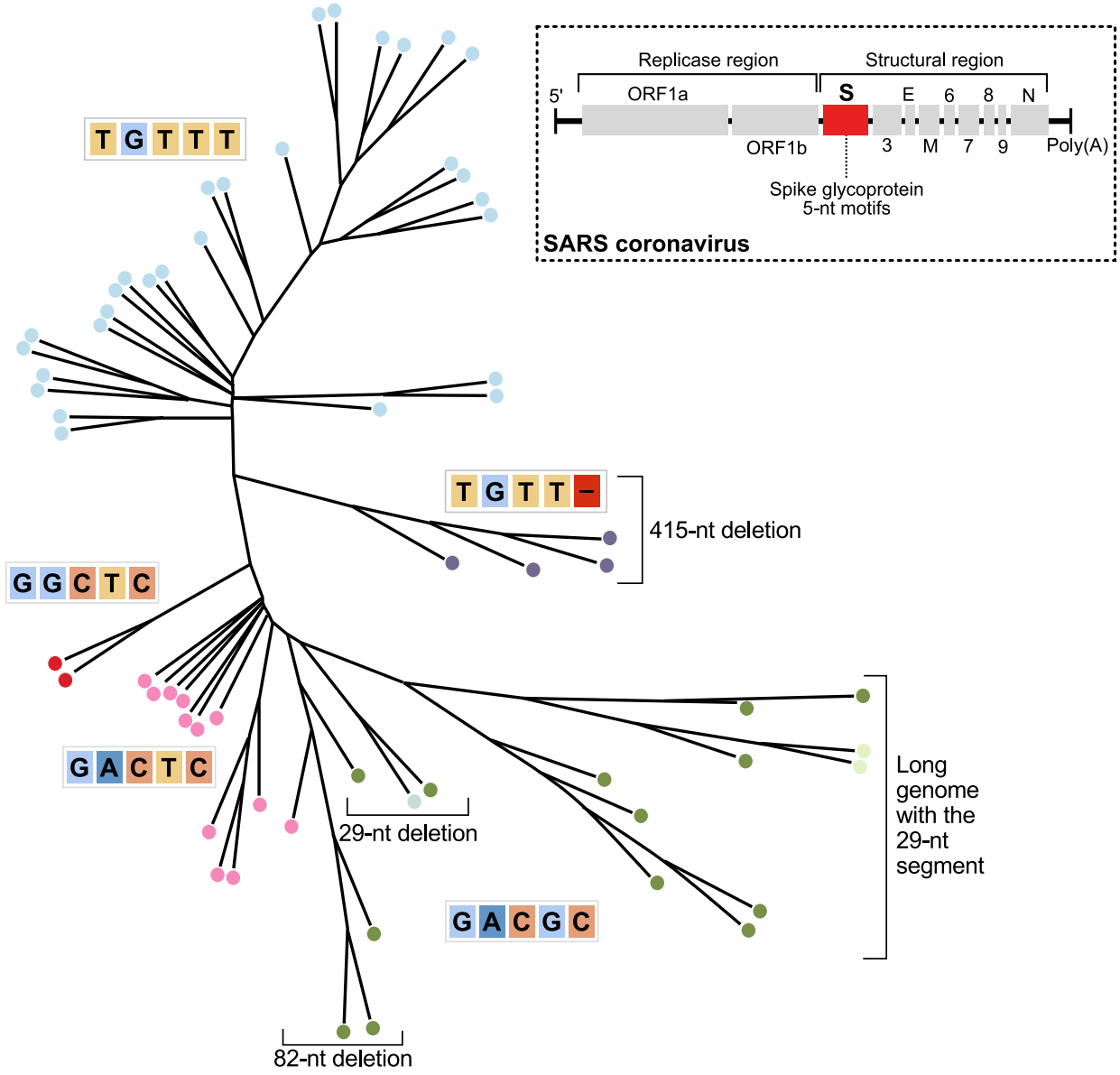
Other human-wildlife interactions such as hunting for bushmeat have led to endemics such as HIV (53). Although interspecies contact has occurred in the past, it was restricted to small human populations rather than highly dense ones that are present today in urban centers.

The evolutionary relation between the previous host and a new potential one may influence the ability of the pathogen to establish and spread. For example, chimpanzees

are the closest related species to humans, which may explain the emergence of SIV in humans, causing the HIV pandemic (48,51). Distant divergence between species does not necessarily limit the transmission ability of viruses, as exposure also plays an important role. Plague and typhus, caused by *Yersinia pestis* (54) and *Rickettsia prowazekii* (55) bacteria and viruses such as HCV appear to have originated from rodents, which are less related to humans compared to other primates. Nevertheless, large quantity of exposure events over time may in the end assist zoonoses to traffic and eventually emerge in human populations.

Genetic barriers can restrict viruses from infecting new hosts. Among other issues, viruses must be capable of recognizing receptor-binding sites and adapt to the host metabolism (56). Upon entry or fusion, they have to have purposeful directionality (*i.e.*, ontology) as well as the ability to replicate and express its own genetic material. Genetic restrictions such as a change in receptor-binding affinity take multiple and complex changes via repeated exposure and mutations before full recognition can occur. Once a virus is able to proliferate and spread among humans, one can infer geographical spread patterns that may eventually uncover the epicenter of the primary virus emergence. Phylogeographic analysis of pandemics can show how viral sequences diversify through mutations and recombination/reassortment as time and spread progresses (57–59). Case studies on Influenza A, SARS coronavirus, canine parvovirus, and HIV all show changes in genetic information after establishing strong binding in human hosts (57,60–62). For example, after SARS established ACE2 binding, a epidemiological study in China showed viral divergence and adaptive genetic changes spread across populations (63) (Figure 2). As with influenza A transition from α-2-3 to α-2-6 syalic acid receptor affinity, a labile receptor activated and bound to upper respiratory receptors allowing aerosol transmission only with a few point mutations (64). Another example, HIV-1 adapted into human populations through binding receptor changes that allowed targeting of a membrane receptor protein in humans (53,65). It is important to survey potential primary infections before a pandemic virus emerges and becomes more virulent, and capable of efficient human-to-human transmission.

Figure 2 – Viral sequences mutate with time, selectively and randomly



Sequence divergence (mutations and deletions) of the SARS coronavirus spike glycoprotein (S) gene over the course of the 2003 epidemic. Circular nodes are colored to show different SARS genotypes. A genome map (top-right) is shown for reference. Nucleotide changes are highlighted and indicated. Figure was modified from (63).

The aforementioned study showcases the importance of following specific point mutations in the context of viral adaptation that may lead to enhanced transmission or pathogenesis. Evolutionary changes in viruses are still poorly understood in terms of quantitative temporal data pertaining to advances in stages leading to human endemics. HIV-1 has a very high mutation rate per replication cycle (0.2 errors/genome per

replication cycle) and generation time of 2.5 days producing 10^{10} - 10^{12} virons (66,67). Although advantageous mutations are rare, strong evidence supports HIV-1 undergoing positive selection to avert immune defenses due to hypervariability in *env* amino acid sequences (62). For rapidly evolving viruses that do allow cross-species infection, changing from one host to another creates a bottleneck effect where only a very small population size will exhibit increased fitness in the new host organism (68). Multiple HIV strains have overcome transmission barriers in populations through successful adaption and intra-host evolution (69).

Recombination and reassortment are significant evolutionary events that certain viruses experience, which can alter their fitness. It is a single step that can impact on cell tropism and host range of some viruses. With one of the highest recombination rates of known viruses (70), HIV-1 demonstrates the difficulties in therapeutically targeting viruses that undergo frequent recombination due to transcriptase template switching (71). Well-described reassortment events caused the emergence of many influenza viruses that proliferated at pandemic levels and ravaged human populations as a result. For example, H3N2 Influenza was shown to harbor multiple genomic segment reassortments from avian viruses. Both HIV-1 and influenza demonstrate the epidemic potential of positive selection and intra-host evolution through mutation and recombination events. Human innate and adaptive immunological defenses also prevent viruses from proliferating (48) and are therefore a strong selective force shaping virus spreading potential. Interferon-γ, cytokines, APOBEC protein and TRIM5α protein all play preventative roles that a virus must overcome to successfully infect this new host. But recently, with the growing number of immune-suppressed individuals (AIDS, transplants, chemotherapy, etc.), many pathogens that would normally fail to replicate and proliferate can now do so due to the absence of innate/adaptive immunological strategies (72,73). It is therefore important to survey the adaptive diversification of potential emergent viruses among particular human populations such as those encroached among zoonotic pools.

2.3 The Evolution of Virus Surveillance and Discovery

Cell culture has long been the gold standard in viral discovery (74,75). Other techniques such as filtration, electron-microscopy and serology have also been used for novel virus discovery (76). The first virus identified using purely genetic methods was Borna disease virus (subtractive cloning) (77). Genetic techniques also enabled the identification of Human immunodeficiency virus (78), simian retrovirus (79) and hepatitis E virus (80). However, novel virus detection using traditional assays has several limitations. In order for a virus to thrive and replicate in a specific host cell culture, the conditions must mimic the natural interactions between the virus and host (42). Culturing and isolation of many uncharacterized viruses is sometimes difficult or impossible (81). For example, culturing of Hepatitis C proved to be immensely difficult. It was finally cultured in a modified human hepatoma cell line in 2005 (82).

Unlike bacteria with 16S RNA conservation, there is no universal conservation marker in viruses given the diversity of genome types and replication strategies (83,84). Although recently developed molecular assays such as degenerative PCR have identified new viruses using sequence variants derived from known viral sequences (84,85), it does not detect all viruses present in samples. Since the start of viral discovery via high-throughput sequencing, viruses have been found to be the most abundant and diverse organisms on Earth and current immunological assays will fail to uncover the substantial unknown diversity present in environmental and human samples (86,87). Therefore, given the huge diversity of viruses and the serious limitations of traditional tools for viral identification and detection, new approaches, such as indiscriminate massive parallel sequencing of samples potentially carrying unknown viruses, are today considered to be a more efficient approach.

2.4 Metagenomics

Metagenomics is sequence-based analysis of the majority of a sample´s genetic material which sidesteps many classical limitations for discovering unique virus

characteristics and origins (88). Metagenomics originally focused only on cloning double stranded DNA genomes (89). Now techniques have expanded to comprehend all types of genomic targets. This approach solves the limitations with isolation and culturing prior to full characterization (72). Unbiased metagenomic approaches yield sequencing data for almost every organism within the sample. A wide variety of applications have been developed using metagenomics since its initial applications in uncultured bacteria and Archaea (88,90). As typical viral genomes are relatively small, applying metagenomics to viral genomes is efficient due to its large coverage of sequence data (deep-sequencing) (89).

Many environmental samples have been collected and sequenced to date. Marine samples have revealed an unprecedented amount of previously unknown genetic diversity, a great part of which is viral. Surprisingly, there are few viral studies performed using human samples. Human metagenomic analyses have shed light on the micro-organism-host interactions that impact normal physiological processes. This includes the microbiota in the digestive tract as well as important interactions in immune responses to cancer and autoimmune diseases (91–93). Fecal samples are primarily used in human metagenomics primarily due to their availability and non-existent ethical concern for obtaining samples (94–97). In 2003, Breitbart *et al*. performed the first metagenomic analysis of the human gut looking for non-pathogenic human DNA viruses using fecal samples. Using partial shotgun sequencing, 59% of the sequences were unknown when compared to all known biological sequencing information (via BLAST). The results also revealed a substantial, previously unidentified bacteriophage population. In 2005, Breitbart *et al.* performed a metagenomic analysis using blood samples from healthy donors and found a novel anellovirus that further research confirmed to be common in human populations (98). Many other metagenomic analyses over the past 15 years have revealed the importance of investigating viruses in potential viral reservoirs in ecological, animal and human clinical patients (99–102).

Clinical applications of viral metagenomics have uncovered many new potential pathogens and supported a global effort for novel virus surveillance. In 2008, an idiopathic transplant-associated disease was uncovered as an arenavirus through metagenomic approaches (103). As a result, human sources such as serum/blood

samples became prevalent in clinical metagenomic studies (104–106). What makes metagenomic approaches more applicable than PCR for viral discovery are their ability to detect the presence of both known and novel viruses without prior genetic information, at concentrations that may not be detected by PCR (107,108).

Deep sequencing of clinical samples yielding millions of sequence reads can also allow for detection of subtle differences in known viruses which would be undetectable through other methodologies (103,109). Knowing key genomic differences allows better clinical inference on the pathogenicity of the virus that would not be tested for otherwise (110). Metagenomic approaches could also expand and thus improve novel viral detection assays, which will eventually inform on what therapeutic approach to use when combating a specific viral infection. Moreover, testing various patients infected with a particular virus may allow inferences on the appropriate treatment measures based on environment or phylogenic profiling. A case study on cystic fibrosis patents and the viral flora present in their lungs revealed underlying conditions that characterize aetiology for the disease states of the virus (111).

CHAPTER 3 – RESULTS

3.1 *In silico* applications

3.1.1 Using HMMs to Discover Novel Viral Sequences

3.1.1.1 Brief Introduction of HMMs and its applications

Discovering novel and uncharacterized viruses poses a new set of problems and strategies. Although it is the most common approach, BLAST-related searches are limited in their ability to account for the certainty or variability of specific residues in a query sequence (72,112). Because of the limited reads from high-throughput sequencing, as well as the occasional sequence errors and incomplete public databases, very few novel viruses are discovered (43). A Hidden-Markov model (HMM) approach characterizes unknown sequences via conserved protein-domains, which are more conserved markers than DNA sequences. The HMM-based alignments method adequately detects remote protein homology (113). HMMER3 (http:// hmmer.wustl.edu/) (Version 3.1b1) uses a Multiple Segment Viberti (MSV) filter which fully utilizes mathematical modeling principles in probabilistic sequence approaches (13). The profile will be based on a target sequence of a known protein that is found in viruses. For example capsid or polymerase proteins are both common in many viruses (42). The profile sequence should reflect sequences or homology found in either a specific or general population of viruses. Providing a metagenomic database to align for the profile will allow detection of highly divergent sequences based on various probabilistic sequences generated from the profile.

3.1.1.2 Discovery of a novel dicistrovirus in sewage

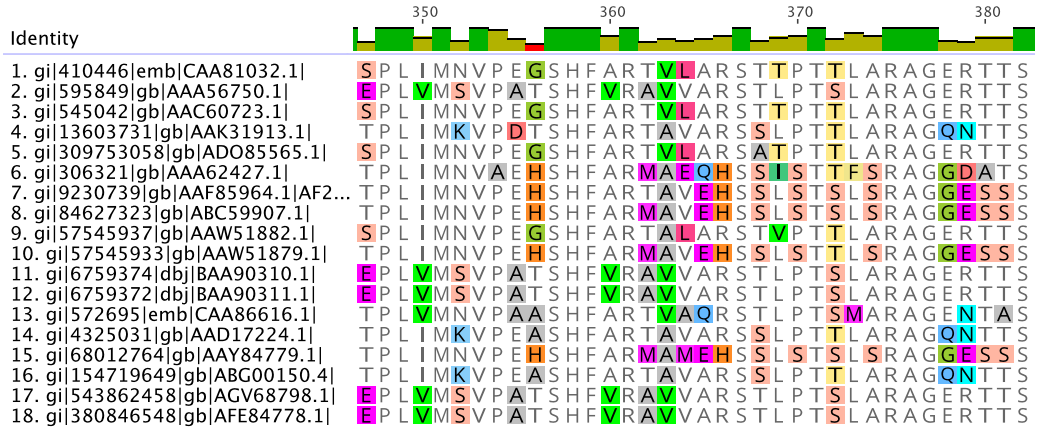
High percentages of currently available metagenomic sequences in databases are unknown with no relevant homology to any known viruses (86,114). Utilizing probabilistic modeling and advanced algorithms, we can find putative novel viruses *in*

silico. Publically available databases provided a great opportunity to apply HMM searching to follow up and uncover additional information. A recent deep-sequencing analysis on sewage (115) collected from 4 locations; San Francisco, Nepal, Maiduguri, and Bangkok (SRA accession numbers SRR545987, SRR545988, SRR545989, SRR545990 respectively) was selected for further investigation to test the effectiveness of our proposed methodology.

For segments of unknown sequences, the group utilized only BLASTx and BLASTn against the GenBank non-redundant database. With over 40% of diarrhea cases having unknown cause and the abundance of unknown enteric infections globally (116), creating a profile HMM can lead to the identification of divergent virus strains in metagenomic samples previously undetectable by BLAST or hard to detect by PCR primers based on known sequences. Therefore, a profile HMM can provide an additional method for uncovering and characterizing the unknown. Astroviruses are positive sense, single-stranded RNA viruses associated with gastroenteris in children around the world (117).

As a crucial step in this proof of concept, an astrovirus RdRp profile was created with many astrovirus amino-acid sequences utilized in Ng *et al.*'s phylogenic analysis but also expanded it with other readily available sequences from GenBank that were not included in their analysis such as in California sea lion and dog strains. RdRps contain conserved motifs widespread in many viral RNA families that are used as an evolutionary signature to qualify divergent virus proteins (118). Using MUSCLE (Version 3.8.3) (38) and SeaView (Version 4.3.1) (119), I aligned the sequence of reference RdRp(s) (Figure 3).

Figure 3 - An example of multiple amino acid alignment used for an astrovirus HMM profile.



Multiple amino acid sequence alignment using Geneious 9.1.3 (120) and MUSCLE (Version 3.8.3) (38). Sequences contain similarity but also enough variability at key positions to allow HMMER3 (Version 3.1) to infer probabilistic parameters to find highly divergent sequences within a database (121).

We then inputted the alignment and built a profile via *hmmbuild* (HMMER3.1) (121). Despite a high percentage of astroviruses found in the San Francisco database, we applied the profile to all four databases with *hmmsearch* (default parameters). Although we composed a search for astrovirus RdRps, the conservation of RdRp motifs and the probabilistic modeling algorithms of HMMER3 pulled several interesting reads from SRR545989. After trimming reads, we were able to uncover several hits that Ng *et al*. covered but also identified several reads containing a very divergent dicistrovirus strain. Dicistroviruses are found in insects and are part of a subclade of picornaviruses (122).

We further investigated one of the *hmmsearch* read hits (SRR545989.527349_6). An initial BLAST of the entire read revealed a 43% identity and E-value of $3e^{-05}$ to a Taura Syndrome virus (TSV) non-structural polyprotein. Approximately 10,000 nucleotides long, TSV has two open-reading frames (ORFs), the first including a helicase, protease and RdRp region (123). The next most similar BLAST hits were Israeli Acute Paralysis virus and Kashmir Bee virus; both retain the same structural form of TSV leading us to hypothesize that the read could infer a novel dicistrovirus present in the database, previously unidentified by Ng *et al*. Using CLC Genomics Workbench 5.5

(CLC Bio, Aarhus, Denmark) and Geneious 6.1 (120,124) (Biomatters, Auckland, New Zealand), we assembled contigs in reference to our target read. Staring from a 368nt read, we were able to extend our consensus sequence to 710 nts before lacking in supporting reads.

Figure 4 - A *de novo* assembly reveals multiple viruses “fished” from original HMM search

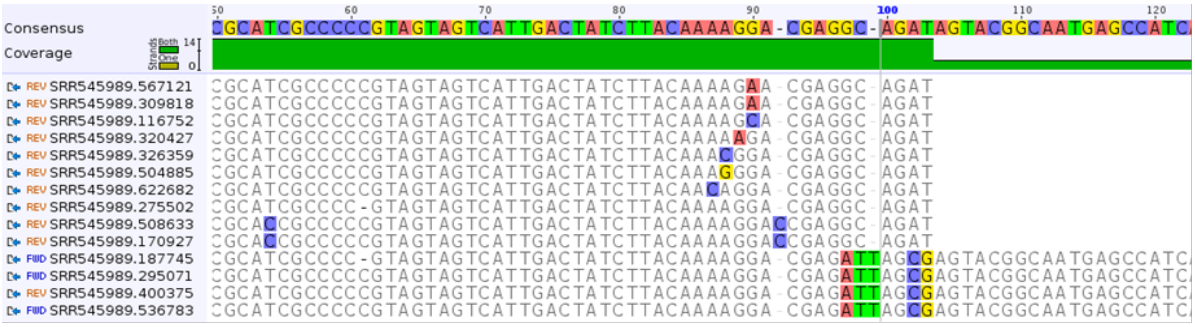
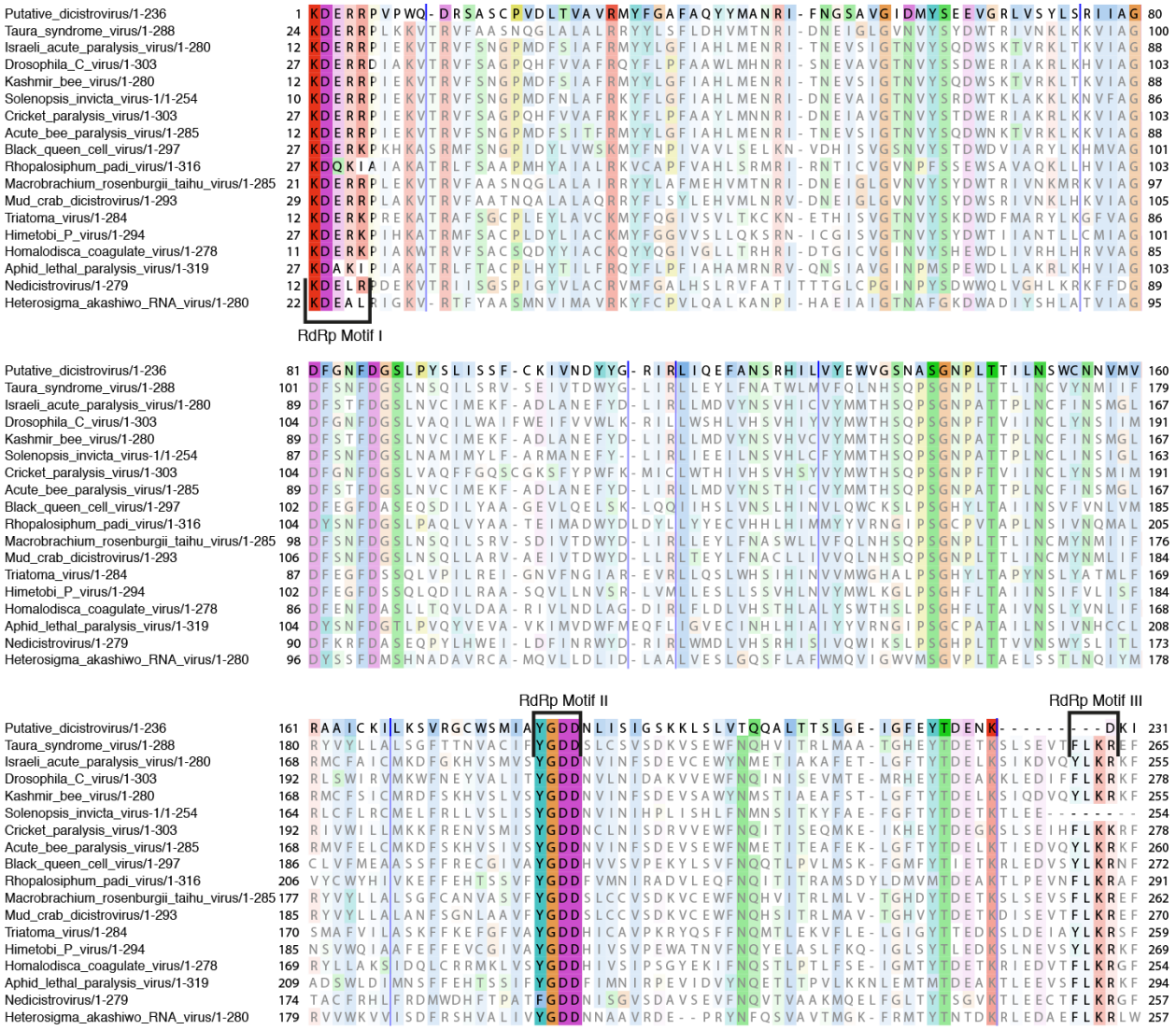


Image from Geneious (Version 6.1 from biomatters) (120,124). The bottom 4 reads represent an alternative sequence pattern. We also see multiple base changes that could also suggest similar viruses and not sequencing errors based on high coverage and other supporting sequence data shown here.

As we extended the contig, we BLASTed extensions to see if the sequences we were adding on still maintained close similarity to previous searches in terms of % identity and supporting e-value. We then mapped the sequence to the entire database as a reference. We extracted contigs that fit our stringent parameters of similarity among reads and gap size to further assemble and extend our sequence. Contigs needed to not only overlap and support previously assembled regions but also had to expand far enough with high quality and quantity to allow consensus sequence extension. This restrictive second round analysis is intended to allow the assembling of different contigs of reads prevenient to more than one genome that maps to the searched region. This possibility is illustrated by Figure 4, which shows groups of reads that have different polymorphisms and possibly include different viral species. Once contigs were chosen, we *de novo* assembled and extended contigs into even longer contigs that matched our sequence.

At a certain point, we decided we did not have enough reads supporting and extending our consensus sequence in both directions. Often, selecting the right contigs was difficult due to the presence of multiple viruses present in contigs that bypassed strict assembly parameters. Although difficult, it demonstrates the ability of our methodology to detect multiple viruses and extend separate contigs on those specific viruses if we chose to do so. After we extended our consensus sequence in both directions to the furthest point with enough support, we used pBLAST to evaluate closest matches in all available GenBank sequences. We then extracted 17 protein sequences that shared close %identify and e-value significance from our search and the sequences from the original study (115) (Figure 5).

Figure 5 – RNA-dependant RNA polymerase multiple alignment of picorna-like viruses

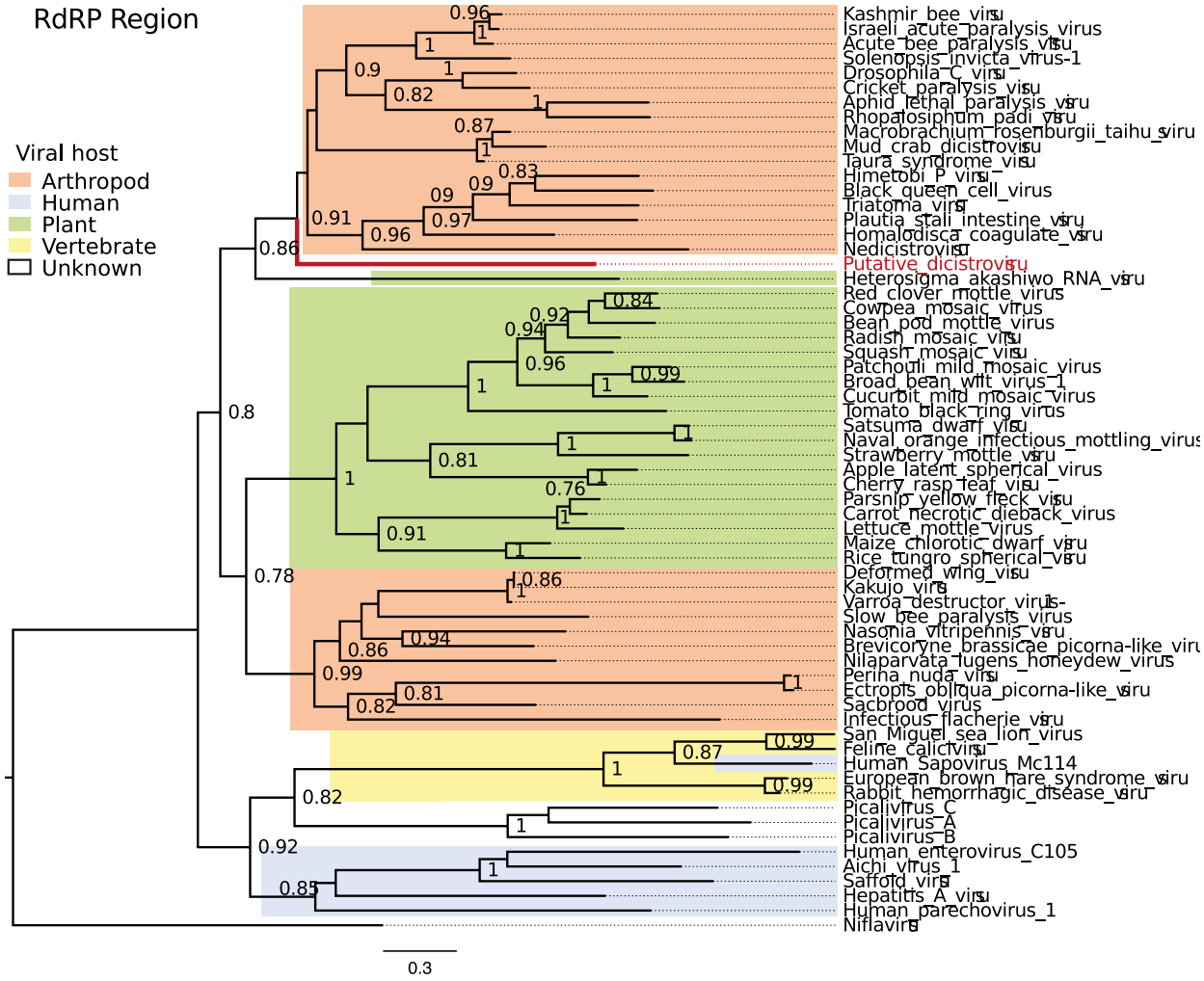


Top sequence is our extended amino acid sequence tentatively named Putative_dicistrovirus. With a final consensus protein sequence almost tripled from its original read, a multiple alignment of 18 dicistrovirus sequences was performed using MUSCLE (Version 3.8.3). Using Clustalw coloring (125), demonstrate our putative novel virus RdRp containing strong conservation in 2 motifs (KDEX and YGDD) and highly conserved amino acid positions (126). We were not able to extend to the 3rd conserved motif. All sequences were downloaded from GenBank.

Three conserved motifs are needed for RdRps to retain function despite RNA viruses' high evolutionary mutation rates (68). We cannot deduce the evolution of picorna-like viruses solely on RdRp evolution (127). However, since RdRp is the only universal protein within the family of picornaviruses and it is the only portion of the virus we have fished out from the database, we performed phylogenic analysis to theoretically place our putative dicistrovirus in reference to known lineages of picornaviruses (122).

With complete repeat of Ng. *et al*. phylogenic profile sequence selection for RdRps, we see a highly divergent novel virus protein from anything currently published and suggest an early divergence from the picorna-like family tree (Figure 6).

Figure 6 – Phylogenetic estimate of a putative dicistrovirus and the picorna-like virus family



Based on the alignment from Figure 6, amino acid sequences from RdRP regions were phylogenically estimated. Maximum-likelihood trees were estimated using FastTree (Version 2.1) (128) with the Whelan and Goldman (WAG) (129) evolutionary model. Maximum-likelihood trees were run with 1000 bootstrap replications; branches with 70% or greater bootstrap values are labeled. Consistency of the tree was measured with the tree from Ng *et al*. study. Niflavirus was selected as an out-group to match their study. Colors represent virus host organism.

We argue that a sample of our own would allow for a second round of sequencing and for the design of primers that could either generate more reads or allow for standard PCR techniques that can finish the putative genome found. If we had the original sample, our next step would be to extract our novel virus protein using unique primers. This would allow us to confirm the presence of a new virus and complement our *in silico*

analysis. This methodological proof of concept not only demonstrates the effectiveness and power of probabilistic modeling but also shows the abundance of novel viruses in unpublished metagenomic databases.

3.1.2 Applications of Codon Adaptation Index in Flavivirus Biology

The following section contains material that was accepted for publication in PLOS One on January 10, 2018.

3.1.2.1 Brief introduction of Codon Adaptation Index and pathogenic flaviviruses

When a virus first jumps into a human host from a non-human vector or reservoir host, conserved changes in viral population sequences can occur that could allow the virus to adapt to a new host (17,18). Part of host adaptation relies on coping with a host's cellular environment, in particular, the translational machinery that will allow a virus to complete its replication cycle. Codon-bias analyses have proven to be effective in extrapolating information from a zoonotic virus infection in a new, vector or reservoir host (130–135). During viral replication, the coding regions of viral genomes undergo both mutational and selective pressures that result in non-synonymous and synonymous changes that may or may not result in adaptive changes. Codon adaptation index (CAI) quantifies how well the codon preferences of the virus match that of the host (16,131). For example, a host cell that uses more frequently AGA to code for Arginine will generally have a greater pool of tRNAs for AGA in its cellular environment, as opposed to a rarer, less used codon for Arginine such as CGU. Although cellular conditions are "tailored" for the host's translational needs, a virus could optimize its replicative and translational kinetics by selecting for codons more abundant in the host's cellular environment. Although it has been shown that the translational efficiency of some persistently infecting viruses is reduced to diminish protein expression to avoid immune surveillance (134), the opposite has been observed with acutely infecting viruses.

The availability of codon usage tables for a host's constitutively expressed or highly expressed genes could allow analysis of adaptive changes of acutely infecting zoonotic viruses (17), as they interact with new hosts. Convergence or divergence of

viral codon adaptation towards or away from that of host can be observed over relatively short periods of time (i.e., years), allowing side by side comparisons with their epidemic history. There is plausibility for increased CAI and the occurrence of human disease from acutely infecting zoonotic viruses. It can be argued that a virus, whose codon usage is calibrated to the host's usage, may have the capability to translate and replicate faster than a virus that has a lower codon adaptation to a specific host (136,137). Because higher replication tends to increase viral titers, higher host viral load could allow for an increased chance of secondary transmission, virulence and higher chance of mortality from the infection (138).

Recently, it was demonstrated that there is an increasing CAI of ZIKV's NS1 protein to human housekeeping genes in Brazilian strains (139) and that this adaptation correlated, as predicted, with changes in its proteins stoichiometric ratios (Freire et al., unpublished). Moreover, it was found that the ZIKV strains circulating in Brazil infects human neural tissues more efficiently, while an African ZIKV strain (MR766) is more adapted to infect non-human primates neural tissues (136), providing empirical support for findings on ZIKV CAI (139). In light of this data, we asked if the codon adaptation of other flaviviruses to various hosts could be observed at the genus and species level.

A requirement for flaviviruses, and arborviruses in general, is the initial need to adapt to the codon choices of a vector species before the virus adapts to the codon usage choices of humans (140). The alternation between mammals and arthropods impose constraints on arbovirus evolution. Specifically in the vector, albeit strong purifying selection, we can observe silent selection, that is changes in 4-fold degenerate codon sites, promoting differential codon usages in viral genomes (141–143).

Yellow Fever virus (YFV), Dengue virus (DENV) and West Nile virus (WNV) are globally prevalent flaviviruses that cause morbidity and mortality in humans. These three flaviviruses have extensive genetic and biological diversity within themselves, i.e. sylvatic and endemic strains of DENV and YFV, or the distinct genetic lineages of WNV (144,145), which vary in reservoir hosts and ability to infect humans in distinct biological behaviors (8,31,146–154). To provide insights into the usage of temporally scaled CAI of YFV, DENV, and YFV in zoonotic emergences, we have performed a large analysis investigating different flavivirus vector-host relationships. We first created two codon

usage tables for specific human genes that could play a role in viral codon adaptation. Next, we looked for broad codon adaptation patterns within the flavivirus genus subgroups. We used an in-depth approach to investigate the codon adaption of YFV, Dengue 2 virus (DENV-2), and WNV to known hosts by using a large set of serially stamped sequences from widespread geographical origins. Our understanding of how viral species modulate translation and adapt to host cellular environments over time is limited. From these data, we show that flaviviruses have distinct patterns of codon adaption linked vector-host subgroups and biology, and we provide evidence that species-specific codon adaptation is observable and possibly linked to the vector-host evolution across time in certain flaviviruses.

3.1.2.2 Results of Codon Adaption Index and Pathogenic flaviviruses

House-keeping genes are constitutively expressed in all human cells (19,155); being therefore a likely venue for viral codon adaptation. Moreover, it is fair to assume that human immune genes shape many viral-host interactions and disease severity (20,25). Usually during a viral acute infection, immune responses are triggered including an increased expression of antiviral genes (22,24,26,156,157). We did not have any evidence supporting that the host's tRNA pools alter during viral infection or if the codon preferences of a virus change when cellular immune processes initiate during a recognized viral infection. The immune antiviral genes table was created to provide an alternate human codon table that has different codon usages than that of human housekeeping genes. Between the two human tables, the average difference between the frequencies per thousand bases among triplets is 4.259 (*s.d.* = ±4.104) with a max difference of 18.1.

We first wanted to show that there are measurable and meaningful viral-host adaptations that could be observed using CAI. As a proof of concept for virus-host codon adaptation, we calculated the CAI of Tobacco Mosaic virus (TMV) using complete coding sequences against the codon preferences of tobacco (*Nicotiana tabacum)* housekeeping genes (158,159), TMV's "natural" host (Figure 7). When TMV sequences were tested against the codon preferences of human genes, there was no evidence for

codon adaptation. The average CAI of TMV to tobacco housekeeping genes was 1.037 (s.d. = ±0.005), showing strong evidence for adaptation. On the other hand, TMV's CAI against human housekeeping and antiviral genes showed no evidence for codon adaptation with an average CAI of 0.993 (s.d. = ±0.003) and 0.981 (s.d. = ±0.003), respectively.

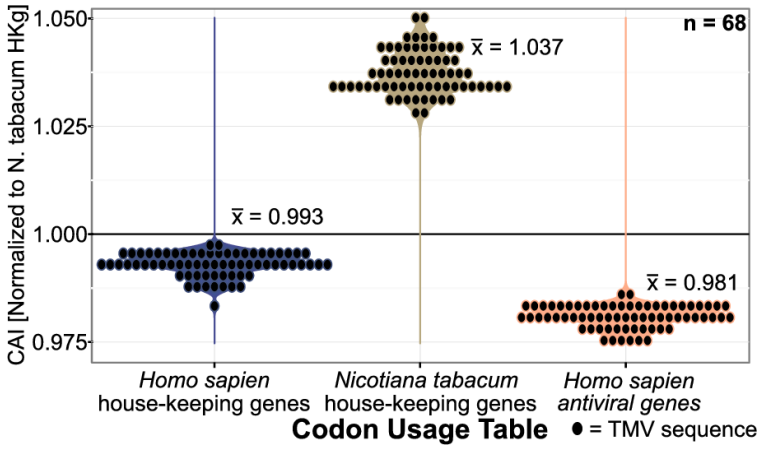


Figure 7 - The CAI of Tobacco Mosaic virus to housekeeping genes.

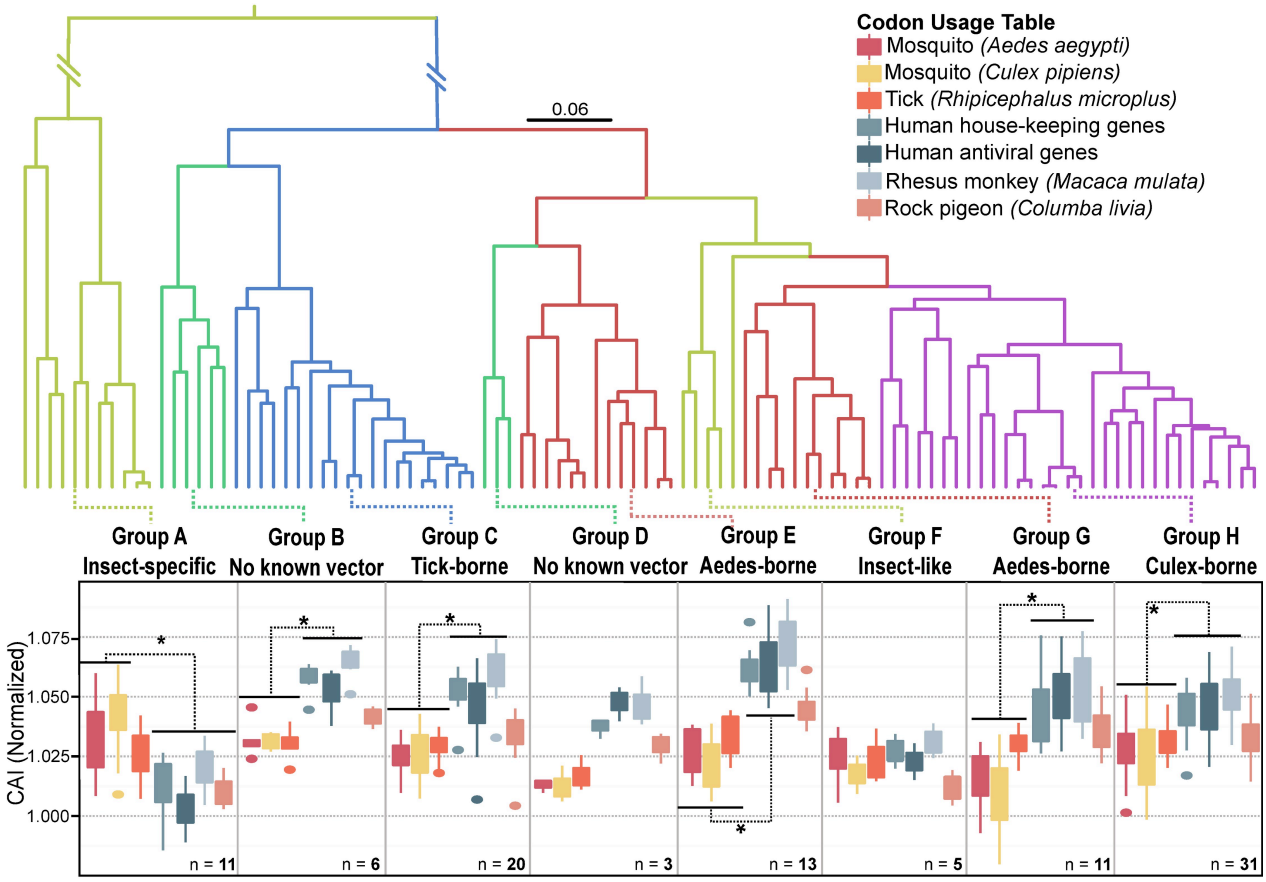
A violin dotplot of the CAI of complete Tobacco Mosaic virus (TMV) coding sequences to the codon usages of human house-keeping genes, antiviral/immune genes, and tobacco (*Nicotiana tabacum*) house-keeping genes. Average CAI values for each group are shown.

After demonstrating that CAI could be used to measure a virus' codon adaptation to specific hosts codon choices, we hypothesized that similar relationships could be observed in flaviviruses. They are arthropod-borne viruses that have the ability to infect and replicate in hosts of different phyla. Therefore, versatility in gene expression and protein synthesis is at premium, and certain changes in the viral RNA genome could affect the fitness of the virus in a specific host. Indeed, viral fitness changes relate to dinucleotide frequencies, codon preferences, and codon pair biases (132,141,142,160,161).

Ecology, different virus-host relationships and biogeographical migrations of flavivirus species may explain observed genetic differences among subgroups (140,142,143,162,163). Phylogenetic studies have acknowledged three groups that reflect the evolutionary and ecological dynamics: tick-borne, mosquito-borne (*Aedes* and *Culex* species), and no known-vector flaviviruses (164–170). We reconstructed the

evolutionary history of the Flavivirus genus based on the complete coding sequences of species used in Moureau *et al*.'s 2015 paper (170), except that we used a nucleotide alignment. As biological and genetic differences are evident between each host/vector subgroup, we calculated the normalized CAI for each sequence in their respective clade against 7 codon usage tables: *i)* human housekeeping, *ii)* highly expressed human antiviral immune genes, *iii) Aedes aegypti*, *iv) Culex pipiens*, *v) Rhipicephalus microplus* (tick), *vi) Macaca mulata* (rhesus monkey) and *vii) Columba livia* (rock dove pigeon) (Figure 8).

Figure 8 - Flavivirus-CAI dynamics to associated species.



(A) Bayesian-inferred phylogenetic tree of the complete ORF using nucleotide sequences. All nodes had a posterior probability > 0.9. Taxa were omitted for clarity. Associated viral vectors and vertebrate host groups are colored. The scale bar represents 0.06 mutations per site. (B) Box and whisker plot of CAI for each flavivirus species subgroup. For each group of sequences, the CAI was normalized by length, GC-percentage and amino-acid content. Black lines with asterisks signify CAI values that are significantly different.

Insect-specific groups A includes Cell fusing agent virus (CFAV), Culex flavivirus (CxFV), Aedes flavivirus (AEFV), Kamiti River virus (KMV) and seven others (S1 Figure): all viruses in this group are thought to be vertebrate-incompetent both *in vitro* and *in vivo* (169,171–177). Both CAI values to human codon and pigeon tables compared to insect vectors were significantly different; with vector CAI values higher (Wilcoxon rank sum test, *p*-values ranged from 1.985×10^{-5} to 0.0457). Although all CAI results for vertebrate CAIs were significantly lower to the *Culex* and *Aedes* tables, we did not find any significant difference between tick and monkey codon adaptation within the group A clade.

The other insect-specific flavivirus group F clade includes 5 species (Chaoyang virus (CHAOV), Donggang virus (DONV), Lammi virus (LAMV), Barkedji virus (BJV) and Ilomantsi virus (ILOV) paralogous to mosquito-borne flaviviruses, supporting a more recent emergence and closer relation to *Aedes* and *Culex*-borne viruses. The general equality across all of the CAIs of group F is different than the other insect-specific group (A), where the CAI of vector species is visibly greater than vertebrate species. The group F species showed no clear relationship between vertebrates and vector species codon usages.

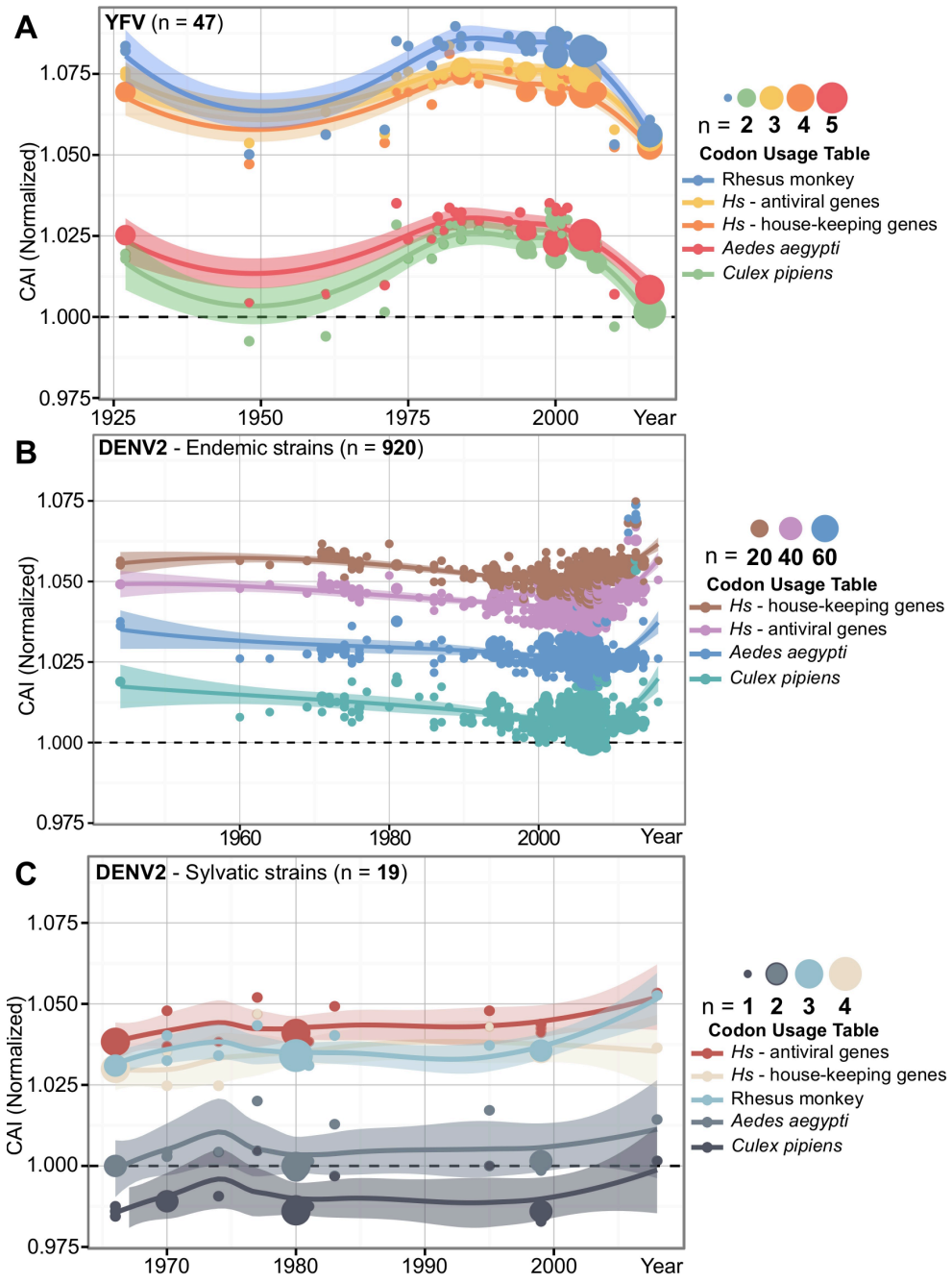
Mosquito-borne and tick-borne groups that have known associations with a vertebrate host revealed a different codon bias. In groups C (tick-borne), E (Yellow fever virus complex), G (Dengue virus complex) and H (West Nile virus complex), viral codon adaptation indexes were significantly higher in mammalian vertebrates than in vector species (Wilcoxon rank sum test, p-values ranged from 5.387×10^{-7} to 0.0192). Additionally, we found that CAI values were on average higher in monkeys than in human genes for vertebrate-associated flavivirus groups, and in some cases differences were significant. A higher CAI average in vertebrates compared to vectors was universal in all vertebrate-associated groups. Interestingly, the group B clade (no known-vector) shared a similar trend to mosquito-borne and tick-borne CAI box plots, where mammalian vertebrate CAIs were significantly higher than vector species (Wilcoxon rank sum test, p-values ranged from 0.002 to 0.008).

The *Culex*-borne flavivirus clade (group H) emerged from *Aedes*-borne flaviviruses thousands of years ago (165,178). *Culex*-borne species have a life cycle involving primarily avian species and a few species have even been isolated from rodents. Like the group G *Aedes*-borne flaviviruses species, CAI values to mammalian vertebrate tables were significantly higher than in mosquito species (Wilcoxon rank sum test, p-values ranged from 5.83×10^{-9} to 1.032×10^{-6}). In the case of the rock pigeon (*Columba livia*), the Culex-borne flaviviruses had a lower CAI distribution to pigeons than to mammalian vertebrates (Wilcoxon rank sum test, p-values ranged from 2.201×10^{-8} to 0.000136). Furthermore, *Culex*-borne species had a significantly higher CAI to pigeons than to *Culex pipiens* (Wilcoxon rank sum test, *p*-value = 0.007) but only averaged higher than *Aedes aegypti* and tick CAI values. Across flavivirus subgroups that are

known to harbor viral species with known capability of infecting mammalian vertebrates (B, C, D, E, G, H), the CAI to pigeons rests in the middle of vector and mammalian CAI values.

After observing CAI differences between the flavivirus genus subgroups, we explored if changes in CAI over time could be seen in specific viral species. Yellow fever virus (YFV, group E) and Dengue 2 virus (DENV-2, group G) are paralogous *Aedes*-borne flaviviruses who boast century-long interactions with human hosts (153,179,180). Both had large numbers of serially stamped complete ORF sequences available, allowing us to investigate the temporal trends of YFV CAI, and compare differences in the CAI of DENV-2 endemic and sylvatic strains (Figure 9).

Figure 9 – CAI changes across time for (A) Yellow Fever virus, (B) endemic strains of Dengue 2 virus, and (C) sylvatic strains of Dengue 2 virus.



For each codon usage table, the CAI was normalized by length, GC% and amino acid content for each dataset. Area of plot points reflects the density of sequences at a specific coordinate. A trend line was generated using LOESS, a non-parametric regression method, with 0.95 confidence interval shading. For B), CAI data to monkeys was removed for clarity, but was positioned in between the human table trend lines.

To start, we calculated their CAI to human gene codon usage datasets, the Rhesus monkey (*Macaca mulata*) *Culex pipiens* and the *Aedes aegypti* codon usage tables. For the YFV dataset (Figure 9A), the 5 CAI trendlines were highly correlated (Spearman's rank correlation test, p-value ≤ 2.2×10^{-16}, *rho* ranged from 0.977 to 0.999). YFV demonstrated codon adaptation for both vectors and humans across all sampled strains. YFV CAI to Rhesus monkey codon usage was significantly higher than human and vector tables (Wilcoxon rank sum test, p-values ≤ 6.375×10^{-7}). We also observed significantly higher CAI for *Aedes aegypti* than in *Culex pipiens* (Wilcoxon rank sum test, p-value = 0.006). Across all three datasets, YFV had a higher CAI to the mammalian codon usage tables, followed by DENV-2 endemic and sylvatic strains respectively.

A higher CAI to its *Aedes*-associated vector species compared to the common *Culex*-borne vector was also true for endemic and sylvatic strains of DENV-2 (Wilcoxon rank sum test, p-values < 1.711×10^{-5}) (Figure 9B and 9C). DENV-2 endemic strains, showed highly correlated trend lines across all CAI tests (Spearman's rank correlation test, p-values < 2.2×10^{-16}, *rho* values ranged from 0.875 to 0.996). Since we were more interested in the interplay of DENV-2 endemic strains with humans, particularly when transmission alters between mosquitoes and humans in an urban setting, we removed the CAI data of DENV-2 endemic strains to the Rhesus monkey table. DENV-2 endemic strains showed a significantly higher CAI to housekeeping genes over time (Wilcoxon rank sum test, p-value < 2.2×10^{-16}). This is divergent to YFV, where CAI over time was greater to the immune genes codon usage table (Wilcoxon rank sum test, p-value = 5.775×10^{-5}).

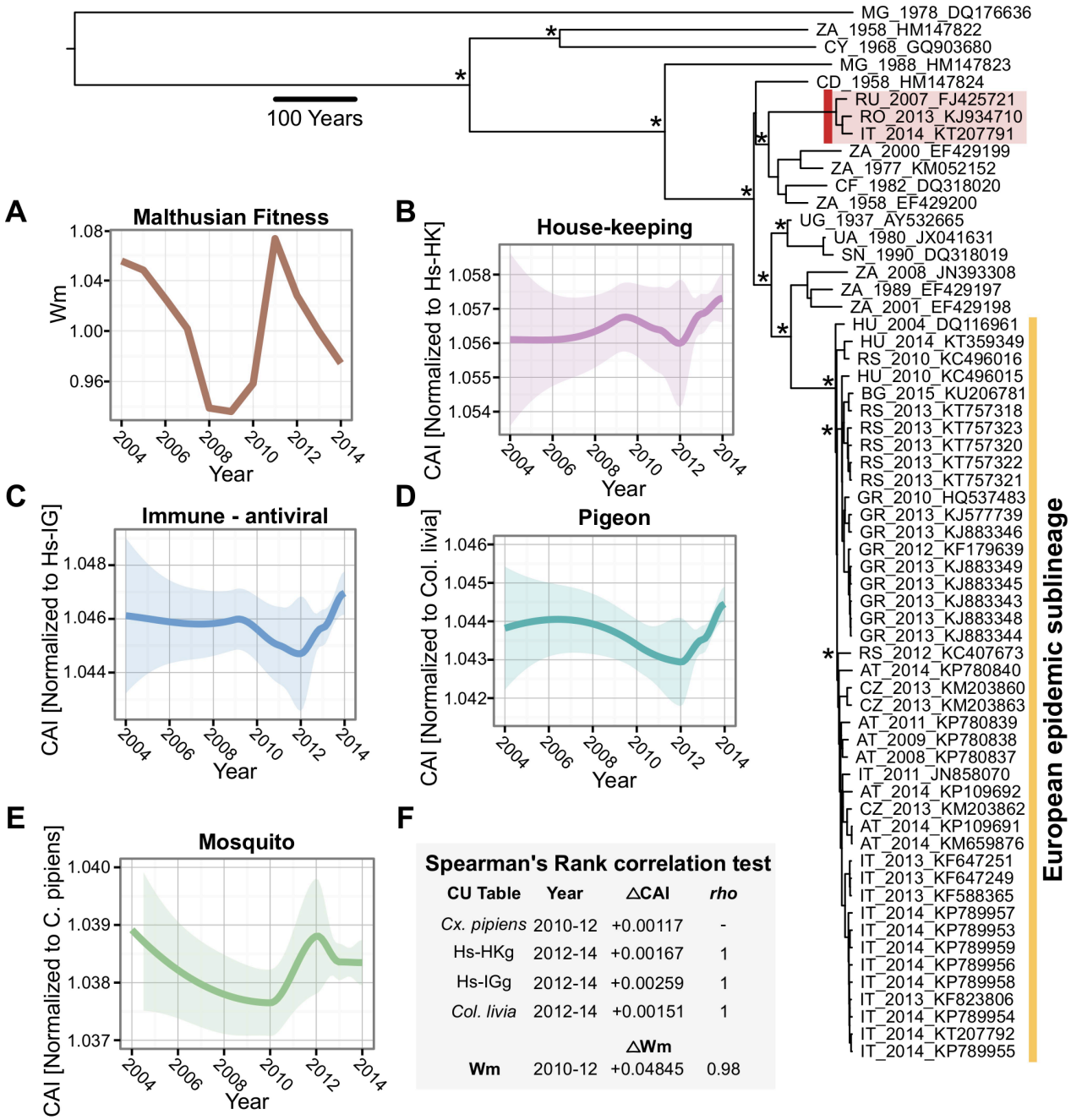
In DENV-2 sylvatic strains, we saw a different pattern (Figure 9C). DENV-2 sylvatic strains had a higher CAI towards immune genes than they did in house-keeping genes (Wilcoxon rank sum test, p-value = 1.919×10^{-5}). Furthermore, sylvatic strains did not show any evidence for codon adaptation towards *Culex pipiens*, as the CAI for all sequences was below 1. Interestingly, codon adaptation to the Rhesus monkey shows an increase in CAI from 1995 to the 2008 sequences, positively correlating with the same increase in CAI that takes place later in DENV-2 endemic strains to human genes from 2003 to 2016 (Spearman's rank correlation test, p-value ≤ 2.2×10^{-16}, *rho* ranged from 0.943 to 0.996). In fact, this same positive correlation was shared in DENV-2

sylvatic strains codon adaptation to *Aedes aegypti* (Spearman's rank correlation test, p-value ≤ 2.2×10^{-16}, *rho* ranged from 0.946 to 0.995).

After exploring the CAI trends of YFV and DENV-2, both *Aedes*-borne flaviviruses, we then asked if a similar result could be found in a *Culex*-borne flavivirus such as West Nile virus (WNV). WNV is transmitted primarily by the *Culex pipiens* mosquito and can cause severe neurological disease in vertebrates: including humans, horses and birds (181). Furthermore, the virus has been isolated in every continent except Antarctica (31). In addition to its vast geographical spread, the WNV species shows high genetic diversity and exists in at least 7 different lineages (144). WNV lineage 2 (L2) was historically thought as a milder, less pathogenic lineage than its lineage 1 counterpart and was believed to be restricted to sub-Saharan Africa (182,183). In 2004, the known boundaries of L2 changed when a Hungarian goshawk (*Accipiter gentilis*) had neurological symptoms that were later identified as WNV disease from a L2 strain (184). L2 was detected in birds, mosquitoes and humans sporadically over the next few years in Eastern European nations (185). In 2010, a large WNV L2 human outbreak in Greece was reported, with 81 confirmed cases of West Nile neuroinvasive disease (186). From 2010 to 2014, WNV human outbreaks were reported in Russia, Czech Republic, Italy, Austria, Serbia, Hungary, Belgium and Romania (187).

With the availability of African and recently sequenced European WNV L2 complete genomes, we investigated the CAI changes of L2 during its recent emergence out of Africa and into Europe. We first show that there were at least two different introductions of L2 into Europe since 2003 (Figure 10A). Interestingly, the epidemic sublineage (n=43) had a higher CAI than the 3 European sequences outside the clade for vertebrate and mosquito tables (Wilcoxon rank sum test, *p*-values ranged from 0.001 to 0.02). We then calculated the Malthusian fitness (*Wm*) for L2 from 2004 to 2014, a time period that coincided with L2 strains isolated and sequenced in Europe (Figure 10B).

Figure 10 - CAI and phylodynamics of West Nile virus lineage 2 sequences.



A) Bayesian maximum clade credibility tree representing a time scaled phylogeny of a WNV lineage 2 polyprotein sequences. Bayesian posterior probabilities > 0.9 are marked with an asterisk at major nodes. Averages for the European epidemic lineage (yellow bar) and a 2nd European lineage (highlighted in red) are shown. B) Malthusian fitness (*Wm*) was calculated from 2004-2014, and compared to LOESS trend lines generated from CAI values to the codon usages of C) human house-keeping genes, D) human immune/antiviral genes, E) pigeon (Columba livia) genes, and F) mosquito (*Culex pipiens*) genes. G) The Spearman's rank correlation test was used to test if there were any correlations between the 2010 to 2012 CAI increase in mosquitoes and the CAI increase in vertebrate species from 2012-2014, as well as Wm. The ΔCAI, Δ*Wm* and *rho* are shown for clarity. # = for all correlations, *p*-values were < 0.05.

We then calculated the CAI of L2 sequences against human, mosquito (*Culex pipiens*) and the pigeon (*Columba livia*) codon usage tables. With *Wm*, a measure of relative fitness over time, we were able to compare two independent measurements of viral fitness that span before and during the European WNV L2 epidemic (Figure 10B-10F). We first looked for any trends of *Wm* or CAI from 2004 to 2014 that were similar. We discovered that the trend lines for mosquitoes and *Wm* were correlated (Wilcoxon rank sum test, p= 0.005, *rho* = 0.8). Notably, the joint increase of *Wm* and mosquito CAI during the 2010 to 2012 period was highly correlated (Wilcoxon rank correlation test, p = 0.017, *rho* = 1). Interestingly, an increase of CAI in pigeon and human tables followed in 2012 to 2014 that was highly correlated to the earlier increase in mosquitoes and Wm (Spearman's rank correlation test, p-value ≤ 0.017, *rho* ≥ 0.98). The increase in vector CAI that precedes the increase in human CAI was similar to what we observed in sylvatic DENV-2 CAI to *Aedes aegypti* and rhesus monkeys, followed by a later increase in the CAI of endemic DENV-2 strains in humans.

3.1.2.3 Discussion of Codon Adaption Index and Pathogenic flaviviruses

Observing changes that may be involved in translational efficiency offers a promising avenue for a better comprehension to how some flaviviruses can infect and replicate in human hosts. Codon usage bias studies are inherently common, but they rarely consider the changes in codon adaptation to hosts that occur over time, as we have done here. Although we observed significant differences in the normalized CAI of different viruses, we still lack the knowledge to infer what biological consequences these small changes could cause. Although the critique of sequencing bias, including the scarcity of aged samples, and lack of non-human sequences can limit CAI analyses, we have shown that the results can be informative.

One advantage of calculating the CAI of flaviviruses is the single open-reading frame that codes for around 3400 amino acids. Using such complete polyprotein reading frames, we were able to observe differences in CAI between previously inferred vector

and vertebrate phylogenetic groups (132,170,188,189). In particular, insect-specific groups favored the codon choices of vector species, while mammal-associated flavivirus groups were better adapted to vertebrate codon usage. Insect-specific flaviviruses are thought to sustain their populations in their respective insect vectors in the absence of mammal reservoirs (190), so lower translational efficiency in vertebrates could be expected.

As expected, some of our results could not be connected to some of the known biological aspects of flaviviruses. This could suggest that there may be uncharacterized features of the translational mechanisms at play. For example, the results we obtained using the tick (*Rhipicephalus microplus*) table did not support the biological and epidemiology of tick-borne viruses (Figure 8). Indeed, tick-borne viruses cause dead-end infections in humans (191), so we did not expect CAI to be as high, especially compared to some mosquito-borne viruses that can cause secondary infections in humans (17). Some tick-borne viruses use rodents as their reservoir hosts, and have associations with other mammalian vertebrates such as dogs, horses, and other large mammals (192), which could have lead to tick-borne viruses having similar codon usages to humans and monkeys. Nevertheless, we did not investigate the CAI of flaviviruses to rodents so far.

The number of insect-specific and arboviral sequences as a whole are rapidly increasing (193,194), which will allow for more in-depth analyses in the future. The capabilities of CAI analysis are limited to the availability of serial-stamped sequencing data, specifically older isolates. Furthermore, the public availability of complete genome sequences has increased as cost for sequencing has decreased and improved over time. Occasionally, we also encountered a limitation as to what inferences we could make on the biology of flaviviruses. For example, in Figure 9C, there is only one complete DENV-2 sylvatic genome isolated in 2008 that suggested a CAI increase in rhesus monkeys and *Aedes aegypti* since the turn of the millennia that was later followed by an increase in human CAI. This was also the case for West Nile virus, where L2 showed an increase in Malthusian fitness and *Culex pipiens* prior to an increase in human CAI (Figure 10B-G). Although there was not enough DENV-2 sylvatic sequences to calculate Malthusian fitness, comparing Malthusian fitness and CAI could help elucidate evolutionary dynamics of an emerging virus (139).

Crucially, just one sequence could be enough to make statistical inferences regarding changes in codon adaptation to a specific host. This is due to the fact that every one of the 3400 codons in the genome is counted, and tested against 500 random sequences with the same length and amino acid content to determine if codon changes are adaptive or random. Therefore, arguably every flavivirus sequence included in our analysis provides substantial amount of data.

During an ongoing epidemic where sufficient sampling and sequencing have been generated, we could use these sequences to observe changes in fitness of a virus that could coincide with fluctuations of CAI to different hosts. This was the case for WNV L2, for which we had access to 46 complete polyprotein sequences sampled in Europe since 2008. Our phylogenetic inference supported results from previous works that suggest multiple entries of WNV into Europe (148,195–197). Interestingly, we found that a small group of European sequences (highlighted in red in Figure 10) had significantly lower CAI values in mammals than the epidemic sublineage composed of the rest of the European L2 sequences. This could be explained by sampling bias or possibly that these strains lacked the translational efficiency to spread with the same force as the other European strains (Figure 10A). We lack *in vivo* experimental data, but we can hypothesize how our results can shed light on the complexity of viral translational adaptation over time in different vectors and hosts.

Coupling CAI analysis with *in vivo* and *in vitro* viral titer quantification, and functional genetics could improve the understanding of the phenotypic impact of synonymous changes, such as in the cases of the comparison between the African and Brazilian strains of ZIKV (136). In a similar manner, a possible follow-up study would be to test *in vivo/in vitro* kinetics (see (198) for an example) for endemic and sylvatic strains of DENV2 from the same sampling location. There are many gaps in our understanding of how synonymous and non-synonymous changes lead to functional changes in viral replication, virulence and host adaptation (199). Nevertheless, temporal analyses of CAI can be informative in better identifying time-stamped sequences with non-synonymous changes that could provide an advantage in different hosts. In sum, we hope more studies will identify housekeeping genes for other peridomestic species capable of

transmitting zoonotic viruses, such as flavivirus vectors and *Culex*-associated migratory birds.

3.2 Molecular approaches

3.2.1 A cryptic outbreak of Human Parvovirus B19 hidden by Dengue fever

3.2.1.1 Brief introduction of DENV surveillance, Guarujá, and metagenomic applications

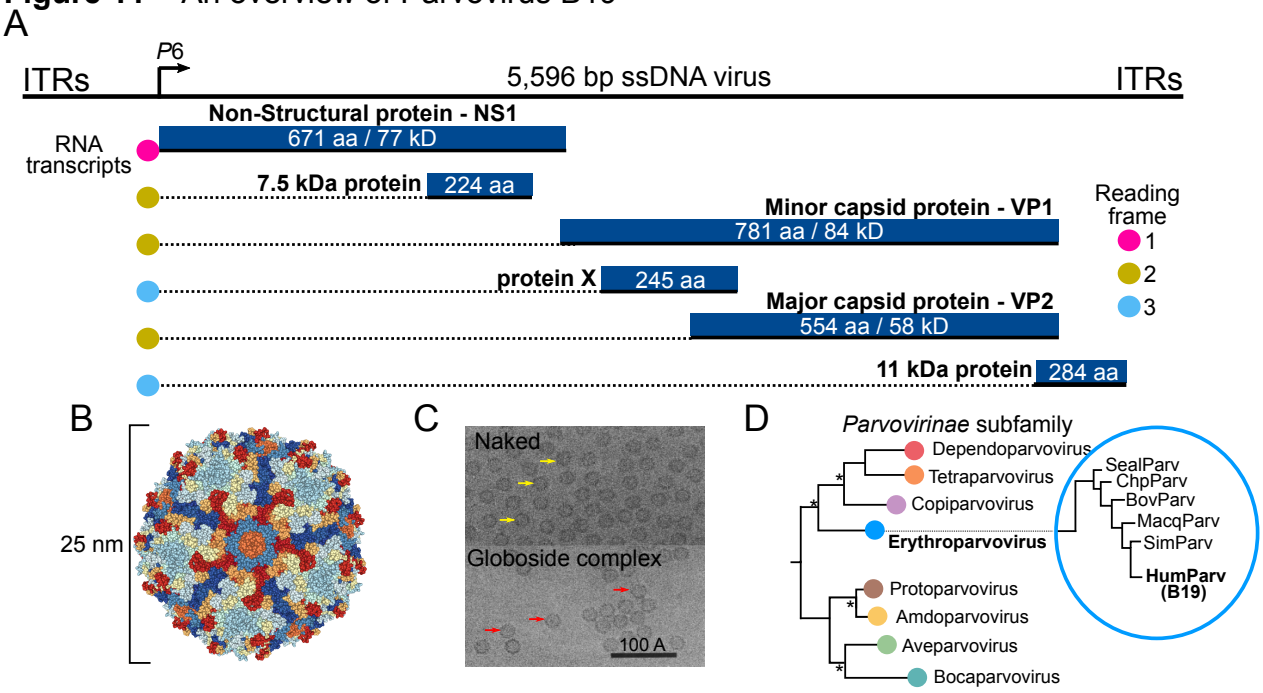
National passive surveillance systems generally rely on weekly clinical reports to administrators from a large network of hospitals and laboratories (200). However, high costs, poor clinical diagnostic accuracy, and general underreporting can limit their effectiveness. As an example, a previous study found that for every 20 dengue patients, only 1 had been reported to Brazil's Notifiable Diseases Information System (SINAN) as having dengue (201). Most Dengue diagnoses are made on clinical and epidemiological grounds where mildly symptomatic and non-specific febrile cases are often discarded (9,201,202).

Few epidemiological studies have investigated or assessed the simultaneous occurrence of different pathogens with similar disease outcome (105,106,203,204). In this case, passive surveillance underestimates Dengue's true rate of infection and ignores the possibility of infection from related viruses *i.e.,* Mayaro, Saint-Louis encephalitis or Zika, but also for distantly related viruses such as Parvovirus B19 (B19V) (203,205,206). B19V infections have been associated with both mild (asymptomatic, Fifth disease, febrile non-specific flu-like illness, rash) and severe (chronic arthralgia, transient aplastic crisis, anemia, hepatitis, myocarditis, and autoimmune syndromes) clinical manifestations (207).

Acute parvovirus B19 infections are characterized by two phases of disease. The first phase of illness consists of a prodromal non-specific febrile illness that can present with fever, myalgia, headache, and chills (208,209), which is a similar disease outcome in Dengue-infected patients (151). However, the Parvovirus B19 disease is usually

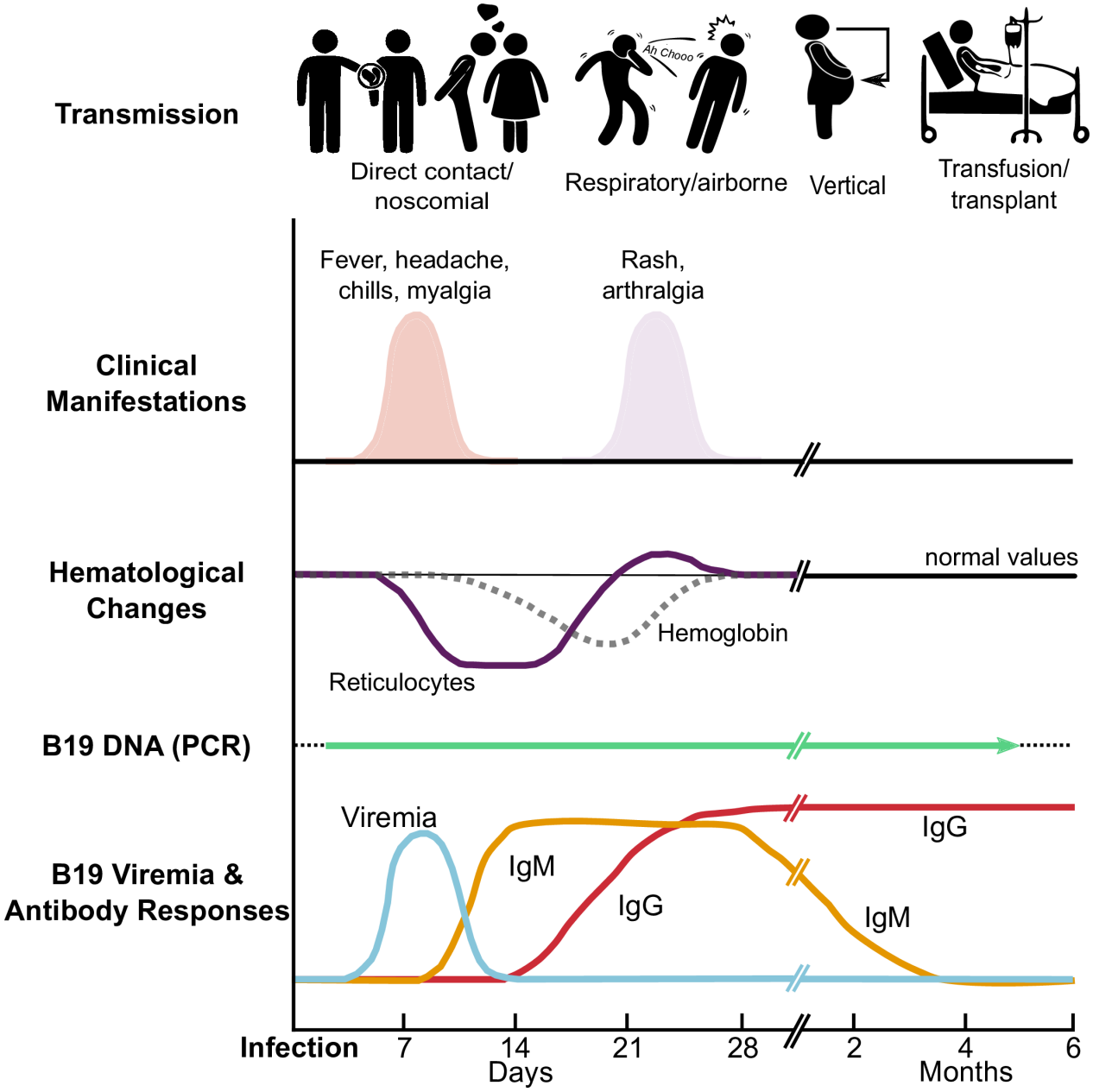
associated with and clinically diagnosed by the second phase of the illness, which may include the hallmark presentations of erythema infectiosum (*i.e.*, Fifth disease), various cutaneous and popular rashes, and arthralgia (joint pain, arthritis) (209). Generally, B19 disease is an asymptomatic or mild pediatric illness with far fewer diagnosed cases in adults; previous childhood infections are thought to provide life-long immunity. However, infections in immunosuppressed individuals and patients with other medical complications can be susceptible to a range of severe disease manifestations and relapsing B19V infections (210,211). The spread and infection of B19V can occur through respiratory secretions, blood-derived products, vertically (from mother to child) and nosocomial transmission (Figures 11 and 12) (212–216).

Figure 11 – An overview of Parvovirus B19



(A) RNA transcription map of B19 with reading frames and resulting protein product. B) A stereoscopic view of a completely assembled B19 capsid complex. C) Cryo-EM images adapted from (217) of empty, frozen and unstained B19 particles (top) and complexed with Globoside (bottom). D) Conjugated maximum likelihood tree of the *Parvoviridae* family and erythroparvovirus genus.

Figure 12 – Clinical and diagnostic considerations of Parvovirus B19



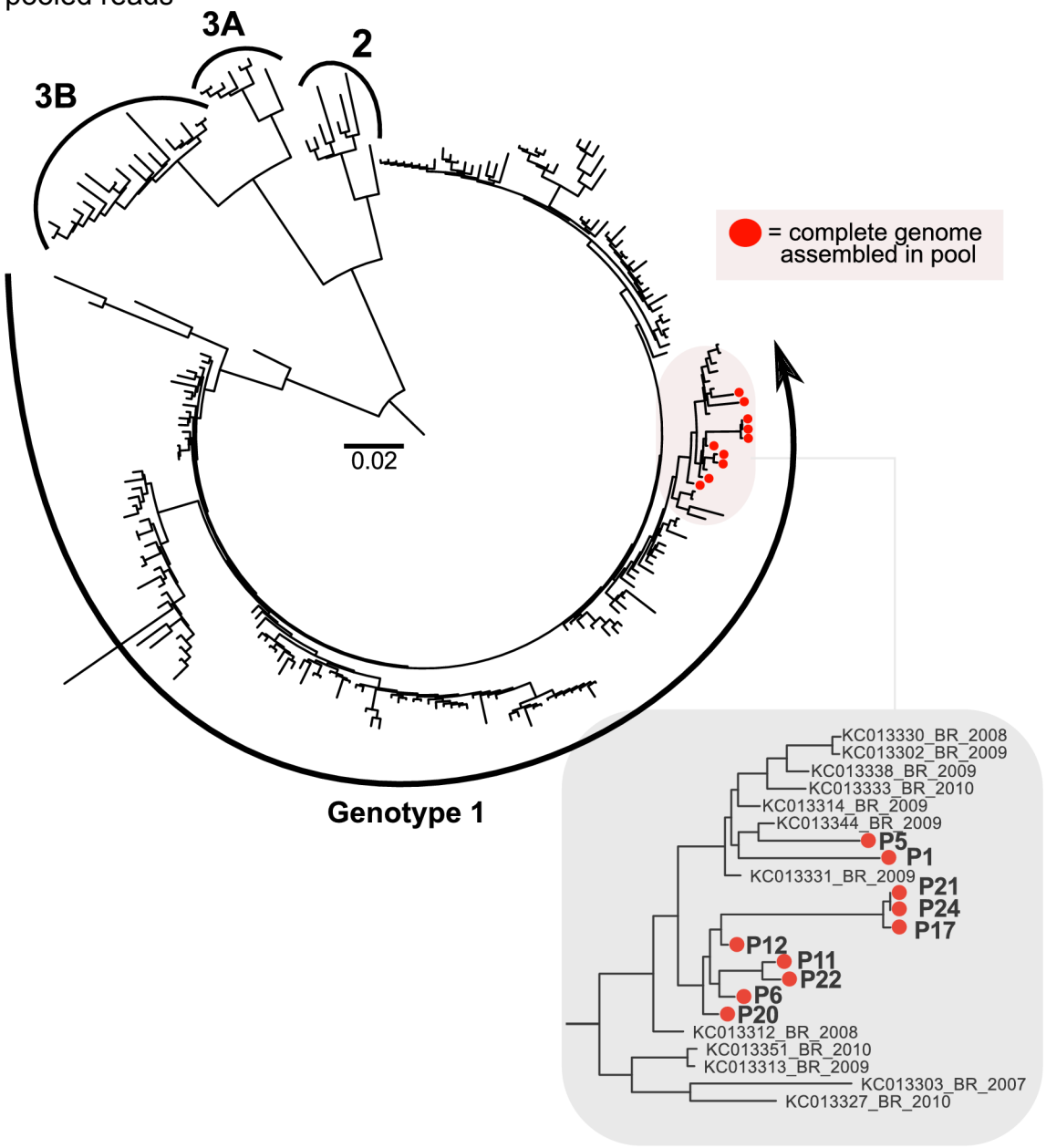
An illustrated overview of Parvovirus B19 transmission, clinical stages of typical manifestations, hematological changes, and presence of viremia and antibodies, all scaled to time. Figure was adapted from (216).

3.2.1.2 Results of Parvovirus B19 cases in Guarujá

From November 2013 to February 2014 during a dengue outbreak, 182 serum samples were collected from patients who visited the city's public health clinics with

symptoms of an acute febrile illness, with some (34.2%) diagnosed with dengue fever but many were undiagnosed. We found that in all these cases, there was no evidence for DENV infection or any other nationally prevalent arboviruses such as Zika virus, Chikungunya virus, Mayaro virus or Yellow Fever virus. From here, samples were grouped into 24 pools and were subjected to total RNA sequencing to screen for other viruses with a similar disease outcome to Dengue. Using reads from individual pools, we were able to *de novo* assemble a complete or nearly complete Parvovirus B19 genome from 10 of the 24 pools (Figure 13). Albeit at a lower frequency, we also found sequences from GB virus C, Hepatitis C virus, Hepatitis A virus, and many different groups of phages and retroviruses.

Figure 13 – Phylogenetic estimation of B19V complete genomes assembled from pooled reads



A radial maximum-likelihood tree with Parvovirus B19 genotypes labeled. Red circles at tree tips indicate a complete genome assembled in the respective pool. A zoom-in (gray box) shows that sequences are closely related to other genotype 1 isolates collected in Brazil.

Once we identified B19V we designed specific primers for the VP1/VP2 region of B19V. Of the 182 patients, 115 (63.2%) were PCR positive for the B19 virus (Table 1). The median age of these patients was 27. In fact, nearly 60% of the patients were between 18 and 39 years of age.

Table 1 – The demographics and diagnostics of Parvovirus B19 in Guarujá

Demographics of patients with B19V DNA^	**Total**
Female sex — no. (%)	66/111 (59)
Median age (range) — yr	27 (6−75)
5−17 yr (%)	19/106 (17.9)
18−39 yr (%)	61/106 (57.5)
≥40 yr (%)	26/106 (24.5)
Laboratory testing[&] — no./total (%)	
PCR-positive	115/182 (63.2)
PCR-positive and IgM-positive, IgG-negative	8/182 (4.4)
PCR-positive and IgG-positive, IgM-negative	47/182 (25.8)
PCR-positive alone	50/182 (27.5)
IgM-positive alone	4/182 (2.2)
IgM-positive and IgG-positive, PCR-negative	0/182 (0)
IgG-positive alone	25/182 (13.7)
PCR-negative, IgM-negative and IgG-negative	23/182 (12.6)
Clinical Diagnosis* — no./total (%)	
DENV NS1-specific ELISA test requested	46/70 (65.7)
Differential diagnosis of dengue	24/70 (34.2)

^Four samples had no gender data, nine had no demographic data
[&]Borderline antibody reaction results were abstained from results
*Medical records for 70 patients were available. However, most records either were incomplete or illegible for further analysis.

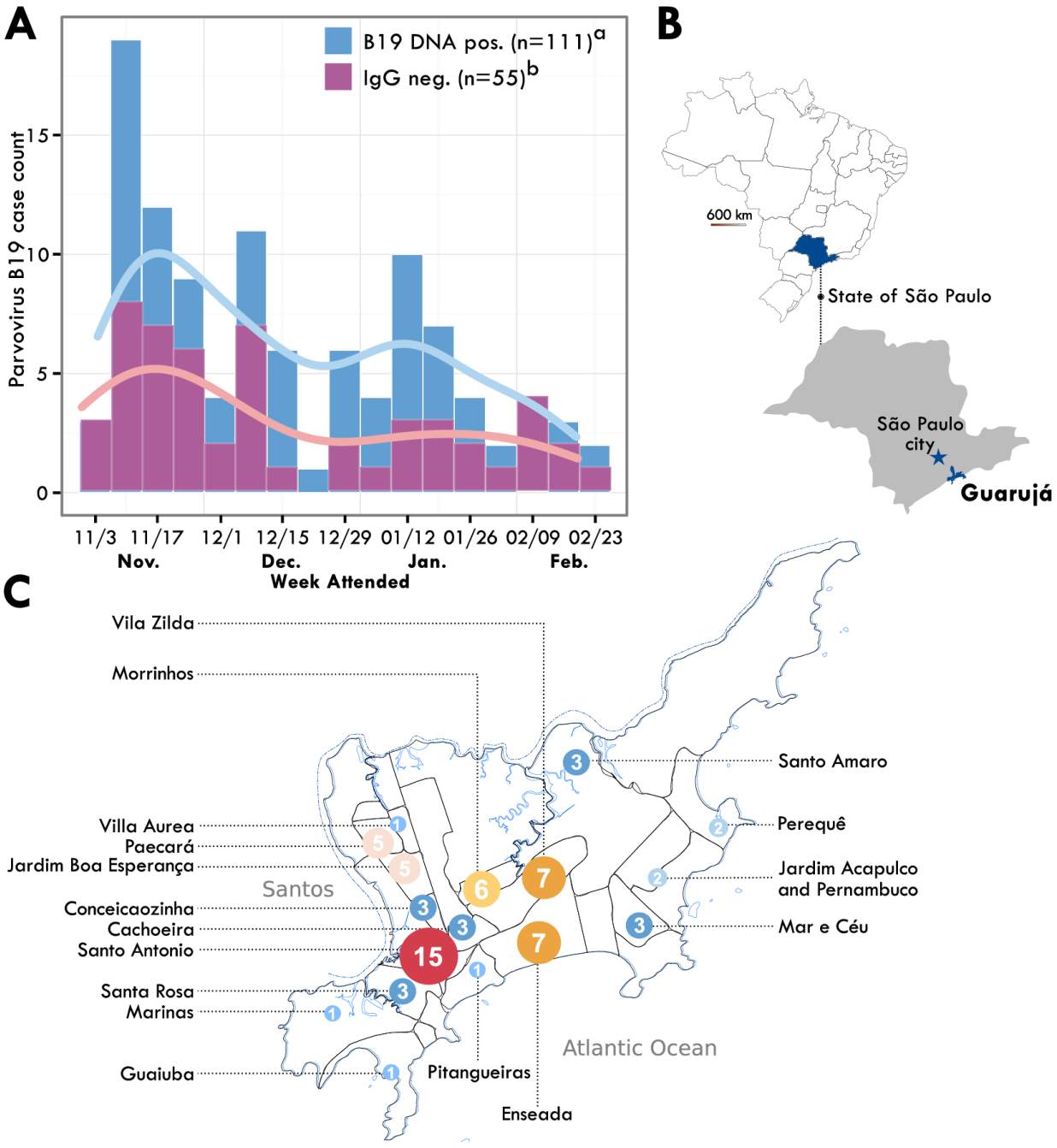
From the patients where DNA was amplified by PCR, eight were also positive for the B19V IgM antibody. Only four samples were PCR negative and IgM positive. A total of 50 patients (27.5%) with B19V DNA were negative for both B19V IgM and IgG, suggesting a first time exposure. Additionally, 47 (25.8%) patients were PCR positive, IgM negative and IgG positive, which is a serological profile that has been associated to a prior, persistent or a reoccurring infection (218–221). Only 23 patients (12.6%) were negative for all three diagnostic assays for B19V.

As contaminations have been reported as a problem for B19V diagnostics, we took precautions to ensure the reliability of our results. For each sample, PCR assays

were performed in duplicate and on separate days to avoid false-positive DNA amplifications. Prior to and upon completion of both assays, we sterilized all materials and used new reagents as well as PCR specific negative controls for each PCR run and associated gel. We also tested 35 serological samples of febrile patients from another Brazilian state, Sergipe, that were suspected of Chikungunya (222), but were found negative for both B19V PCR and IgM antibodies.

We then retrieved the patient's date of attendance at a public clinic from 111 of the 115 patients positive for B19V DNA. Cases ranged from November 4th, 2013 to February 25th, 2014. We mapped the B19V positive patient's date of attendance by week to discover whether was the time where case incidence was highest (Figure 14A). Peak incidence occurred during the week of November 10th, 2013. We also obtained the addresses for 68 of the PCR positive patients and mapped the distribution of cases around the city (Figure 14C). We found that the majority of cases were in the Santo Antônio district where the main administrative, mass transportation and commercial facilities are located. Highly populated districts as Morrinhos (6 cases), Enseada (7 cases) and Vila Zilda (7 cases) also had a large proportion of localized cases.

Figure 14 – The 2013 to 2014 temporal and geographic spread of Parvovirus B19 cases in the city of Guarujá.



A) Weekly sampling of serum samples where Parvovirus B19 DNA was detected (royal blue) and the proportion of samples that were first-time infections (purple). B) The geographical location of the coastal city of Guarujá. C) The geographical distribution of Parvovirus B19 cases in Guarujá.

For 70 patients, we were able to retrieve a more detailed medical record with a range of incomplete information including symptoms, hematology, complications,

diagnoses, and drug recommendations. Several medical records were discarded because they were either lacking data or were completely illegible. However, all patients included in our study presented one or more dengue-like symptom(s), but not all medical records included patient's symptoms during their attendance (see footnote of Table 1). Nevertheless, we annotated the symptoms and complications presented from 56 patients who were also viremic for B19V during their clinical attendance (Table 2 and Table 3). Fever (70%), myalgia (61%) and headaches (43%) were the most prevalent clinical presentations in patients. Only 12 of these patients (21%) had presentations of arthralgia (6 patients) or rash (6 patients). Eight patients (14.3%) described retro-orbital pain at their attendance, a symptom to the best of our knowledge that has no current association with acute or persistent parvovirus B19 infection.

We also obtained data from patients with complications such as vomiting and diarrhea. Additionally, 25 medical records included patient platelet count information that helped us determine the prevalence of thrombocytopenia (platelets < 150,000/mm^3) in viremic patients. In total, 15 patients (60%) presented with thrombocytopenia during their clinical attendance.

Because people can be persistently infected by B19V, the sole existence of B19V DNA in a patient's serum can be misleading for a clinical diagnosis of an acute infection without any supporting clinical information, especially if paired sampling or a quantitative PCR or ELISA analysis were not done (223–226). The presence of the B19V IgM antibody is a strong indicator of an acute infection (227,228). Therefore, we looked to see if these patients presented different symptoms in relation to their serological profile (Table 2) by analyzing the occurrence of patient symptoms and clinical complications with their serology. Only two patients with a detailed medical record were IgM positive. Nevertheless, 22 patients (with medical records) were IgM negative and IgG positive, which could indicate a persistent or reoccurring infection (227,229,230). The other 24 patients (with medical records) were negative for both antibodies, suggesting a first-time and recent infection. Given that 50.4% of our (PCR only or PCR and IgM positive) patients appeared to be acutely infected it is fair to assume that we sampled from an ongoing B19V outbreak. We did not find any significant differences between serological profiles and clinical presentations.

Table 2 – Serology and clinical presentation in suspected Parvovirus B19 patients

		IgM+/IgG+	IgM-/IgG+			IgM-/IgG-		
Symptoms	**No./total no. (%)**	Total	M	F	Total	M	F	Total
Fever	39/56 (69.6)	1 (50)	7 (70)	9 (75)	16 (73)	9 (69)	9 (82)	18 (75)
Arthralgia	6/56 (10.7)	0	2 (20)	0	2 (9)	1 (8)	2 (18)	3 (13)
Headache	24/56 (42.9)	0	4 (40)	5 (42)	9 (41)	5 (38)	7 (64)	12 (50)
Generalized Rash	6/56 (10.7)	1 (50)	2 (20)	0	2 (9)	0	1 (9)	1 (4)
Retro-orbital pain	8/56 (14.3)	0	1 (10)	3 (25)	4 (18)	3 (23)	0	3 (13)
Myalgia	34/56 (60.7)	1 (50)	8 (80)	7 (58)	15 (68)	6 (46)	8 (73)	14 (58)
Odynophagia	2/56 (3.6)	0	0	0	0	0	1 (9)	1 (4)
Complications								
Vomiting	6/56 (8.9)	0	1 (10)	0	1 (4)	2 (15)	2 (18)	4 (17)
Diarrhea	5/56 (9.3)	0	2 (20)	0	2 (9)	2 (15)	0	2 (8)
Thrombocytopenia	15/25 (60.0)	1 (50)	1/1 (100)	2/5 (40)	3/6 (50)	2/4 (50)	2/5 (40)	4/9 (44)
	Total	2	10	12	22	13	11	24

Next we compared clinical presentations by the age and sex of the patient (Table 3). We grouped patients by those who were 5-17 years of age (children) and those who were over 18 (adults) at the time of their attendance. Notably, we found that the prevalence of myalgia was significantly higher in adult patients compared to younger patients (Fisher's exact test, *p*-value = 0.02). Likewise, we also observed that only adults presented with symptoms of arthralgia, odynophagia (difficulty swallowing), and diarrhea. In disagreement to some previous works (231,232), we did not see any significant differences in symptoms between men and women. However, we did find that only men reported having diarrhea at their clinical attendance (Fisher's exact test, *p*-value = 0.02).

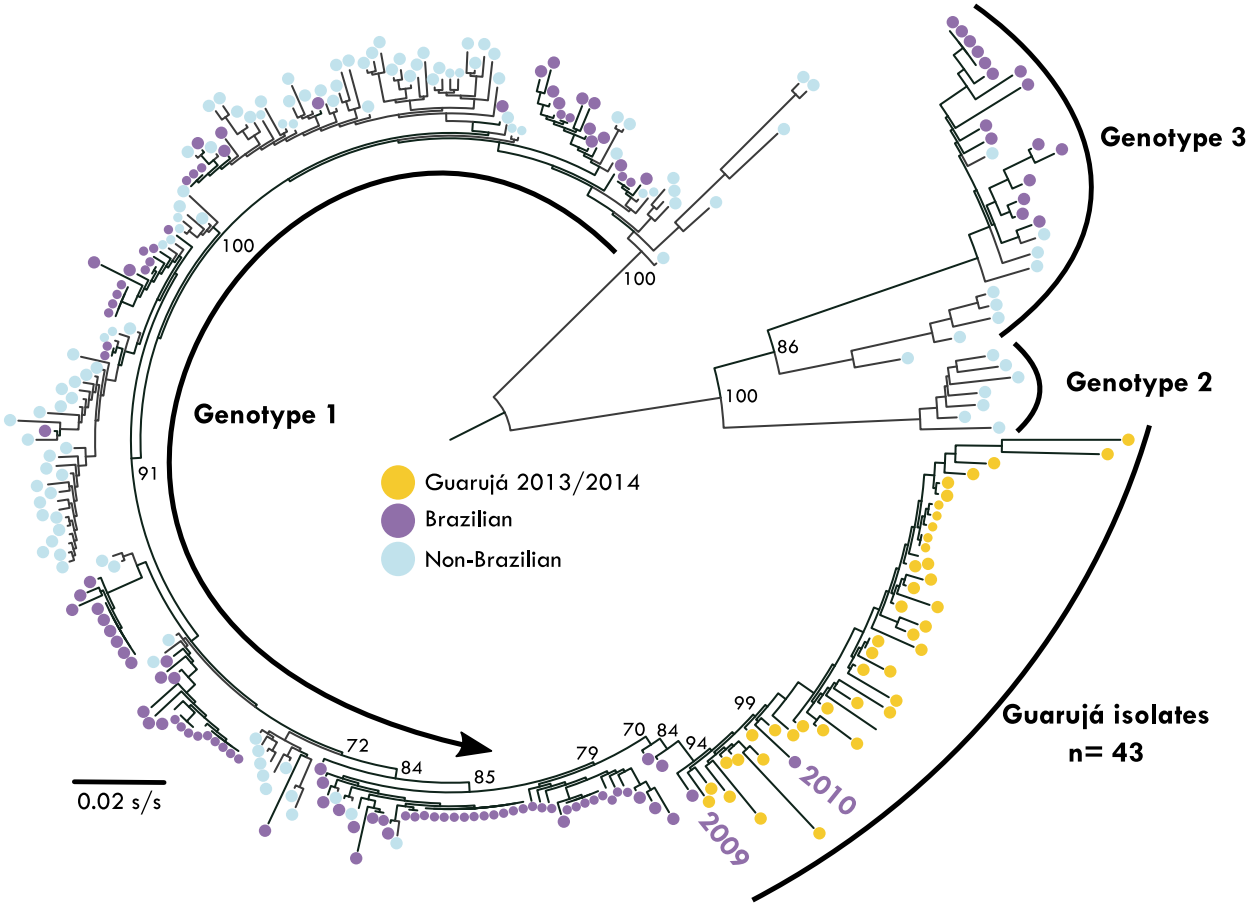
Table 3 – Age and sex of patients and their clinical presentations

		Age			Sex		
Symptoms	**No./total no. (%)**	5-17	>18	*p*-value^	M	F	*p*-value^
Fever	39/56 (69.6)	11/13 (84)	28/43 (65)	0.3029	19/27 (68)	20/28 (71)	1
Arthralgia	6/56 (10.7)	0/13 (0)	6/43 (14)	0.3181	4/27 (14)	2/28 (7)	0.6687
Headache	24/56 (42.9)	5/13 (38)	19/43 (44)	0.7603	12/27 (43)	12/28 (43)	1
Generalized Rash	6/56 (10.7)	3/13 (23)	3/43 (7)	0.13	2/27 (7)	4/28 (14)	0.6695
Retro-orbital pain	8/56 (14.3)	1/13 (8)	7/43 (16)	0.6652	5/27 (18)	3/28 (11)	0.4688
Myalgia	34/56 (60.7)	4/13 (31)	30/43 (70)	0.02135	17/27 (71)	17/28 (61)	1
Odynophagia	2/56 (3.6)	0/13 (0)	2/43 (5)	1	1/27* (4)	1/28 (4)	1
Complications							
Vomiting	6/56 (8.9)	2/13 (15)	4/43 (9)	0.6286	4/27 (14)	2/28 (7)	0.4216
Diarrhea	5/56 (9.3)	0/13 (0)	5/43 (12)	0.58	5/27 (18)	0/28 (0)	0.02321
Thrombocytopenia	15/25 (60.0)	5/9 (55)	10/16 (63)	1	8/11 (73)	7/14 (50)	0.4139

*No serological data was available for this patient
^Statistical inferences were performed using Fisher's exact test

To explore the genetic diversity of B19V in Guarujá, we amplified a small region (sequences ranged from 427 to 560 nucleotides in length) of the VP1/VP2 gene that has been used for genotyping in other works (233,234). We then aligned our sequences (Genbank accession numbers: MG765329-MG765371) with all publically available B19V sequences with the country of isolation and collection date available from Genbank. Our phylogenetic estimation showed that our sequences (orange tips) are most similar to the B19V Genotype 1 sequences (Figure 15). It is possible that there was more than one emergence of B19V in Guarujá. Our sequences were genetically similar to two sequenced isolates from the city of São Paulo: one from 2010 (Genbank accession: KC013305) and one from 2009 (Genbank accession: KC013344). We also indicated (violet tips) all published Brazilian and non-Brazilian sequences (light blue tips) of B19V, which are widely spread across Genotype 1 and Genotype 3B. Genotype 2 sequences from Brazil are only available for the NS1 gene.

Figure 15 – Phylogenetic estimation of the VP1/VP2 Parvovirus B19 gene fragment using a maximum-likelihood tree.



The Shimodaira-Hasegawa bootstrap values greater than 70% are shown at major nodes. Nodes were placed and colored at tree trips to indicate the origin of sequences. Scale bar (s/s) is represented by substitutions per site.

3.2.1.3 Discussion of Parvovirus B19 cases in Guarujá

In this study, we retrospectively studied dengue-suspected cases and found a large prevalence (63.2%) of B19V in patient's serum. Despite several seroprevalence studies (235,236) and some sporadic cases in the state of Rio de Janeiro (237), the public attention for Parvovirus B19 in Brazil is limited. To our knowledge, B19V is rarely tested for or even diagnosed in public health centers. Moreover, its febrile prodromal disease (*i.e.*, pre-rash and pre-arthritis symptoms) is often mistaken for more prevalent pathogens such as dengue.

Dengue infections occur all year round in Brazil, but explosive outbreaks tend to happen during the summer months of November to April (238–240). A high number of dengue cases lead to clinical diagnoses based on epidemiological grounds, which we showed here can lead to cryptic outbreaks of pathogens with similar disease outcomes. We found that during the 2013/2014 B19V outbreak in Guarujá, infections peaked in the late spring and early summer months of November through January (Figure 14), which agreed with other reports of erythema infectiosum outbreaks (228,232,241–243), and occurred in a similar time frame of locally reported dengue outbreaks (239,240). Furthermore, clinical and hematological manifestations of B19V infected patients were also found to be similar to classical dengue disease presentations like fever, headache, myalgia, and retro-orbital pain (151). This is especially true for the eight patients who presented with retro-orbital pain—a symptom that has no known association with the Parvovirus B19 disease—and with thrombocytopenia, a common feature and diagnostic marker of dengue.

Since Parvovirus B19's discovery, diagnostic methods for detecting acute infections in serum have involved electron microscopy, direct DNA hybridization, antigen enzyme-linked immunosorbent assays (ELISA) *i.e.* detection of IgM, IgG, IgA, and NS1-specific antibodies, PCR and hemagglutination (229,243–253). Detection of B19V DNA in serum is circumstantial evidence for an infection, but cannot alone determine whether the viremia in question is acute or persistent because of its prolonged presence after infection (227,254). The lack of any IgM or IgG antibodies in 50 symptomatic cases could be a strong indication for a first time infection, which is surprising since the majority of patients were adults and B19V exposure typically happens during childhood (255). Moreover, we only detected IgM antibodies in 12 cases, which is much less than the number of PCR positive cases. In these cases, it is possible that the majority of patients were attended very early in their illness and before their IgM antibody response could be detected (209,216).

We also found 47 symptomatic patients who were PCR and IgG positive, but IgM negative. We are aware of four explanations that could account for this: *i*) an acute infection where the patient could not elicit an IgM response, *ii*) a prolonged acute infection that lingered after the disappearance of IgM antibodies and was not completely

neutralized by IgG (209), *iii*) a persistent infection where a secondary viral reactivation or relapsing infection led to secondary clinical manifestations (218–221), and *iv*) the presence of low levels of B19V DNA from a previous infection that have no correlation to the current clinical manifestations (227). Even with clinical records for some patients, we were uncertain how to determine which of the above-mentioned possibilities were most likely.

None of the patients in this study were pregnant or younger than 5 years of age. However, it is of high concern that B19V infections in pregnant women are linked to hydrops fetalis (*i.e.*, fluid collects in two or more fetal compartments), miscarriages, microcephaly, hydrocephalus, contractures, and intracranial and hepatic calcification (214,256). Nevertheless, few studies have investigated the prevalence of B19V in babies with birth defects suspected of ZIKV congenital syndrome in Brazil (257,258). This is especially worrisome as congenital infections of ZIKV and B19V can both cause hydrops fetalis and microcephaly (214,256,259,260). Furthermore, Brazil harbors all three genotypes of *Erythroparvovirus*, as well as recombinant variants, attesting to the large endemic diversity within the country and abundant case studies of severe complications associated with the virus (Figure 15) (233,234,261).

Next generation sequencing approaches have led to better awareness and detection of several pathogens; its success in accomplishing this has been demonstrated over the last decade (203,262–264). In this case, it allowed us to discover what underlying etiological agents were behind a large number of sick patients. This outbreak served as an example that current passive surveillance measures could be vulnerable to missing a mild pathogen like B19V, which in turn, also raises concerns that we could be overlooking more severe pathogens too. Here, we found high amounts of B19V in dengue-suspected patients where no diagnosis could be previously confirmed. Significantly, we found that despite a negative DENV diagnosis, 34% of patients with medical records were still clinical diagnosed for dengue (Table 1). In one case, an official case notification for dengue without any diagnostic evidence was found on the patient's medical record. This further illustrates that even known viruses could circulate camouflaged underneath diagnostic detection. Although relatively expensive, we advocate for the situational use of NGS for symptomatic but negatively diagnosed

patients during large seasonal outbreaks of dengue or other febrile illnesses. Our findings provide an example that a medically important virus like Parvovirus B19 can be neglected by passive surveillance systems during a dengue outbreak. Crucially, it demands greater attention in surveillance systems to viruses showing similar clinical presentations.

3.2.2 Sequencing and characterization of Yellow Fever virus isolated from a convalescent patient's urine

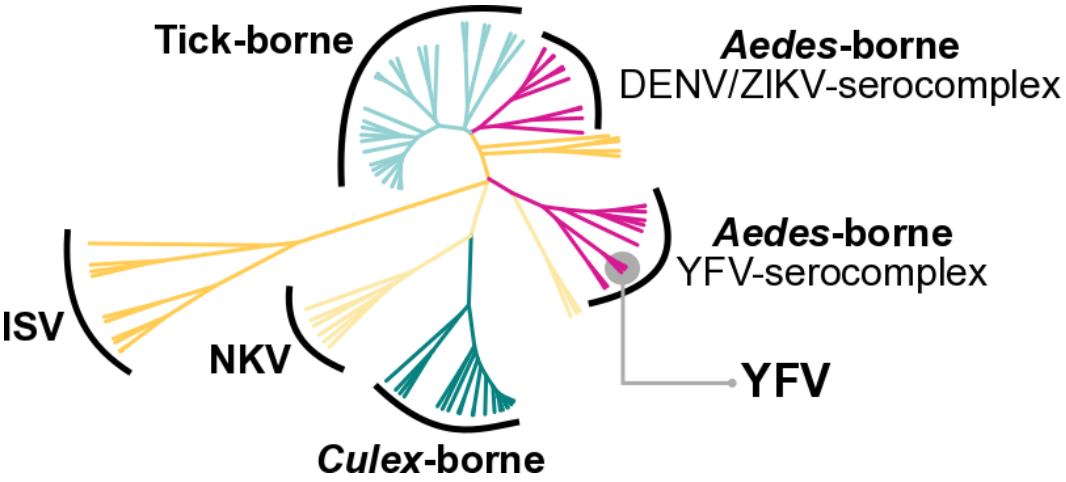
The following section contains material that was published in Emerging Infectious Diseases on January 1, 2018. (265)

3.2.2.1 Brief Introduction of Yellow Fever Virus

The Yellow fever virus (YFV) is a single-stranded positive-sense RNA virus that is classified as a member of the *Flavivirus* genus (Figure 16) in the *Flaviviridae* family (266). YFV populations are maintained in two replication cycles: a sylvatic cycle between non-human primates and *Culicidae* mosquitoes, and an urban endemic cycle where transmission alternates between humans and *Aedes aegypti* mosquitoes (267).

Despite the availability of a vaccine, YFV cases are estimated to exceed over 200,000 cases annually, with greater than 30,000 case-fatalities per year (153). In early 2017, a large outbreak of Yellow Fever occurred in Brazil (Figure 17) with 792 confirmed cases and 274 case fatalities (35% case fatality rate) (268).

Figure 16 – The genus Flavivirus and Yellow Fever virus

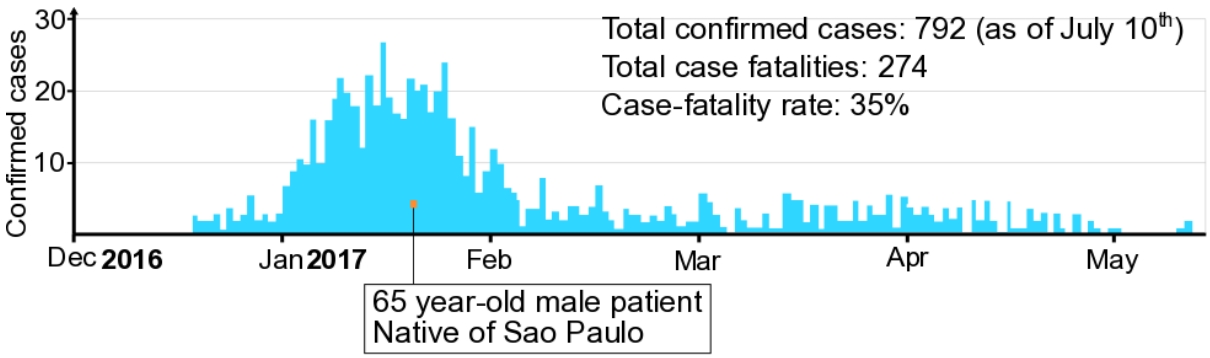


A radial Bayesian-estimated phylogenetic tree of the subgroups of the genus Flavivirus colored by their predominant vector host genus. Two subgroups of species utilize the Aedes mosquito vector to infect mammalian species. Yellow Fever virus (YFV) is labeled. The tree branches are not too scale.

YFV can cause Yellow Fever disease in humans, which is characterized by fever, prostration, hepatic, renal and myocardial injury that can lead to death in 20-50% of cases (269). Clinical confirmation of YFV infections is based on the detection of viral RNA in serum by reverse transcription PCR or antigen-based Enzyme-linked immunosorbent assays (ELISA).

Recently, viral diagnostics using urine has been used for confirming flavivirus infection, such as West Nile virus (270), Zika virus (271) and Dengue virus (272). In these studies, the genomes of the three viruses were detected for longer periods of time in urine than in serum. In the case of WNV and ZIKV, recovered viral particles from urine were still infectious (270,273).

Figure 17 – Yellow Fever virus outbreak in Brazil, 2017



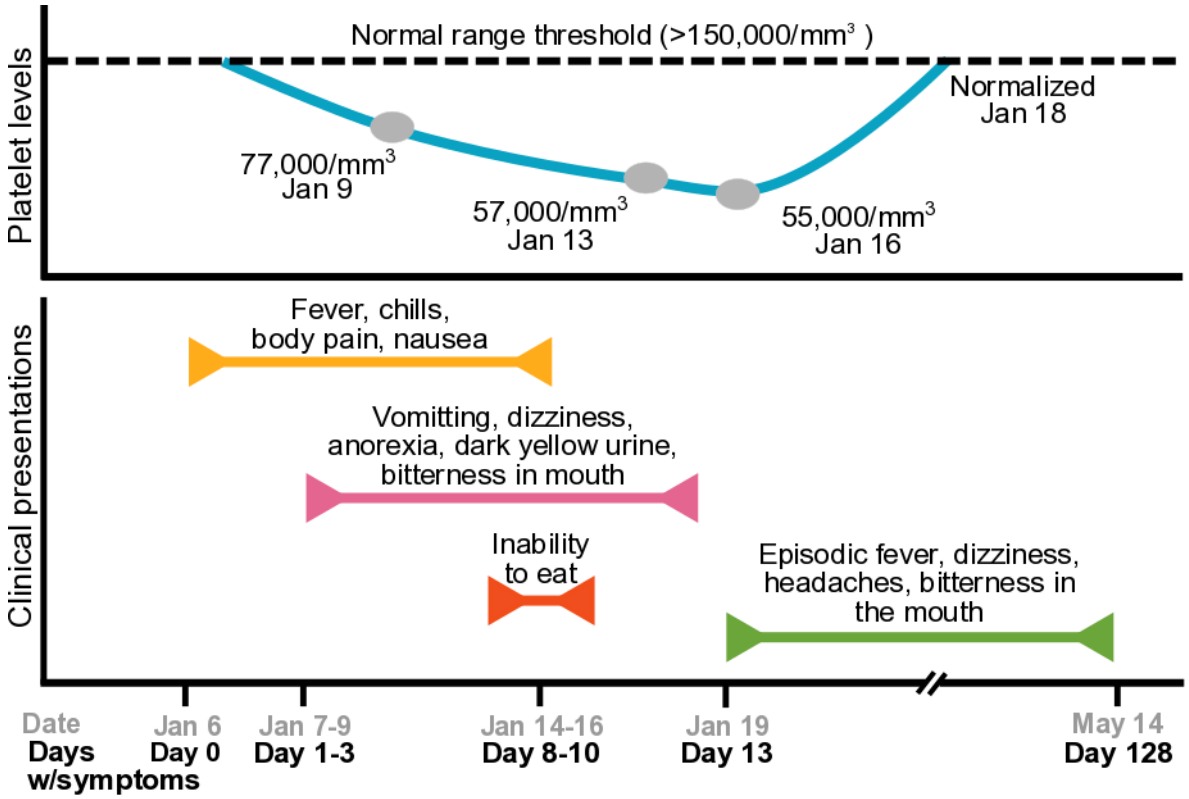
The distribution of confirmed cases that were notified to the Brazilian MS/SVS since May 31st, 2017, adapted from (268). The patient (orange square) from this study was reported on Jan 19th, 2017.

3.2.2.2 Brief Introduction of a clinical case of yellow fever

Here, we report the case of a 65 year-old male who is a native of Diadema, São Paulo (Figures 18 and 19). On December 28th, 2016, the patient traveled to Januária city, which resides in the neighboring state of Minas Gerais, north of São Paulo. On January 6th, 2017, he became ill showing general symptoms such as fever, myalgia and nausea. As his condition worsened until the 9th of January, he checked himself into the local hospital. A NS1-antigen ELISA assay for Dengue virus was negative. On January 13th, he returned to São Paulo city and visited a public hospital in Diadema, where a second Dengue NS1-antigen ELISA assay was found negative.

From January 14th to the 16th, severe symptoms persisted in the patient; his urine was a dark yellow color and was feeling very weak. The patient was no longer able to eat solid food during this 3 day period. On January 18th, the normalization of his platelet count was observed on January 18th. The patient's general condition began to improve. From January 19th to May 14th, the patient reported having episodes of fever, dizziness, headaches and bitterness in the mouth, especially during physical excursion. He also described a loss of taste.

Figure 18 – Clinical Presentation of the patient diagnosed with Yellow Fever

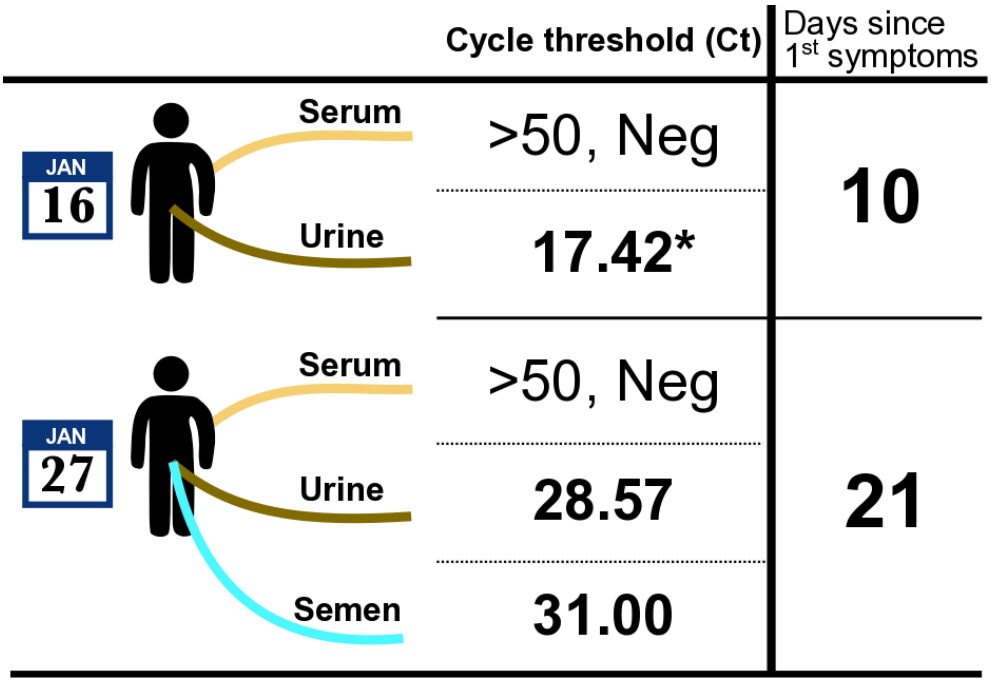


Over the course of his illness, the patient experienced severe thrombocytopenia and different phases of YFV symptoms. The patient presented YF-related symptoms 128 days after his initial presentation.

3.2.2.3 Sequencing and Phylogenetic Results

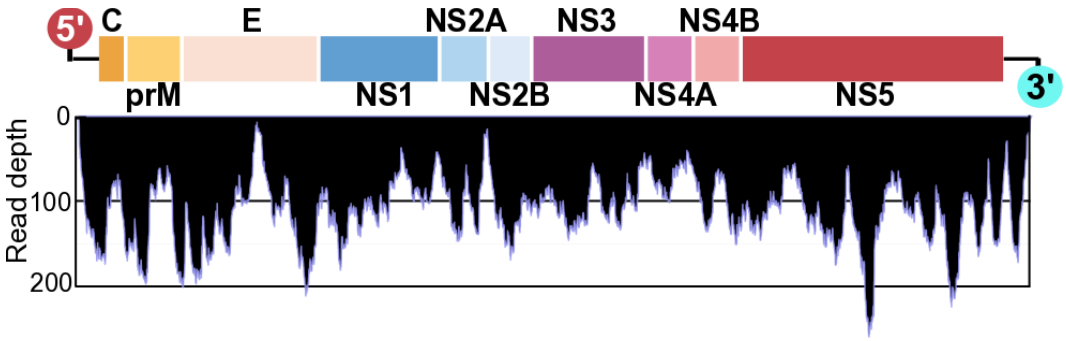
On January 16th,and January 27th, samples were collected and tested for YFV using a qRT-PCR assay. RNA from the urine sample was extracted and sequenced to obtain a complete YFV genome (Figure 20).

Figure 19 – Diagnostic summary of the Yellow Fever case



Clinical samples were collected from the patient 11 days apart. A qRT-PCR assay testing for YFV was performed for all samples. YFV infection was confirmed by a qualitative assay carried out with capture IgM ELISA using a specific viral antigen. *Complete genome was sequenced from this isolate.

Figure 20 – Genome mapping of the sequence obtained from the patient on January 17th, 2017

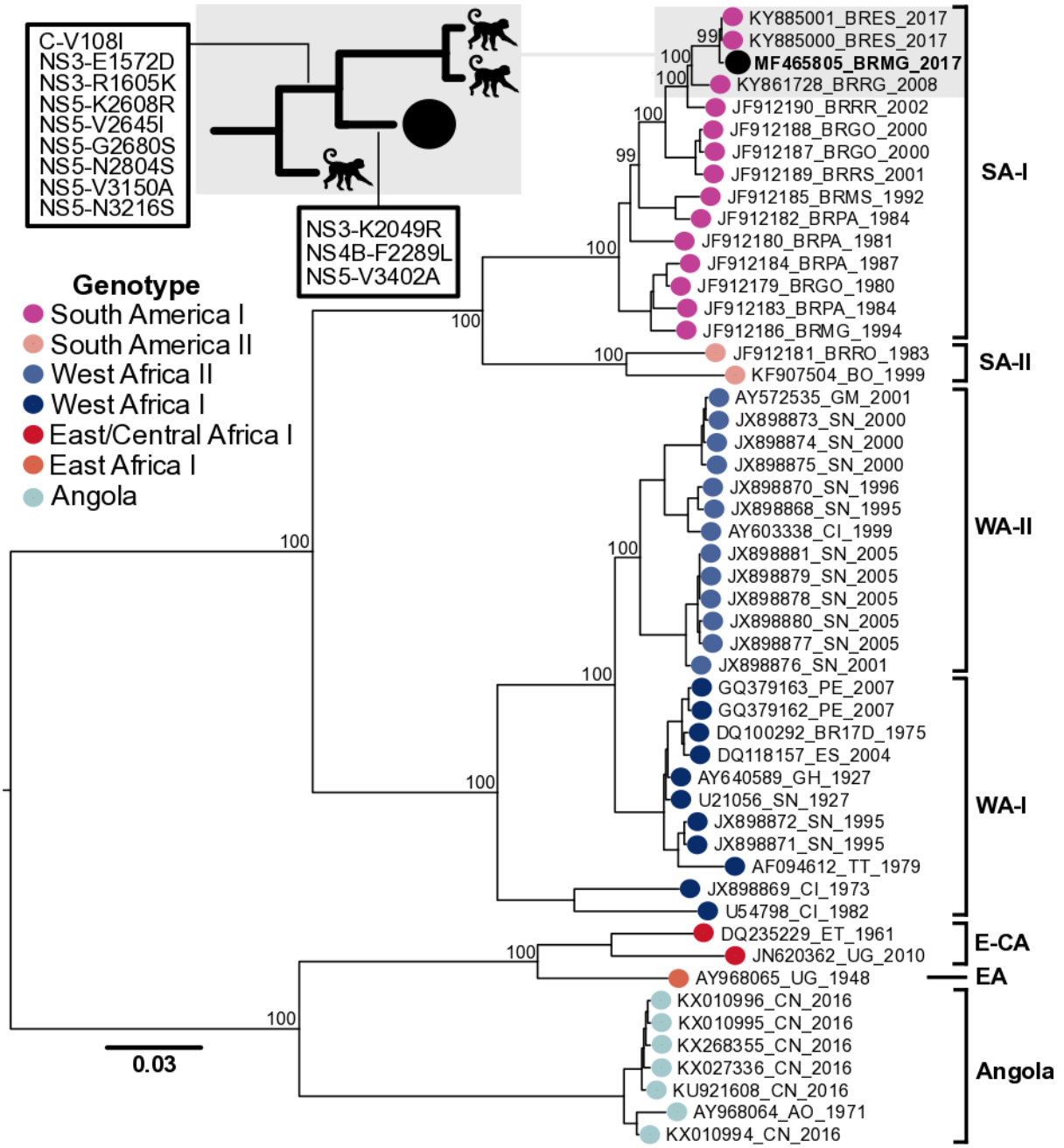


Complete genome map of the Yellow Fever virus. The consensus sequence had a length of 10,948 with 112.6X average depth and a 99.5% breadth of coverage of the reference genome (KY885001), with bases lacking only at the extremes of the untranslated region.

3.2.2.4 Discussion and Implications for Future Yellow Fever virus diagnostics

The sequenced Brazilian isolate from Minas Gerais clustered with the South America I isolates, including other two sequences isolated in Espírito Santo state in 2017, a neighboring state of Minas Gerais, and others sequences isolated in previous years in Brazil (Figure 21). No insertions or deletions were observed in the Brazil-2017 nucleotide sequences compared with the others South America I strains. We found several synapomorphic changes in our sequence, most of them located in the NS5 protein (RNA-dependent RNA polymerase) that has important activity in viral replication.

Figure 21 – Maximum-likelihood estimate of the January 17th Yellow Fever virus isolate



Maximum-likelihood phylogenetic tree rooted at midpoint based on full length genomes of YFV showing major YFV genotypes. Percent bootstrap values after 10000 replicates for major branches are indicated near nodes. Sequences used are labeled as follows: accession number/country/year of isolation. Most parsimonious reconstructions (MPR) of synapomorphic changes detected in the NS3, NS4B and NS5 of our isolate (shown as a black circle) and those of the monkey-derived sister taxa are shown in the top left insert.

We are currently testing the infectiousness of the YFV isolated from the patient's semen and urine. If the YFV particles from the semen are indeed infectious, the public health implications could be considerable. We hope to explore differences between YFV isolates and further improve diagnostics for future outbreaks.

CHAPTER 4 – DISCUSSION

Viruses know no borders. It is a spinoff of a more popular quote but the lesson is the same. Emerging and re-emerging viral threats never cease to catch us unprepared. As of late, several examples come to mind: Dengue in 2014, Chikungunya in 2015, Zika in 2016, and Yellow Fever in 2017 (268,274–276). These are just in Brazil, as countless examples exist overseas.

Brazil is experiencing a rural–urban migration that ultimately fuels haphazard expansion of urban areas; there have also been landscape changes in the Amazonian and massive recent increases in cropland in the Cerrado (tropical savanna) (277). Given these changes, disease may turn out to be an environmental issue: in changing environments, novel pathogens are more likely to spill over into human populations (48,49,51). It was previously shown that inadequate management of the natural ecosystem and expansion of cities without proper urban-environmental planning influence the incidence of disease (278). This concern is reflected in a recent work reinforcing the theory that human activities such as selective logging, construction of new roads, and poor forest and landscape management are risk factors for malaria and other infectious diseases in the Brazilian Amazon (279). As the propensity for emerging viruses augments with increasing population densities, global mobility, identifying and characterizing emerging threats before they reach critical levels is of the utmost importance.

Pure *in silico* approaches can be used to identify and characterize emerging viruses without any costs or sample restrictions, making them highly accessible and economical. As others and I have shown, discoveries of divergent viral species can be made using creative algorithms and publically available sequencing data (280–282). In this work, the RdRp gene of a novel dicistrovirus was identified from a published sewage metagenomic study (115). Although we only found a partial sequence of the virus, it can lead to a variety of follow-up analyses: re-sequencing the original sample, searching for orphan domains, and codon bias analysis. All are forms are characterization once a target virus has been identified.

Characterizing viruses using *in silico* methods are improving with time, especially with the growing availability of powerful operating systems, software, and published virus sequences with attached experimental data. As an example, codon usage bias could be useful in explaining differences in replication dynamics for viruses in different hosts (137). In this work, it was explored thoroughly across the flavivirus genus with a few specific examples *i.e.*, Dengue virus, Yellow Fever virus and West Nile virus. The work has a disadvantage as it has no attached *in vitro* or *in vivo* data. However, theoretical works can lead to experiments that can test data and theories developed *in silico*.

There is a constant balance of *a priori* justification and empirical data that drives research. In this case, both approaches contributed to the discovery of a Parvovirus B19 outbreak hidden by dengue fever. Patients who were undiagnosed but presented with symptoms of dengue were expected to be sick from some infectious agent. As an example, dengue-suspected patients in Nicaragua were discovered to be infected with viruses from over 6 different viral families (203), which served as a good proof-of-concept to make a similar investigation here. The generation of complete B19V genomes and similar studies in different geographical regions could improve B19V surveillance and help us better understand the phylodynamics and the epidemiology of the infectious agent.

It has been shown that a large number of clinical cases can be simultaneously characterized using a metagenomics approach. The same method was applied to a single patient. For the first time, an adult male from Minas Gerais was diagnosed with Yellow fever from his urine and semen. Then, the complete YFV genome was sequenced from his urine. Using this data, unique changes were mapped across the genome and compared to sequences from the ongoing and past isolates from Brazil. Metagenomic sequencing has characterized viruses during epidemics and facilitated the development of new diagnostic tools (264,283,284). Future works could further investigate phylogenetic and replicative differences between YFV in serum, urine and semen.

In summary, there is a necessity for more active surveillance in an increasingly globalized world. As shown, both molecular and bioinformatics approaches can lead to the discovery and characterization of emerging and re-emerging viruses. The recent

outbreaks in Brazil have tested their capabilities to respond to emerging threats. Threats will always continue to emerge and our vigilance should continue to grow in turn.

CHAPTER 5 – CONCLUSIONS

(i) Identified a highly divergent dicistrovirus in metagenomic sequencing data from sewage using an HMM approach

(ii) Demonstrated that codon adaptation index can be applied to flavivirus subgroups and individual flavivirus species across a number of hosts

(iii) Discovered and characterized an outbreak of Parvovirus B19 in the summer months of 2013/2014 in Guarujá, Brazil

(iv) Sequenced and characterized a Yellow Fever virus isolate from a convalescent patient's urine

REFERENCES*

1. Di Paola N, Zanotto PMDA. Where there is no overlap , there is a gap. Rev Soc Bras Med Trop. 2014;47(6):675–6.
2. Gatherer D. The 2014 Ebola virus disease outbreak in west Africa. J Gen Virol [Internet]. 2014 May 2 [cited 2014 Jul 15];1619–24. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24795448
3. Althaus CL. Estimating the Reproduction Number of Ebola Virus (EBOV) During the 2014 Outbreak in West Africa. PLoS Curr Outbreaks. 2014;1–9.
4. Figueiredo MLG de, Figueiredo LTM. Emerging alphaviruses in the Americas : Chikungunya and Mayaro. Rev Soc Bras Med Trop. SciELO Brasil; 2014;47(October):677–83.
5. Coimbra TLM, Santos CLS, Suzuki A, Petrella SMC, Bisordi I, Nagamori AH, et al. Mayaro virus: imported cases of human infection in São Paulo State, Brazil. Rev Inst Med Trop Sao Paulo [Internet]. 2007;49(4):221–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17823750
6. Figueiredo LTM. Saint Louis encephalitis virus and other arboviruses in the differential diagnosis for dengue. Rev Soc Bras Med Trop. SciELO Brasil; 2014;47(August):541–2.
7. West TE, Arnim A von SA. Clinical Presentation and Management of Severe Ebola Virus Disease. Ann Am … [Internet]. 2014 [cited 2014 Nov 11];1–44. Available from: http://www.atsjournals.org/doi/abs/10.1513/AnnalsATS.201410-481PS
8. Messina JP, Brady OJ, Scott TW, Zou C, Pigott DM, Duda K a, et al. Global spread of dengue virus types: mapping the 70 year history. Trends Microbiol [Internet]. Elsevier Ltd; 2014 Mar [cited 2014 Apr 29];22(3):138–46. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3946041&tool=pmcentrez&rendertype=abstract
9. Teixeira MG, Siqueira JB, Ferreira GLC, Bricks L, Joint G. Epidemiological trends of dengue disease in Brazil (2000-2010): a systematic literature search and analysis. PLoS Negl Trop Dis [Internet]. 2013 Jan [cited 2014 May 5];7(12):e2520. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3871634&tool=pmcentrez&rendertype=abstract
10. Pan American Health Organization [Internet]. Pan American Health Organization. 2014. Available from: http://www.paho.org
11. Daszak P, Lipkin WI. The search for meaning in virus discovery. Curr Opin Virol [Internet]. Elsevier B.V.; 2011 Dec [cited 2014 Apr 7];1(6):620–3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22440920
12. Mbonye A, Wamala J, Nanyunja M. Ebola viral hemorrhagic disease outbreak in west Africa-lessons from Uganda. African Heal … [Internet]. 2014 [cited 2014 Nov 11];14(3). Available from: http://www.ajol.info/index.php/ahs/article/view/107213
13. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol [Internet]. 2011 Oct [cited 2013 May 21];7(10):e1002195. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3197634&tool=pmcentrez&rendertype=abstract
14. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7 : Improvements in Performance and Usability Article Fast Track. Mol Biol Evol. 2013;30(4):772–80.
15. Puigbò P, Bravo IG, Garcia-Vallvé S. E-CAI: a novel server to estimate an expected value of Codon Adaptation Index (eCAI). BMC Bioinformatics. 2008;9:65.
16. Sharp PM, Li W-H. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res [Internet]. 1987;15(3):1281–95. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/15.3.1281
17. Bahir I, Fromer M, Prat Y, Linial M. Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. Mol Syst Biol [Internet]. Nature Publishing Group; 2009 Oct 13;5(311):311. Available from: http://msb.embopress.org/cgi/doi/10.1038/msb.2009.71
18. Longdon B, Brockhurst MA, Russell CA, Welch JJ, Jiggins FM. The Evolution and Genetics of Virus Host Shifts. PLoS Pathog. 2014;10(11).
19. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. Trends Genet [Internet]. Elsevier Ltd; 2013;29(10):569–74. Available from: http://dx.doi.org/10.1016/j.tig.2013.05.010
20. Shay T, Jojic V, Zuk O, Rothamel K, Puyraimond-Zemmour D, Feng T, et al. Conservation and

*De acordo com:
International Committee of Medical Journal Editors. [Internet]. Uniform requirements for manuscripts submitted to biomedical journals. [2011 Jul 15]. Available from: http://www.nlm.nih.gov/bsd/uniform_requirements.htlm

divergence in the transcriptional programs of the human and mouse immune systems. Proc Natl Acad Sci [Internet]. 2013 Feb 19;110(8):2946–51. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3581886&tool=pmcentrez&rendertype=abstract
21. Pruitt KD, Harrow J, Harte RA, Wallin C, Diekhans M, Maglott DR, et al. The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. Genome Res. 2009;19(7):1316–23.
22. Taniguchi T, Takaoka A. The interferon-α/β system in antiviral responses: A multimodal machinery of gene regulation by the IRF family of transcription factors. Curr Opin Immunol. 2002;14(1):111–6.
23. Yoneyama M, Kikuchi M, Natsukawa T, Shinobu N, Imaizumi T, Miyagishi M, et al. The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. Nat Immunol. 2004;5(7):730–7.
24. Yoneyama M, Kikuchi M, Matsumoto K, Imaizumi T, Miyagishi M, Taira K, et al. Shared and Unique Functions of the DExD/H-Box Helicases RIG-I, MDA5, and LGP2 in Antiviral Innate Immunity. J Immunol [Internet]. 2005 Sep 1;175(5):2851–8. Available from: http://www.jimmunol.org/cgi/doi/10.4049/jimmunol.175.5.2851
25. Akira S, Uematsu S, Takeuchi O. Pathogen recognition and innate immunity. Cell [Internet]. 2006 Feb 24;124(4):783–801. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0092867406001905
26. Thio CL. Host Genetic Factors and Antiviral Immune Responses to Hepatitis C Virus. Clin Liver Dis [Internet]. 2008 Aug;12(3):713–26. Available from: http://linkinghub.elsevier.com/retrieve/pii/S108932610800041X
27. Liu ZY, Li XF, Jiang T, Deng YQ, Ye Q, Zhao H, et al. Viral RNA switch mediates the dynamic control of flavivirus replicase recruitment by genome cyclization. Elife. 2016;5(OCTOBER2016):1–27.
28. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature [Internet]. 2015;526(7571):68–74. Available from: http://www.nature.com/nature/journal/v526/n7571/fig_tab/nature15393_SF1.html%5Cnhttp://dx.doi.org/10.1038/nature15393%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/26432245%5Cnhttp://www.nature.com/doifinder/10.1038/nature15393
29. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One. 2010 Jan;5(3):e9490.
30. Bakonyi T, Hubálek Z, Rudolf I, Nowotny N. Novel Flavivirus or New Lineage of West Nile Virus, Central Europe. Emerg Infect Dis [Internet]. 2005 Feb;11(2):225–31. Available from: http://dx.doi.org/10.3201/eid1102.041028
31. May FJ, Davis CT, Tesh RB, Barrett ADT. Phylogeography of West Nile virus: from the cradle of evolution in Africa to Eurasia, Australia, and the Americas. J Virol [Internet]. 2011 Mar [cited 2013 Oct 1];85(6):2964–74. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3067944&tool=pmcentrez&rendertype=abstract
32. Li WLS, Drummond AJ. Model averaging and Bayes factor calculation of relaxed molecular clocks in Bayesian phylogenetics. Mol Biol Evol. 2012;29(2):751–61.
33. Minin VN, Bloomquist EW, Suchard MA. Smooth Skyride through a Rough Skyline : Bayesian Coalescent-Based Inference of Population Dynamics. Mol Biol Evol [Internet]. 2000 Apr 3;25(7):1459–71. Available from: http://mbe.oxfordjournals.org/cgi/doi/10.1093/molbev/msn090
34. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol [Internet]. 2007 Jan [cited 2013 Aug 1];7:214. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2247476&tool=pmcentrez&rendertype=abstract
35. Day T, Otto SP. Fitness. In: Encyclopedia of Life Sciences [Internet]. Chichester, UK: John Wiley & Sons, Ltd; 2001. p. 1–6. Available from: http://dx.doi.org/10.1038/npg.els.0001745
36. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. Oxford University Press; 2014;30(15):2114–20.
37. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA; 2012;19(5):455–

77.
38. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res [Internet]. 2004 Jan [cited 2013 Aug 7];32(5):1792–7. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=390337&tool=pmcentrez&rendertype=abstract
39. Rambaut A. Se-Al: sequence alignment editor. 1996;
40. Durigon EL, Erdman DD, Gary GW, Pallansch MA, Torok TJ, J.Anderson L. Multiple primer pairs for polymerase chain reaction (PCR) amplification of human parvovirus B19 DNA. J Virol Methods. 1993;44(2):155–65.
41. Erdman DD, Durigon EL, Wang Q-Y, Anderson LJ. Genetic diversity of human parvovirus B19: sequence analysis of the VP1/VP2 gene from multiple isolates. J Gen Virol. Microbiology Society; 1996;77(11):2767–74.
42. Delwart EL. Viral metagenomics. 2007;(February):115–31.
43. Wang L. Discovering novel zoonotic viruses. N S W Public Health Bull [Internet]. 2011 [cited 2014 Mar 13];113–7. Available from: http://www.publish.csiro.au/?paper=NB10078
44. Lipkin WI, Osterhaus a DME. Emerging viruses. Curr Opin Virol [Internet]. Elsevier B.V.; 2013 Apr [cited 2013 Jul 30];3(2):168–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23578937
45. Zaki A. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. New Engl … [Internet]. 2012 [cited 2013 Aug 9]; Available from: http://www.nejm.org/doi/full/10.1056/NEJMoa1211721
46. Dobson AP, Carper ER. Infectious Diseases and Human Population History. Bioscience [Internet]. 1996 Feb;46(2):115–26. Available from: http://www.jstor.org/stable/info/10.2307/1312814
47. Taylor LH, Latham SM, Woolhouse ME. Risk factors for human disease emergence. Philos Trans R Soc Lond B Biol Sci [Internet]. 2001 Jul 29 [cited 2013 Aug 9];356(1411):983–9. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1088493&tool=pmcentrez&rendertype=abstract
48. Parrish CR, Holmes EC, Morens DM, Park E-C, Burke DS, Calisher CH, et al. Cross-species virus transmission and the emergence of new epidemic diseases. Microbiol Mol Biol Rev [Internet]. 2008 Sep [cited 2013 Aug 15];72(3):457–70. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2546865&tool=pmcentrez&rendertype=abstract
49. Wolfe N, Daszak P. Bushmeat hunting, deforestation, and prediction of zoonotic disease. … Infect Dis [Internet]. 2005 [cited 2014 Apr 16];11(12). Available from: http://wwwnc.cdc.gov/eid/article/11/12/04-0789.htm
50. Diamond J. Evolution, consequences and future of plant and animal domestication. Nature [Internet]. 2002 Aug 8;418(6898):700–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12167878
51. Wolfe ND, Dunavan CP, Diamond J. Origins of major human infectious diseases. Nature. 2007;447:279–83.
52. Plowright RK, Parrish CR, McCallum H, Hudson PJ, Ko AI, Graham AL, et al. Pathways to zoonotic spillover. Nat Rev Microbiol [Internet]. Nature Publishing Group; 2017 May 30;1–29. Available from: http://dx.doi.org/10.1038/nrmicro.2017.45
53. Heeney JL, Dalgleish AG, Weiss R a. Origins of HIV and the evolution of resistance to AIDS. Science [Internet]. 2006 Jul 28 [cited 2013 Aug 6];313(5786):462–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16873637
54. Donaldson EF, Haskew AN, Gates JE, Huynh J, Moore CJ, Frieman MB. Metagenomic analysis of the viromes of three North American bat species: viral diversity among different bat species that share a common habitat. J Virol [Internet]. 2010 Dec [cited 2013 Aug 8];84(24):13004–18. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3004358&tool=pmcentrez&rendertype=abstract
55. Bozeman F, Masiello S, Williams M, Elisberg B. Epidemic typhus rickettsiae isolated from flying squirrels. Nature [Internet]. 1975 [cited 2013 Aug 20]; Available from: http://www.nature.com/nature/journal/v255/n5509/abs/255545a0.html
56. Woolhouse MEJ, Haydon DT, Antia R. Emerging pathogens: the epidemiology and evolution of species jumps. Trends Ecol Evol [Internet]. 2005 May [cited 2013 Aug 9];20(5):238–44. Available

from: http://www.ncbi.nlm.nih.gov/pubmed/16701375
57. Parrish CR, Kawaoka Y. The origins of new pandemic viruses: the acquisition of new host ranges by canine parvovirus and influenza A viruses. Annu Rev Microbiol. 2005;59:553–86.
58. Holmes EC. What can we predict about viral evolution and emergence? Curr Opin Virol [Internet]. Elsevier B.V.; 2013 Apr [cited 2013 Aug 9];3(2):180–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23273851
59. Holmes EC, Grenfell BT. Discovering the Phylodynamics of RNA Viruses. Fraenkel E, editor. PLoS Comput Biol [Internet]. 2009 Oct 26;5(10):e1000505. Available from: http://dx.plos.org/10.1371/journal.pcbi.1000505
60. Li W, Zhang C, Sui J, Kuhn JH, Moore MJ, Luo S, et al. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. EMBO J [Internet]. 2005 Apr 20 [cited 2013 Aug 20];24(8):1634–43. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1142572&tool=pmcentrez&rendertype=abstract
61. Wu K, Peng G, Wilken M, Geraghty RJ, Li F. Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. J Biol Chem [Internet]. 2012 Mar 16 [cited 2013 Aug 20];287(12):8904–11. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3308800&tool=pmcentrez&rendertype=abstract
62. Williamson S. Adaptation in the env gene of HIV-1 and evolutionary theories of disease progression. Mol Biol Evol [Internet]. 2003 Aug [cited 2013 Sep 3];20(8):1318–25. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12777505
63. Material SO, Press H, York N, Nw A. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. Science [Internet]. 2004 Mar 12 [cited 2013 Aug 9];303(5664):1666–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14752165
64. Matrosovich M, Tuzikov A. Early Alterations of the Receptor-Binding Properties of H1 , H2 , and H3 Avian Influenza Virus Hemagglutinins after Their Introduction into Mammals. J … [Internet]. 2000 [cited 2013 Aug 17]; Available from: http://jvi.asm.org/content/74/18/8502.short
65. Okie S. Fighting HIV--lessons from Brazil. N Engl J Med [Internet]. 2006 May 11;354(19):1977–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16687709
66. Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science (80-). 1996;271:1582–6.
67. Roberts J D Bebenek K Kunkel TA. Accuracy Transcriptase. Science (80-). 1988;1:1986–8.
68. Duffy S, Shackelton L a, Holmes EC. Rates of evolutionary change in viruses: patterns and determinants. Nat Rev Genet [Internet]. 2008 Apr [cited 2013 May 24];9(4):267–76. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18319742
69. Rambaut A, Posada D, Crandall K a, Holmes EC. The causes and consequences of HIV evolution. Nat Rev Genet [Internet]. 2004 Jan [cited 2013 Aug 14];5(1):52–61. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14708016
70. Zhuang J, Jetzt AE, Sun G, Yu H, Klarmann G, Ron Y, et al. Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. J Virol [Internet]. 2002 Nov [cited 2013 Aug 29];76(22):11273–82. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=136766&tool=pmcentrez&rendertype=abstract
71. Simon-Loriere E, Holmes EC. Why do RNA viruses recombine? Nat Rev Microbiol [Internet]. Nature Publishing Group; 2011 Aug [cited 2013 Aug 16];9(8):617–26. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3324781&tool=pmcentrez&rendertype=abstract
72. Fancello L, Raoult D, Desnues C. Computational tools for viral metagenomics and their application in clinical research. Virology [Internet]. Elsevier; 2012 Dec 20 [cited 2013 May 27];434(2):162–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23062738
73. Lipkin WI, Firth C. Viral surveillance and discovery. Curr Opin Virol [Internet]. Elsevier B.V.; 2013 Apr [cited 2013 May 30];3(2):199–204. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23602435
74. Tapparel C, Cordey S, Van Belle S, Turin L, Lee W-M, Regamey N, et al. New molecular detection tools adapted to emerging rhinoviruses and enteroviruses. J Clin Microbiol [Internet]. 2009 Jun

[cited 2013 Aug 20];47(6):1742–9. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2691104&tool=pmcentrez&rendertype=abstract
75. Lipkin WI. The changing face of pathogen discovery and surveillance. Nat Rev Microbiol [Internet]. 2013 Feb [cited 2013 Aug 9];11(2):133–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23268232
76. Mokili JL, Rohwer F, Dutilh BE. Metagenomics and future perspectives in virus discovery. Curr Opin Virol [Internet]. Elsevier B.V.; 2012 Feb [cited 2013 Aug 9];2(1):63–77. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22440968
77. Lipkin WI, Travis GH, Carbone KM, Wilson MC. Isolation and characterization of Borna disease agent cDNA clones. Proc Natl Acad Sci U S A [Internet]. 1990 Jun;87(11):4184–8. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=54072&tool=pmcentrez&rendertype=abstract
78. Mokili JLK, Rogers M, Carr JK, Simmonds P, Bopopi JM, Foley BT, et al. Identification of a novel clade of human immunodeficiency virus type 1 in Democratic Republic of Congo. AIDS Res Hum Retroviruses [Internet]. 2002 Jul 20 [cited 2013 Sep 3];18(11):817–23. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12167273
79. Clewley JP, Lewis JCM, Brown DWG, Gadsby EL. A novel simian immunodeficiency virus (SIVdrl) pol sequence from the drill monkey, Mandrillus leucophaeus. J Virol. 1998;72:10305–9.
80. Reyes GR, Yarbough PO, Tam AW, Purdy MA, Huang CC, Kim JS, et al. Hepatitis E virus (HEV): the novel agent responsible for enterically transmitted non-A, non-B hepatitis. Gastroenterol Jpn. 1991;26 Suppl 3:142–7.
81. Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev. 1995;59:143–69.
82. Bartenschlager R, Pietschmann T. Efficient hepatitis C virus cell culture system: what a difference the host cell makes. Proc Natl Acad Sci U S A [Internet]. 2005 Jul 12;102(28):9739–40. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1175013&tool=pmcentrez&rendertype=abstract
83. Staheli JP, Boyce R, Kovarik D, Rose TM. CODEHOP PCR and CODEHOP PCR primer design. Methods Mol Biol [Internet]. 2011 Jan [cited 2013 Sep 3];687:57–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20967601
84. Rose TM, Schultz ER, Henikoff JG, Pietrokovski S, McCallum CM, Henikoff S. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. Nucleic Acids Res [Internet]. 1998 Apr 1;26(7):1628–35. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=147464&tool=pmcentrez&rendertype=abstract
85. Johne R, Enderlein D, Nieper H, Müller H. Novel polyomavirus detected in the feces of a chimpanzee by nested broad-spectrum PCR. J Virol [Internet]. 2005 [cited 2014 Apr 7]; Available from: http://jvi.asm.org/content/79/6/3883.short
86. Suttle C a. Viruses in the sea. Nature [Internet]. 2005 Sep 15 [cited 2013 May 21];437(7057):356–61. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16163346
87. Culley AI, Lang AS, Suttle CA. Metagenomic analysis of coastal RNA virus communities. Science (80-). 2006;312:1795–8.
88. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol. 1998;5:R245–9.
89. Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, et al. Genomic analysis of uncultured marine viral communities. Proc Natl Acad Sci U S A [Internet]. 2002 Oct 29;99(22):14250–5. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=137870&tool=pmcentrez&rendertype=abstract
90. Hugenholtz P, Tyson GW. Metagenomics. Nat Q&A. 2008;455(August 2006):2006–8.
91. Faust K, Raes J. Microbial interactions: from networks to models. Nat Rev Microbiol [Internet]. Nature Publishing Group; 2012 Aug [cited 2013 Aug 8];10(8):538–50. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22796884

92. Claesson MJ, Jeffery IB, Conde S, Power SE, O'Connor EM, Cusack S, et al. Gut microbiota composition correlates with diet and health in the elderly. Nature. 2012;488:1–8.
93. Peterson DA, Frank DN, Pace NR, Gordon JI. Metagenomic approaches for defining the pathogenesis of inflammatory bowel diseases. Cell Host Microbe [Internet]. 2008 Jun 12 [cited 2014 Feb 27];3(6):417–27. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2872787&tool=pmcentrez&rendertype=abstract
94. Breitbart M, Hewson I, Felts B. Metagenomic analyses of an uncultured viral community from human feces. J … [Internet]. 2003 [cited 2013 Sep 4];185(20). Available from: http://jb.asm.org/content/185/20/6220.short
95. Breitbart M, Haynes M, Kelley S, Angly F, Edwards RA, Felts B, et al. Viral diversity and dynamics in an infant gut. Res Microbiol [Internet]. 2008 Jun [cited 2013 Sep 3];159(5):367–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18541415
96. Esposito JJ, Sammons S a, Frace a M, Osborne JD, Olsen-Rasmussen M, Zhang M, et al. Genome sequence diversity and clues to the evolution of variola (smallpox) virus. Science [Internet]. 2006 Aug 11 [cited 2013 Jul 27];313(5788):807–12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16873609
97. Zhang Y, Sun Y. HMM-FRAME: accurate protein domain classification for metagenomic sequences containing frameshift errors. BMC Bioinformatics. 2011;12:198.
98. Breitbart M, Rohwer F. Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing. Biotechniques [Internet]. 2005 Nov [cited 2013 Aug 19];39(5):729–36. Available from: http://www.biotechniques.com/article/05395RR03
99. Munang'andu HM, Mugimba KK, Byarugaba DK, Mutoloki S, Evensen Ø. Current advances on virus discovery and diagnostic role of viral metagenomics in aquatic organisms. Front Microbiol. Frontiers Media SA; 2017;8.
100. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature. Nature Research; 2016;537(7622):689–93.
101. Simmonds P, Adams MJ, Benkő M, Breitbart M, Brister JR, Carstens EB, et al. Consensus statement: Virus taxonomy in the age of metagenomics. Nat Rev Microbiol. Nature Research; 2017;15(3):161–8.
102. Zhang W, Li L, Deng X, Kapusinszky B, Delwart E. What is for dinner? Viral metagenomics of US store bought beef, pork, and chicken. Virology [Internet]. Elsevier; 2014;468–470(Supplement C):303–10. Available from: http://www.sciencedirect.com/science/article/pii/S004268221400405X
103. Palacios G, Druce J, Du L, Tran T. A New arenavirus in a cluster of fatal transplant-associated diseases. New Engl J … [Internet]. 2008 [cited 2013 Aug 10];991–8. Available from: http://www.nejm.org/doi/full/10.1056/NEJMoa073785
104. Sullivan PF, Allander T, Lysholm F, Goh S, Persson B, Jacks A, et al. An unbiased metagenomic search for infectious agents using monozygotic twins discordant for chronic fatigue. BMC Microbiol [Internet]. 2011 Jan [cited 2013 Aug 20];11:2. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3022642&tool=pmcentrez&rendertype=abstract
105. Briese T, Paweska JT, McMullan LK, Hutchison SK, Street C, Palacios G, et al. Genetic Detection and Characterization of Lujo Virus, a New Hemorrhagic Fever–Associated Arenavirus from Southern Africa. PLoS Pathog. Public Library of Science; 2009;5(5):8.
106. McMullan LK, Frace M, Sammons SA, Shoemaker T, Balinandi S, Wamala JF, et al. Using next generation sequencing to identify yellow fever virus in Uganda. Virology [Internet]. 2012 Jan 5 [cited 2013 Sep 3];422(1):1–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21962764
107. Nakamura S, Yang C-S, Sakon N, Ueda M, Tougan T, Yamashita A, et al. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. PLoS One [Internet]. 2009 Jan [cited 2013 Jul 31];4(1):e4219. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2625441&tool=pmcentrez&rendertype=abstract
108. Su Z, Ning B, Fang H, Hong H, Perkins R, Tong W, et al. Next-generation sequencing and its applications in molecular diagnostics. Expert Rev Mol Diagn [Internet]. 2011 Apr [cited 2013 Sep

3];11(3):333–43. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21463242
109. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science (80-). 2008;319:1096–100.
110. Kiechle IFL. Point-of-Care Testing : Past , Present and Future. Point Care J NearPatient Test Technol. 2010;7:97–9.
111. Willner D, Furlan M, Haynes M, Schmieder R, Angly FE, Silva J, et al. Metagenomic Analysis of Respiratory Tract DNA Viral Communities in Cystic Fibrosis and Non-Cystic Fibrosis Individuals. PLoS One. 2009;4:12.
112. Breitbart M, Miyake JH, Rohwer F. Global distribution of nearly identical phage-encoded DNA sequences. FEMS Microbiol Lett [Internet]. 2004 Jul 15 [cited 2013 Jun 7];236(2):249–56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15251204
113. Karplus K, Barrett C, Hughey R. Hidden Markov models for detecting remote protein homologies. Bioinformatics [Internet]. 1998 Jan;14(10):846–56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9927713
114. Hambly E, Suttle CA. The viriosphere, diversity, and genetic exchange within phage communities. Curr Opin Microbiol. 2005 Aug;8(4):444–50.
115. Ng TFF, Marine R, Wang C, Simmonds P, Kapusinszky B, Bodhidatta L, et al. High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. J Virol [Internet]. 2012 Nov [cited 2013 Aug 16];86(22):12161–75. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3486453&tool=pmcentrez&rendertype=abstract
116. Scallan E, Hoekstra RM, Widdowson M-A, Hall AJ, Griffin PM. Foodborne illness acquired in the United States. Vol. 17, Emerging Infectious Diseases. 2011. p. 1339–40.
117. Brown DW, Gunning KB, Henry DM, Awdeh ZL, Brinker JP, Tzipori S, et al. A DNA oligonucleotide microarray for detecting human astrovirus serotypes. J Virol Methods. 2008;147:86–92.
118. Poch O, Sauvaget I, Delarue M, Tordo N. Identification of four conserved motifs among the RNA-dependant polymerase encoding elements. 1989;8(12):3867–74.
119. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol [Internet]. 2010 Feb [cited 2013 Aug 6];27(2):221–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19854763
120. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics [Internet]. 2012 Jun 15 [cited 2013 Aug 13];28(12):1647–9. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3371832&tool=pmcentrez&rendertype=abstract
121. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res [Internet]. 2011 Jul [cited 2013 Aug 7];39(Web Server issue):W29-37. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3125773&tool=pmcentrez&rendertype=abstract
122. Koonin E V, Wolf YI, Nagasaki K, Dolja V V. The Big Bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups. Nat Rev Microbiol [Internet]. 2008 Dec [cited 2013 May 24];6(12):925–39. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18997823
123. Mari J, Poulos BT, Lightner D V, Bonami J-R. Shrimp Taura syndrome virus: genomic characterization and similarity with members of the genus Cricket paralysis-like viruses. J Gen Virol [Internet]. 2002 Apr;83(Pt 4):915–26. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11907342
124. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003;52:696–704.
125. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res [Internet]. 1994 Nov 11;22(22):4673–80. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=308517&tool=pmcentrez&rendertype=abstract
126. Koonin E V, Dolja V V. Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. Crit Rev Biochem Mol Biol. 1993;28:375–430.

127. Zanotto PM, Gibbs MJ, Gould E a, Holmes EC. A reevaluation of the higher taxonomy of viruses based on RNA polymerases. J Virol [Internet]. 1996 Sep;70(9):6083–96. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=190630&tool=pmcentrez&rendertype=abstract
128. Price MN, Dehal PS, Arkin AP. FastTree : Neighbor-Joining with Profiles instead of a Dis- tance Matrix.
129. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol [Internet]. 2001 May;18(5):691–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11319253
130. Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. Nat Rev Genet [Internet]. Nature Publishing Group; 2011 Jan [cited 2013 Sep 16];12(1):32–42. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3074964&tool=pmcentrez&rendertype=abstract
131. Sharp PM, Tuohy TM, Mosurski KR. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res [Internet]. 1986 Jul 11;14(13):5125–43. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=311530&tool=pmcentrez&rendertype=abstract
132. Jenkins GM, Pagel M, Gould E a, Zanotto PM de A, Holmes EC. Evolution of Base Composition and Codon Usage Bias in the Genus Flavivirus. J Mol Evol [Internet]. 2001 Apr [cited 2013 Sep 18];52(4):383–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11343134
133. Butt AM, Nasrullah I, Qamar R, Tong Y. Evolution of codon usage in Zika virus genomes is host and vector specific. Emerg Microbes Infect [Internet]. Nature Publishing Group; 2016;5(10):e107. Available from: http://www.nature.com/doifinder/10.1038/emi.2016.106
134. Shin YC, Bischof GF, Lauer WA, Desrosiers RC. Importance of codon usage for the temporal regulation of viral gene expression. Proc Natl Acad Sci [Internet]. 2015;112(45):201515387. Available from: http://www.pnas.org/content/early/2015/10/22/1515387112.abstract?sid=565a050f-9f4c-45d6-b4dd-552e819e89c5
135. Urrutia a O, Hurst LD. Codon usage bias covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection. Genetics [Internet]. 2001 Nov;159(3):1191–9. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1461876&tool=pmcentrez&rendertype=abstract
136. Cugola FR, Fernandes IR, Russo FB, Freitas BC, Dias JLM, Guimarães KP, et al. The Brazilian Zika virus strain causes birth defects in experimental models. Nature [Internet]. 2016 Jun 9;534(7606):267–271. Available from: http://www.nature.com/doifinder/10.1038/nature18296
137. Andersen KG, Shapiro BJ, Matranga CB, Sealfon R, Lin AE, Moses LM, et al. Clinical Sequencing Uncovers Origins and Evolution of Lassa Virus. Cell. 2015;162(4):738–50.
138. Faye O, Andronico A, Faye O, Salje H, Bo??lle PY, Magassouba N, et al. Use of Viremia to Evaluate the Baseline Case Fatality Ratio of Ebola Virus Disease and Inform Treatment Studies: A Retrospective Cohort Study. PLoS Med. 2015;12(12):1–14.
139. Freire CC de M, Iamarino A, Neto DF de L, Sall AA, Zanotto PM de A. Spread of the pandemic Zika virus lineage is associated with NS1 codon usage adaptation in humans [Internet]. Vol. 2015. 2015 Nov. Available from: http://biorxiv.org/lookup/doi/10.1101/032839
140. Vasilakis N, Deardorff ER, Kenney JL, Rossi SL, Hanley KA, Weaver SC. Mosquitoes Put the Brake on Arbovirus Evolution: Experimental Evolution Reveals Slower Mutation Accumulation in Mosquito Than Vertebrate Cells. Andino R, editor. PLoS Pathog [Internet]. 2009 Jun 5;5(6):e1000467. Available from: http://dx.plos.org/10.1371/journal.ppat.1000467
141. Behura SK, Severson DW. Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes. Biol Rev [Internet]. 2013 Feb;88(1):49–61. Available from: http://doi.wiley.com/10.1111/j.1469-185X.2012.00242.x
142. Behura SK, Severson DW. Bicluster pattern of codon context usages between flavivirus and vector mosquito Aedes aegypti: relevance to infection and transcriptional response of mosquito genes. Mol Genet Genomics [Internet]. 2014 Oct 18;289(5):885–94. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022202X15370834

143. Shen SH, Stauft CB, Gorbatsevych O, Song Y, Ward CB, Yurovsky A, et al. Large-scale recoding of an arbovirus genome to rebalance its insect versus mammalian preference. Proc Natl Acad Sci [Internet]. 2015 Apr 14;112(15):4749–54. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4403163&tool=pmcentrez&rendertype=abstract
144. Pachler K, Lebl K, Berer D, Rudolf I, Hubalek Z, Nowotny N, et al. Putative New West Nile Lineage in Uranotaenia unguiculata Mosquitoes, Austria, 2013. Emerg Infect Dis. 2014;20(12):2119–22.
145. Fall G, Diallo M, Loucoubar C, Faye O, Sall AA. Vector competence of Culex neavei and Culex quinquefasciatus (Diptera: Culicidae) from Senegal for lineages 1, 2, Koutango and a putative new lineage of West Nile virus. Am J Trop Med Hyg. 2014;90(4):747–54.
146. Peterson D, Frank D, Pace N, Gordon J. Metagenomic approaches for defining the pathogenesis of inflammatory bowel diseases. Cell Host Microbe [Internet]. 2008 [cited 2014 Mar 13];3(6):417–27. Available from: http://www.sciencedirect.com/science/article/pii/S1931312808001492
147. Murray KO, Mertens E, Despres P. West Nile virus and its emergence in the United States of America. Vet Res. 2010;41(6):67.
148. McMullen AR, Albayrak H, May FJ, Davis CT, Beasley DWC, Barrett ADT. Molecular evolution of lineage 2 West Nile virus. J Gen Virol [Internet]. 2013 Feb [cited 2014 Nov 3];94(Pt 2):318–25. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3709619&tool=pmcentrez&rendertype=abstract
149. Villabona-Arenas CJ, Zanotto PM de A. Worldwide spread of Dengue virus type 1. PLoS One [Internet]. 2013;8(5):e62649. Available from: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0062649%5Cnhttp://dx.plos.org/10.1371/journal.pone.0062649
150. Gubler DJ. The changing epidemiology of yellow fever and dengue, 1900 to 2003: Full circle? Comp Immunol Microbiol Infect Dis. 2004;27(5):319–30.
151. Gubler DJ. Dengue and dengue hemorrhagic fever. Clin Microbiol Rev. Am Soc Microbiol; 1998;11(3):480–96.
152. Gubler DJ, Gubler DJ, Reiter P, Reiter P, Ebi KL, Ebi KL, et al. Climate variability and change in the United States: potential impacts on vector- and rodent-borne diseases. Environ Health Perspect [Internet]. 2001;109 Suppl(May):223–33. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11359689
153. Bryant JE, Holmes EC, Barrett ADT. Out of Africa: A molecular perspective on the introduction of yellow fever virus into the Americas. PLoS Pathog. 2007;3(5):0668–73.
154. Beck A, Guzman H, Li L, Ellis B, Tesh RB, Barrett ADT. Phylogeographic Reconstruction of African Yellow Fever Virus Isolates Indicates Recent Simultaneous Dispersal into East and West Africa. PLoS Negl Trop Dis. 2013;7(3).
155. Eisenberg E, Levanon EY. Human housekeeping genes are compact. TRENDS Genet. Elsevier; 2003;19(7):362–5.
156. Saffert RT, Kalejta RF. Inactivating a cellular intrinsic immune defense mediated by Daxx is the mechanism through which the human cytomegalovirus pp71 protein stimulates viral immediate-early gene expression. J Virol [Internet]. 2006;80(8):3863–71. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1440479&tool=pmcentrez&rendertype=abstract
157. Zanotto PM, Kallas EG, de Souza RF, Holmes EC. Genealogical evidence for positive selection in the nef gene of HIV-1. Genetics [Internet]. 1999 Nov;153(3):1077–89. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10545443
158. Schmidt GW, Delaney SK. Stable internal reference genes for normalization of real-time RT-PCR in tobacco (Nicotiana tabacum) during development and abiotic stress. Mol Genet Genomics [Internet]. 2010 Mar 23;283(3):233–41. Available from: http://link.springer.com/10.1007/s00438-010-0511-1
159. Nicot N. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. J Exp Bot [Internet]. 2005 Oct 10;56(421):2907–14. Available from: https://academic.oup.com/jxb/article-lookup/doi/10.1093/jxb/eri285
160. Behura SK, Severson DW. Comparative analysis of codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes. PLoS One. Public Library of Science;

2012;7(8):e43111.
161. Simmen MW. Genome-scale relationships between cytosine methylation and dinucleotide abundances in animals. Genomics. Elsevier; 2008;92(1):33–40.
162. Coffey LL, Forrester N, Tsetsarkin K, Vasilakis N, Weaver SC. Factors shaping the adaptive landscape for arborviruses: implications for the emergence of disease. Future Microbiol. 2013;155–76.
163. Coffey LL, Vasilakis N, Brault AC, Powers AM, Tripet F, Weaver SC. Arbovirus evolution in vivo is constrained by host alternation. Proc Natl Acad Sci U S A. 2008;105(19):6970–5.
164. Grard G, Moureau G, Charrel RN, Lemasson JJ, Gonzalez JP, Gallian P, et al. Genetic characterization of tick-borne flaviviruses: New insights into evolution, pathogenetic determinants and taxonomy. Virology. 2007;361(1):80–92.
165. Gaunt MW, Sall AA, Lamballerie X De, Gould E a, Falconar AKI, Dzhivanian TI, et al. Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography. J Gen Virol [Internet]. 2001 Aug 1;82(8):1867–76. Available from: http://dx.doi.org/10.1099/0022-1317-82-8-1867
166. Grard G, Moureau G, Charrel RN, Holmes EC, Gould EA, de Lamballerie X. Genomics and evolution of Aedes-borne flaviviruses. J Gen Virol [Internet]. 2010 Jan 1;91(1):87–94. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19741066
167. Kuno G, Chang GJ, Tsuchiya KR, Karabatsos N, Cropp CB. Phylogeny of the genus Flavivirus. J Virol [Internet]. 1998 Jan [cited 2012 Mar 5];72(1):73–83. Available from: http://jvi.asm.org/content/72/1/73.short
168. Cook S, Moureau G, Kitchen A, Gould EA, de Lamballerie X, Holmes EC, et al. Molecular evolution of the insect-specific flaviviruses. J Gen Virol [Internet]. 2012 Feb 1;93(2):223–34. Available from: http://dx.doi.org/10.1099/vir.0.036525-0
169. Blitvich B, Firth A. Insect-Specific Flaviviruses: A Systematic Review of Their Discovery, Host Range, Mode of Transmission, Superinfection Exclusion Potential and Genomic Organization [Internet]. Vol. 7, Viruses. 2015. 1927-1959 p. Available from: http://www.mdpi.com/1999-4915/7/4/1927/
170. Moureau G, Cook S, Lemey P, Nougairede A, Forrester NL, Khasnatinov M, et al. New Insights into Flavivirus Evolution, Taxonomy and Biogeographic History, Extended by Analysis of Canonical and Alternative Coding Sequences. PLoS One [Internet]. 2015;10(2):e0117849. Available from: http://dx.plos.org/10.1371/journal.pone.0117849
171. Hoshino K, Takahashi-Nakaguchi A, Isawa H, Sasaki T, Higa Y, Kasai S, et al. Entomological surveillance for flaviviruses at migratory bird stopover sites in Hokkaido, Japan, and a new insect flavivirus detected in Aedes galloisi (Diptera: Culicidae). J Med Entomol. The Oxford University Press; 2012;49(1):175–82.
172. Hoshino K, Isawa H, Tsuda Y, Sawabe K, Kobayashi M. Isolation and characterization of a new insect flavivirus from Aedes albopictus and Aedes flavopictus mosquitoes in Japan. Virology. Elsevier; 2009;391(1):119–29.
173. Bolling BG, Eisen L, Moore CG, Blair CD. Insect-specific flaviviruses from Culex mosquitoes in Colorado, with evidence of vertical transmission. Am J Trop Med Hyg. ASTMH; 2011;85(1):169–77.
174. Vázquez A, Sánchez-Seco M-P, Palacios G, Molero F, Reyes N, Ruiz S, et al. Novel flaviviruses detected in different species of mosquitoes in Spain. Vector-Borne Zoonotic Dis. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA; 2012;12(3):223–9.
175. Crabtree MB, Nga PT, Miller BR. Isolation and characterization of a new mosquito flavivirus, Quang Binh virus, from Vietnam. Arch Virol [Internet]. 2009;154. Available from: http://dx.doi.org/10.1007/s00705-009-0373-1
176. Crabtree MB, Sang RC, Stollar V, Dunster LM, Miller BR. Genetic and phenotypic characterization of the newly described insect flavivirus, Kamiti River virus. Arch Virol. Springer; 2003;148(6):1095–118.
177. Kim DY, Guzman H, Bueno R, Dennett JA, Auguste AJ, Carrington C V. Characterization of Culex Flavivirus (Flaviviridae) strains isolated from mosquitoes in the United States and Trinidad. Virology [Internet]. 2009;386. Available from: http://dx.doi.org/10.1016/j.virol.2008.12.034
178. Pettersson JHO, Fiz-Palacios O. Dating the origin of the genus Flavivirus in the light of Beringian biogeography. J Gen Virol. 2014;95(PART 9):1969–82.

179. Zanotto PM, Gould E a, Gao GF, Harvey PH, Holmes EC. Population dynamics of flaviviruses revealed by molecular phylogenies. Proc Natl Acad Sci U S A. 1996;93(2):548–53.
180. Sall AA, Faye O, Diallo M, Firth C, Kitchen A, Holmes EC. Yellow fever virus exhibits slower evolutionary dynamics than dengue virus. J Virol [Internet]. 2010 Jan 15;84(2):765–72. Available from: http://jvi.asm.org/cgi/doi/10.1128/JVI.01738-09
181. Marra PP, Griffing SM, McLean RG. West Nile Virus and Wildlife. Emerg Infect Dis. 2003;9(7):7–9.
182. Petersen LR, Roehrig JT. West Nile virus: a reemerging global pathogen. Rev Biomed. 2001;12(3).
183. Beasley DWC, Li L, Suderman MT, Barrett ADT. Mouse Neuroinvasive Phenotype of West Nile Virus Strains Varies Depending upon Virus Genotype. Virology [Internet]. Elsevier; 2002 Apr;296(1):17–23. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0042682202913723
184. Bakonyi TT, Ivanics E, Erdleyi K, Ursu K, Ferenczi EE, Weissenbock H, et al. Lineage 1 and 2 strains of encephalitic West Nile virus, Central Europe. Emerg Infect Dis. 2006;12(4):618–23.
185. Bakonyi T, Ferenczi E, Erdélyi K, Kutasi O, Csörgő T, Seidel B, et al. Explosive spread of a neuroinvasive lineage 2 West Nile virus in Central Europe, 2008/2009. Vet Microbiol [Internet]. 2013 Jul;165(1–2):61–70. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0378113513001648
186. Papa A, Danis K, Baka A, Bakas A, Dougas G, Lytras T, et al. Ongoing outbreak of West Nile virus infections in humans in Greece, July–August 2010. Age (Omaha). 2010;20:20–9.
187. ECDC. West Nile fever maps [Internet]. European Centre for Disease Prevention and Control. 2017 [cited 2017 Jun 27]. Available from: http://ecdc.europa.eu/en/healthtopics/west_nile_fever/West-Nile-fever-maps/Pages/historical-data.aspx
188. Billoir F, de CR, Tolou H, de MP, Gould EA, de LX. Phylogeny of the genus flavivirus using complete coding sequences of arthropod-borne viruses and viruses with no known vector. J Gen Virol [Internet]. 2000;81 Pt 9. Available from: http://dx.doi.org/10.1099/0022-1317-81-9-2339
189. Gould EA, Zanotto PM de A, Holmes EC. The genetic evolution of flaviviruses. 1997.
190. Kuno G. Host range specificity of flaviviruses: correlation with in vitro replication. J Med Entomol [Internet]. 2007;44. Available from: http://dx.doi.org/10.1093/jmedent/41.5.93
191. Labuda M, Nuttall P a. Tick-borne viruses. Parasitology. 2004;129 Suppl:S221–45.
192. Dobler G. Zoonotic tick-borne flaviviruses. Vet Microbiol. 2010;140(3–4):221–8.
193. Shi M, Lin X-D, Vasilakis N, Tian J-H, Li C-X, Chen L-J, et al. Divergent viruses discovered in arthropods and vertebrates revise the evolutionary history of the *Flaviviridae* and related viruses. J Virol [Internet]. 2015;90(2):JVI.02036-15. Available from: http://jvi.asm.org/lookup/doi/10.1128/JVI.02036-15
194. Cook S, Moureau G, Kitchen A, Gould EA, de Lamballerie X, Holmes EC, et al. Molecular evolution of the insect-specific flaviviruses. J Gen Virol. 2012;93(2):223–34.
195. Bagnarelli P, Marinelli K, Trotta D, Monachetti A, Tavio M, Gobbo R Del, et al. Human case of autochthonous West Nile virus lineage 2 infection in Italy, September 2011. Eurosurveillance. 2011;16(43):1–4.
196. Papa A, Bakonyi TT, Xanthopoulou K, Vázquez A, Tenorio A, Nowotny N, et al. Genetic Characterization of West Nile Virus Lineage 2, Greece, 2010. Emerg Infect Dis [Internet]. 2011 May;17(5):920–2. Available from: http://wwwnc.cdc.gov/eid/article/17/5/10-1759_article.htm
197. Sambri V, Capobianchi M, Charrel R, Fyodorova M, Gaibani P, Gould E, et al. West Nile virus in Europe: emergence, epidemiology, diagnosis, treatment, and prevention. Clin Microbiol Infect [Internet]. 2013 Aug;19(8):699–704. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1198743X14614154
198. Fall G, Di Paola N, Faye M, Dia M, Freire CC de M, Loucoubar C, et al. Biological and phylogenetic characteristics of West African lineages of West Nile virus. PLoS Negl Trop Dis [Internet]. Public Library of Science; 2017 Nov 8;11(11):e0006078. Available from: https://doi.org/10.1371/journal.pntd.0006078
199. Plowright RK, Parrish CR, McCallum H, Hudson PJ, Ko AI, Graham AL, et al. Pathways to zoonotic spillover -Supplementary. Nat Rev Microbiol [Internet]. 2017 May 30;1–3. Available from: http://www.nature.com/doifinder/10.1038/nrmicro.2017.45
200. World Health Organization. National Passive Surveillance [Internet]. 2017 [cited 2017 Dec 28]. Available from: http://www.who.int/immunization/monitoring_surveillance/burden/vpd/surveillance_type/passive/en/

201. Silva MMO, Rodrigues MS, Paploski IAD, Kikuti M, Kasper AM, Cruz JS, et al. Accuracy of Dengue Reporting by National Surveillance System, Brazil. Emerg Infect Dis J [Internet]. 2016;22(2):336. Available from: http://wwwnc.cdc.gov/eid/article/22/2/15-0495
202. Nunes-Araújo FRF, Ferreira MS, Nishioka SDE a. Dengue fever in Brazilian adults and children: assessment of clinical findings and their validity for diagnosis. Ann Trop Med Parasitol [Internet]. 2003 Jun [cited 2014 May 7];97(4):415–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12831527
203. Yozwiak NL, Skewes-Cox P, Stenglein MD, Balmaseda A, Harris E, DeRisi JL. Virus identification in unknown tropical febrile illness cases using deep sequencing. PLoS Negl Trop Dis [Internet]. 2012 Jan [cited 2013 Aug 6];6(2):e1485. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3274504&tool=pmcentrez&rendertype=abstract
204. Lysholm F, Wetterbom A, Lindau C, Darban H, Bjerkner A, Fahlander K, et al. Characterization of the viral microbiome in patients with severe lower respiratory tract infections, using metagenomic sequencing. PLoS One. Public Library of Science; 2012;7(2):e30875.
205. Terzian ACB, Mondini A, Bronzoni RVDM, Drumond BP, Ferro BP, Cabrera EMS, et al. Detection of Saint Louis encephalitis virus in Dengue-suspected cases during a dengue 3 outbreak. Vector Borne Zoonotic Dis [Internet]. 2011 Mar;11(3):291–300. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20645866
206. da Silva Pessoa Vieira CJ, da Silva DJF, Barreto ES, Siqueira CEH, Colombo TE, Ozanic K, et al. Detection of Mayaro virus infections during a dengue outbreak in Mato Grosso, Brazil. Acta Trop [Internet]. Elsevier B.V.; 2015;147:12–6. Available from: http://dx.doi.org/10.1016/j.actatropica.2015.03.020
207. Qiu J, Söderlund-Venermo M, Young NS. Human Parvoviruse. Clin Microbiol Rev. 2016;30(1):43–113.
208. Waza K, Inoue K, Matsumura S. Symptoms Associated with Parvovirus B19 Infection in Adults: A Pilot Study. Intern Med. 2007;46(24):1975–8.
209. Anderson MJ, Higgins PG, Davis LR, Willman JS, Jones SE, Kidd IM, et al. Experimental parvoviral infection in humans. J Infect Dis. The University of Chicago Press; 1985;152(2):257–65.
210. Young NS, Brown KE. Parvovirus B19. N Engl J Med [Internet]. 2004;350(6):586–97. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMra030840
211. Lefrére JJ, Servant-Delmas A, Candotti D, Mariotti M, Thomas I, Brossard Y, et al. Persistent B19 infection in immunocompetent individuals: Implications for transfusion safety. Blood. 2005;106(8):2890–5.
212. Cohen BJ, Beard S, Knowles WA, Ellis JS, Joske D, Goldman JM, et al. Chronic anemia due to parvovirus B19 infection in a bone marrow transplant patient after platelet transfusion. Transfusion. Wiley Online Library; 1997;37(9):947–52.
213. Erdman DD, Anderson BC, Török TJ, Finkel TH, Anderson LJ. Possible transmission of parvovirus B19 from intravenous immune globulin. J Med Virol. Wiley Online Library; 1997;53(3):233–6.
214. Von Kaisenberg CS, Jonat W. Fetal parvovirus B19 infection. Ultrasound Obstet Gynecol [Internet]. 2001 Sep;18(3):280. Available from: http://www.blackwell-synergy.com/rd.asp?code=UOG&vol=18&page=280&goto=abstract
215. Lui SL, Luk WK, Cheung CY, Chan TM, Lai KN, Peiris JS. Nosocomial outbreak of parvovirus B19 infection in a renal transplant unit. Transplantation [Internet]. 2001 Jan 15;71(1):59–64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11211196
216. Heegaard ED, Brown KE. Human parvovirus B19. Clin Microbiol Rev [Internet]. 2002;15(3):485–505. Available from: pm:12097253
217. Chipman PR, Agbandje-McKenna M, Kajigaya S, Brown KE, Young NS, Baker TS, et al. Cryo-electron microscopy studies of empty capsids of human parvovirus B19 complexed with its cellular receptor. Proc Natl Acad Sci [Internet]. 1996 Jul 23;93(15):7502–6. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.93.15.7502
218. Kaufmann J, Buccola JM, Stead W, Rowley C, Wong M, Bates CK. Secondary symptomatic parvovirus B19 infection in a healthy adult. J Gen Intern Med. 2007;22(6):877–8.
219. Kurtzman GJ, Cohen B, Meyers P, Amunullah A, Young NS. Persistent B19 parvovirus infection as a cause of severe chronic anaemia in children with acute lymphocytic leukaemia. Lancet [Internet]. 1988;2(8621):1159–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2903376

220. Jacobson SK, Daly JS, Thorne GM, McIntosh K. Chronic Parvovirus B19 Infection Resulting in Chronic Fatigue Syndrome: Case History and Review. Clin Infect Dis [Internet]. 1997;24(6):1048–51. Available from: https://academic.oup.com/cid/article-lookup/doi/10.1086/513627
221. Musiani M, Manaresi E, Gallinella G, Cricca M, Zerbini M. Recurrent Erythema in Patients with Long-Term Parvovirus B19 Infection. Clin Infect Dis. 2005;40(June):117–9.
222. dos Passos Cunha M, dos Santos CA, de Lima Neto DF, Schanoski AS, Pour SZ, Passos SD, et al. Outbreak of chikungunya virus in a vulnerable population of Sergipe, Brazil—A molecular and serological survey. J Clin Virol. Elsevier; 2017;97:44–9.
223. Söderlund M, Von Essen R, Haapasaari J, Kiistala U, Kiviluoto O, Hedman K. Persistence of parvovirus B19 DNA in synovial membranes of young patients with and without chronic arthropathy. Lancet. 1997;349(9058):1063–5.
224. Cassinotti P, Burtonboy G, Fopp M, Siegl G. Evidence for persistence of human parvovirus B19 DNA in bone marrow. J Med Virol. 1997;53(3):229–32.
225. Foley B, O'Mahony J, Hill C, Morgan JG. Evidence for persistence of parvovirus B19 DNA in livers of adults. J Med Virol. 2001;65(2):395–401.
226. Enders M, Weidner A, Rosenthal T, Baisch C, Hedman L, Söderlund-Venermo M, et al. Improved diagnosis of gestational parvovirus B19 infection at the time of nonimmune fetal hydrops. J Infect Dis [Internet]. 2008;197(1):58–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18171285
227. Gallinella G, Zuffi E, Gentilomi G, Manaresi E, Venturoli S, Bonvicini F, et al. Relevance of B19 markers in serum samples for a diagnosis of parvovirus B19-correlated diseases. J Med Virol. 2003;71(1):135–9.
228. Hoebe CJPA, Claas ECJ, Steenbergen JE van, Kroes ACM. Confirmation of an outbreak of parvovirus B19 in a primary school using IgM ELISA and PCR on thumb prick blood samples. J Clin Virol. 2002;25(3):303–7.
229. Erdman DD, Usher MJ, Tsou C, Caul EO, Gary GW, Kajigaya S, et al. Human parvovirus B19 specific IgG, IgA, and IgM antibodies and DNA in serum specimens from persons with erythema infectiosum. J Med Virol. 1991;35(2):110–5.
230. Manaresi E, Gallinella G, Zuffi E, Bonvicini F, Zerbini M, Musiani M. Diagnosis and quantitative evaluation of parvovirus B19 infections by real-time PCR in the clinical laboratory. J Med Virol. 2002;67(2):275–81.
231. Woolf AD, Campion G V, Chishick A, Wise S, Cohen BJ, Klouda PT, et al. Clinical manifestations of human parvovirus B19 in adults. Arch Intern Med [Internet]. 1989 May;149(5):1153–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2541666
232. Oliveira SA de, Camacho LAB, Pereira AC de M, Faillace TF, Setubal S, Nascimento JP do. Clinical and epidemiological aspects of human parvovirus B19 infection in an urban area in Brazil (Niterói city area, State of Rio de Janeiro, Brazil). Mem Inst Oswaldo Cruz [Internet]. 2002 Oct;97(7):965–70. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12471422
233. de Freitas RB, Durigon EL, Oliveira D de S, Romano CM, de Freitas MRC, Linhares A da C, et al. The "pressure pan" evolution of human erythrovirus B19 in the Amazon, Brazil. Virology. 2007;369(2):281–7.
234. da Costa AC, Bendit I, de Oliveira ACS, Kallas EG, Sabino EC, Sanabani SS. Investigation of human parvovirus B19 occurrence and genetic variability in different leukaemia entities. Clin Microbiol Infect [Internet]. European Society of Clinical Infectious Diseases; 2013;19(1):E31–43. Available from: http://dx.doi.org/10.1111/1469-0691.12058
235. GIDEON Informatics I, Berger S. Parvovirus B19: Global Status: 2017 edition [Internet]. GIDEON Informatics, Incorporated; 2017. Available from: https://books.google.com.br/books?id=KnIWDgAAQBAJ
236. de Freitas RB, Wong D, Boswell F, de Miranda MFR, Linhares AC, Shirley J, et al. Prevalence of human parvovirus (B19) and rubellavirus infections in urban and remote rural areas in Northern Brazil. J Med Virol. Wiley Online Library; 1990;32(4):203–8.
237. Cubel Garcia R de CN, Pereira RFA, Azevedo KML de, Castro TX de, Mello FCA, Setubal S, et al. Molecular diversity of human parvovirus B19 during two outbreaks of erythema infectiosum in Brazil. Brazilian J Infect Dis [Internet]. Sociedade Brasileira de Infectologia; 2017 Jan;21(1):102–6. Available from: http://dx.doi.org/10.1016/j.bjid.2016.11.002
238. Romano CM, de Matos AM, Araujo ESA, Villas-Boas LS, da Silva WC, Oliveira OM, et al. Characterization of Dengue virus type 2: new insights on the 2010 Brazilian epidemic. PLoS One.

Public Library of Science; 2010;5(7):e11811.
239. Romano CM, Lauck M, Salvador FS, Lima CR, Villas-Boas LS, Araújo ES a., et al. Inter- and Intra-Host Viral Diversity in a Large Seasonal DENV2 Outbreak. López-Galíndez C, editor. PLoS One [Internet]. 2013 Aug 2 [cited 2013 Aug 7];8(8):e70318. Available from: http://dx.plos.org/10.1371/journal.pone.0070318
240. Villabona-Arenas CJ, de Oliveira JL, de Sousa-Capra C, Balarini K, Pereira da Fonseca CRT, Zanotto PM de A. Epidemiological dynamics of an urban Dengue 4 outbreak in São Paulo, Brazil. PeerJ [Internet]. 2016;4:e1892. Available from: https://peerj.com/articles/1892
241. Greenwald P, Bashe Jr WJ. An Epidemic of Erythema Infectiosum. Arch Pediatr Adolesc Med. 1963;107(1):68–72.
242. Oiwa H, Shimada T, Hashimoto M, Kawaguchi A, Ueda T, Sugiyama E, et al. Clinical findings in parvovirus B19 infection in 30 adult patients in Kyoto. Mod Rheumatol [Internet]. 2011 Feb;21(1):24–31. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20680378
243. Anderson MJ, Lewis E, Kidd IM, Hall SM, Cohen BJ. An outbreak of erythema infectiosum associated with human parvovirus infection. J Hyg (Lond) [Internet]. 1984;93(1):85–93. Available from: http://www.journals.cambridge.org/abstract_S0022172400060964
244. Cohen BJ, Buckley MM, Clewley JP, Jones VE, Puttick AH, Jacoby RK. Human parvovirus infection in early rheumatoid and inflammatory arthritis. Ann Rheum Dis. BMJ Publishing Group Ltd; 1986;45(10):832–8.
245. Anderson MJ, Jones SE, Minson AC. Diagnosis of human parvovirus infection by dot-blot hybridization using cloned viral DNA. J Med Virol. Wiley Online Library; 1985;15(2):163–72.
246. Anderson LJ, Tsou C, Parker RA, Chorba TL, Wulff H, Tattersall P, et al. Detection of antibodies and antigens of human parvovirus B19 by enzyme-linked immunosorbent assay. J Clin Microbiol. Am Soc Microbiol; 1986;24(4):522–6.
247. Schwarz TF, Roggendorf M, Deinhardt F. Human parvovirus B19: ELISA and immunoblot assays. J Virol Methods. Elsevier; 1988;20(2):155–68.
248. Nobuo Y, Hiroyuki S, Kohtaro T, Akira Y, Kazuo S. Enzyme-linked immunosorbent assay for IgG and IgM antibodies against human parvovirus B19: use of monoclonal antibodies and viral antigen propagated in vitro. J Virol Methods. Elsevier; 1989;26(2):171–81.
249. Brown KE, Buckley MM, Cohen BJ, Samuel D. An amplified ELISA for the detection of parvovirus B19 IgM using monoclonal antibody to FITC. J Virol Methods. Elsevier; 1989;26(2):189–98.
250. Clewley JP. Polymerase chain reaction assay of parvovirus B19 DNA in clinical specimens. J Clin Microbiol. Am Soc Microbiol; 1989;27(12):2647–51.
251. Salimans MM, Van de Rijke FM, Raap AK, van Elsacker-Niele AM. Detection of parvovirus B19 DNA in fetal tissues by in situ hybridisation and polymerase chain reaction. J Clin Pathol. BMJ Publishing Group; 1989;42(5):525–30.
252. Pattison JR, Jones SE, Hodgson J, Davis LR, White JM, Stroud CE, et al. Parvovirus infections and hypoplastic crisis in sickle-cell anaemia. Lancet. Elsevier; 1981;317(8221):664–5.
253. Rogers BB, Singer DB, Mak SK, Gary SW, Fikrig MF, McMILLAN PN. Detection of human parvovirus B19 in early spontaneous abortuses using serology, histology, electron microscopy, in situ hybridization, and the polymerase chain reaction. Int J Gynecol Obstet. No longer published by Elsevier; 1993;43(2):232.
254. Söderlund-Venermo M, Hokynar K, Nieminen J, Rautakorpi H, Hedman K. Persistence of human parvovirus B19 in human tissues. Pathol Biol (Paris) [Internet]. 2002;50(5):307–16. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12116849
255. (CDC C for DC. Risks associated with human parvovirus B19 infection. MMWR Morb Mortal Wkly Rep. 1989;38(6):81.
256. Weiner CP, Grose CF, Naides SJ. Diagnosis of fetal infection in the patient with an ultrasonographically detected abnormality but a negative clinical history. Am J Obstet Gynecol [Internet]. Mosby; 1993;168(1):6–11. Available from: http://dx.doi.org/10.1016/S0002-9378(12)90876-5
257. Panchaud A, Stojanov M, Ammerdorffer A, Vouga M, Baud D. Emerging role of Zika virus in adverse fetal and neonatal outcomes. Clin Microbiol Rev. Am Soc Microbiol; 2016;29(3):659–94.
258. Calvet G, Aguiar RS, Melo ASO, Sampaio SA, de Filippis I, Fabri A, et al. Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a case study. Lancet Infect Dis. Elsevier; 2016;16(6):653–60.

259. Sarno M, Sacramento GA, Khouri R, do Rosário MS, Costa F, Archanjo G, et al. Zika virus infection and stillbirths: a case of hydrops fetalis, hydranencephaly and fetal demise. PLoS Negl Trop Dis. Public Library of Science; 2016;10(2):e0004517.
260. Cugola FR, Fernandes IR, Russo FB, Freitas BC, Dias JLM, Guimarães KP, et al. The Brazilian Zika virus strain causes birth defects in experimental models. Nature [Internet]. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2016 Jun 9;534(7606):267–71. Available from: http://dx.doi.org/10.1038/nature18296
261. Sanabani S, Neto WK, Pereira J, Sabino EC. Sequence variability of human erythroviruses present in bone marrow of Brazilian patients with various parvovirus B19-related hematological symptoms. J Clin Microbiol. Am Soc Microbiol; 2006;44(2):604–6.
262. Beer M, Conraths FJ, van der Poel WHM. "Schmallenberg virus"--a novel orthobunyavirus emerging in Europe. Epidemiol Infect [Internet]. 2013 Jan [cited 2014 Apr 7];141(1):1–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23046921
263. Grard G, Fair JN, Lee D, Slikas E, Steffen I, Muyembe J-J, et al. A novel rhabdovirus associated with acute hemorrhagic fever in central Africa. PLoS Pathog [Internet]. 2012 Sep [cited 2014 Feb 6];8(9):e1002924. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3460624&tool=pmcentrez&rendertype=abstract
264. Yozwiak NL, Skewes-Cox P, Gordon A, Saborio S, Kuan G, Balmaseda A, et al. Human enterovirus 109: a novel interspecies recombinant enterovirus isolated from a case of acute pediatric respiratory illness in Nicaragua. J Virol [Internet]. 2010 Sep [cited 2013 Aug 22];84(18):9047–58. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2937614&tool=pmcentrez&rendertype=abstract
265. Barbosa CM, Paola N Di, Cunha MP, Rodrigues-Jesus MJ, Araujo DB, Silveira VB, et al. Yellow Fever Virus DNA in Urine and Semen of Convalescent Patient, Brazil. Emerg Infect Dis J [Internet]. 2018;24(1):176. Available from: http://wwwnc.cdc.gov/eid/article/24/1/17-1310
266. Zanotto PM de A, Gould EA. Molecular epidemiology, evolution and dispersal of the genus Flavivirus. Mol Epidemiol Hum viruses. Springer; 2002;167.
267. Gardner CL, Ryman KD. Yellow Fever: A Reemerging Threat. Clin Lab Med [Internet]. 2010;30(1):237–60. Available from: http://www.sciencedirect.com/science/article/pii/S0272271210000028
268. PAHO, WHO. Epidemiological Update: Yellow Fever July 10th, 2017 [Internet]. 2017. Available from: http://reliefweb.int/sites/reliefweb.int/files/resources/2017-jul-10-phe-epi-update-yellow-fever.pdf
269. Beasley DWC, McAuley AJ, Bente DA. Yellow fever virus: Genetic and phenotypic diversity and implications for detection, prevention and therapy. Antiviral Res. 2015;115(February 2016):48–70.
270. Barzon L, Pacenti M, Franchin E, Pagni S, Martello T, Cattai M, et al. Excretion of West Nile Virus in Urine During Acute Infection. J Infect Dis [Internet]. 2013 Oct 1;208(7):1086–92. Available from: http://dx.doi.org/10.1093/infdis/jit290
271. Gourinat A-C, O'Connor O, Calvez E, Goarant C, Dupont-Rouzeyrol M. Detection of Zika Virus in Urine. Emerg Infect Dis [Internet]. 2015 Jan;21(1):84–6. Available from: http://wwwnc.cdc.gov/eid/article/21/1/14-0894
272. Andries A-C, Duong V, Ly S, Cappelle J, Kim KS, Lorn Try P, et al. Value of Routine Dengue Diagnostic Tests in Urine and Saliva Specimens. PLoS Negl Trop Dis [Internet]. Public Library of Science; 2015 Sep 25;9(9):e0004100. Available from: https://doi.org/10.1371/journal.pntd.0004100
273. Fonseca K, Meatherall B, Zarra D, Drebot M, MacDonald J, Pabbaraju K, et al. First Case of Zika Virus Infection in a Returning Canadian Traveler. Am J Trop Med Hyg. 2014;91(5).
274. Villabona-Arenas CJ, de Oliveira JL, Capra CDS, Balarini K, Loureiro M, Fonseca CRTP, et al. Detection of four dengue serotypes suggests rise in hyperendemicity in urban centers of Brazil. PLoS Negl Trop Dis [Internet]. 2014 Feb [cited 2014 May 15];8(2):e2620. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3937275&tool=pmcentrez&rendertype=abstract
275. Morrison TE. Re-emergence of chikungunya virus. J Virol [Internet]. 2014;88(20):11644–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25078691
276. Lazear HM, Diamond MS. Zika Virus: New Clinical Syndromes and its Emergence in the Western

Hemisphere. J Virol [Internet]. 2016;90(10):JVI.00252-16. Available from: http://jvi.asm.org/lookup/doi/10.1128/JVI.00252-16
277. Lapola DM, Martinelli L a., Peres C a., Ometto JPHB, Ferreira ME, Nobre C a., et al. Pervasive transition of the Brazilian land-use system. Nat Clim Chang [Internet]. Nature Publishing Group; 2013 Dec 20 [cited 2014 May 1];4(1):27–35. Available from: http://www.nature.com/doifinder/10.1038/nclimate2056
278. Ferreira Gonçalves MJ, Alecrim WD. Non-planed urbanization as a contributing factor for malaria incidence in Manaus-Amazonas, Brazil. Rev Salud Publica (Bogota) [Internet]. 2004;6(2):156–66. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15382454
279. Hahn MB, Gangnon RE, Barcellos C, Asner GP, Patz J a. Influence of deforestation, logging, and fire on malaria in the Brazilian Amazon. PLoS One [Internet]. 2014 Jan [cited 2014 May 5];9(1):e85725. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3880339&tool=pmcentrez&rendertype=abstract
280. Alves JMP, de Oliveira AL, Sandberg TOM, Moreno-Gallego JL, de Toledo MAF, de Moura EMM, et al. GenSeed-HMM: A Tool for Progressive Assembly Using Profile HMMs as Seeds and its Application in Alpavirinae Viral Discovery from Metagenomic Data. Front Microbiol [Internet]. 2016;7(March):1–15. Available from: http://journal.frontiersin.org/article/10.3389/fmicb.2016.00269
281. Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GGZ, Boling L, et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat Commun [Internet]. Nature Publishing Group; 2014;5:1–11. Available from: http://www.nature.com/doifinder/10.1038/ncomms5498
282. Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S, et al. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. Nat Biotechnol [Internet]. 2014;32(8):822–8. Available from: http://dx.doi.org/10.1038/nbt.2939%5Cnhttp://www.cbs.dtu.dk/databases/CAG/
283. Simon-Loriere E, Faye OO, Faye OO, Koivogui L, Magassouba N, Keita S, et al. Distinct lineages of Ebola virus in Guinea during the 2014 West African epidemic. Nature [Internet]. 2015 Jun 24;524(7563):102–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26106863
284. Li L, Victoria J, Kapoor A, Naeem A, Shaukat S, Sharif S, et al. Genomic Characterization of Novel Human Parechovirus Type. Emerg Infect Dis [Internet]. 2009 Feb [cited 2014 Apr 7];15(2):288–91. Available from: http://wwwnc.cdc.gov/eid/article/15/2/08-0341_article.htm

APPENDIX A – Published Manuscripts

Revista da Sociedade Brasileira de Medicina Tropical 47(6):675-676, Nov-Dec, 2014
http://dx.doi.org/10.1590/0037-8682-0294-2014

Editorial

Where there is no overlap, there is a gap

***Nicholas Di Paola*[1] *and Paolo Marinho de Andrade Zanotto*[1]**

[1]. Laboratório de Evolução Molecular e Bioinformática, Departamento de Microbiologia, Instituto de Ciências Biomédicas II, Universidade de São Paulo, São Paulo, SP.

Over the past 60 years, Brazil has seen extensive rural-urban migration, leading to human settlements encroaching on tropical ecosystems. Persistent human interaction with vectors from diverse tropical environments has led to the emergence of zoonotic pathogens varying in pathogenicity and infectivity. Over time, zoonotic viruses adapt to humans, eventually leading to secondary transmission in some cases. However, most zoonotic pathogens result in asymptomatic and/or dead-end cases, rarely leading to sustained human-human transmission. In very rare cases, a single transmission can explode into an outbreak with sustained networks of human-human transmission. This is the current situation in West African nations, initiated by a single emergence of the Zaire Ebola virus (EBOVZ) in Guéckédou in Guinea's eastern rainforest region[1]. Although EBOVZ is a known virus, this was the first record of a case in this region, surprising unprepared health authorities with no experience of encountering Ebola viruses. Owing to ideal conditions, the 25th Ebola virus outbreak quickly spread to the neighboring countries, Sierra Leone and Liberia. As of December 2, 2014, the Centers for Disease Control (CDC) and World Health Organization (WHO) had conjunctly reported 16,933 cases and 6,002 deaths. Critically, the virus itself is not necessarily the main factor explaining the current high EBOVZ case count, as it is behaving similarly to the ones isolated during the first fully recorded outbreak of EBOVZ, isolated in 1976. This time, the virus is not experiencing restrictions with respect to the availability of susceptible hosts, as it shows high transmission in densely populated metropolitan settings. Accordingly, the basic reproductive rate (R) has been estimated to be around 2.56 in areas with the highest transmission rates, which is comparable to fomite-transmitted viruses such as small pox and influenza[2]. Indeed, as we cannot predict the success of a pathogen or the time or location of emergence, we should focus on what we can control, *i.e.*, setting up efficient health systems to report and isolate cases and contacts and maximizing surveillance efforts and resources to prevent future outbreaks. The arrival of all dengue virus serotypes (DENV-1, -2, -3, and -4), West Nile virus, and more recently, the Chikungunya virus (CHIKV) in the Americas is a reminder that even with a strong healthcare foundation, a novel emerging or re-emerging virus can cause morbidity and mortality in major metropolitan settings if it is not identified and reported quickly.

A recent review in *Revista da Sociedade Brasileira de Medicina Tropical*/Journal of the Brazilian Society of Tropical Medicine discussed the introduction of the Mayaro virus (MAYV) and more recently, CHIKV in Brazil[3]. Both have the potential to cause increased morbidity and mortality in Brazil because of arbovirus infections, and therefore, require our attention.

Originally isolated in the 1950s, CHIKV has migrated through the Caribbean and Americas since then, with imported cases reported in Brazil since 2010. Autochthonous transmission was first recorded in September 2014 in the State of Amapá. CHIKV has already experienced sustained transmission in the States of Bahia and Minas Gerais and may have spread widely already. Both competent *Aedes aegypti* and *Aedes albopictus* mosquito vectors of CHIKV in an urban cycle, and they are widespread across Brazil. Moreover, the three genotypes of CHIKV can be transmitted efficiently by all 35 strains of *Aedes aegypti* and *Aedes albopictus*, justifying the suggestion that outbreaks in other states are inevitable[3].

Mayaro virus is another pathogen responsible for sporadic cases in Brazil, mainly in the Amazonian region. With MAYV residing in primate reservoirs, *Haemagogus* and *Aedes aegypti* mosquitos initiate transmission via a pseudo-sylvatic cycle, similar to yellow fever[3]. Although there have been no large outbreaks, health authorities should preemptively prepare for a re-emergence in unexpected localities.

Symptoms of CHIKV and MAYV are similar to those of other common infectious agents, such as the dengue virus (DENV), and less common agents, such as Rocio, Oropouche, Saint Louis encephalitis, and yellow fever viruses[4]. Importantly, the availability of rapid and efficient assays precisely identifying viral agents is crucial to diagnose patients and analyze epidemiological data. In addition, for medical personnel to follow proper treatment protocols, diagnostics should be readily available, accurate, sensitive, and economical to differentiate between these agents without a *coincidental* assumption (*i.e.*, during a dengue outbreak, patients with similar symptoms may be infected by other arboviruses but not subjected to formal serological assay). Interestingly, this mirrors initial diagnostic scenarios in Guinea. According to the CDC, in the early stages

Address to: Dr. Paolo Marinho de Andrade Zanotto. LEMB/ICB II/USP. Av. Prof. Lineu Prestes 1374, 05508-000 São Paulo, SP, Brasil.
Phone: 55 11 3091-8453; **Fax:** 55 11 3091-7354
e-mail: pzanotto@usp.br
Received 3 December 2014
Accepted 9 December 2014

of Ebola virus (EBOV) infection, patients could present with fever, severe headache, fatigue, muscle pain, vomiting, diarrhea, abdominal pain, or unexplained hemorrhaging. In West Africa, this symptom collection could indicate a number of differential diagnoses including malaria, typhoid fever, leptospirosis, rickettsiosis, African trypanosomiasis, Lassa fever, cholera, Marburg virus, and Crimean-Congo hemorrhagic fever[5]. Therefore, in the early stages of an outbreak, lack of positive confirmatory serological tests could delay health authorities' responses, depending on the disease severity. Logically, health structure, resources, and sociological qualities differ between West African nations and the Americas. However, we must continuously remind ourselves that a similar scenario can occur in Brazil. Are Brazil's diagnostic and response capabilities prepared for the emergence of a novel pathogen?

It has previously been argued that many arbovirus infections have been neglected in Brazil, with few epidemiological studies conducted to understand their distribution. For instance, dengue fever is difficult to distinguish from other acute febrile illnesses, which emphasizes the importance of a differential diagnosis[4-6]. Indeed, most diagnoses in the country are made on clinical and epidemiological grounds. In 2014 (as of November 2), 547,612 clinical cases of dengue fever were reported, but 8,423 (1.54%) were laboratory confirmed[7,8]. These numbers do not represent the countless mildly symptomatic or nonspecific febrile cases. Therefore, the true rates of infection and transmission are underestimated, and the possibility of infection from other pathogens is ignored. With respect to the CDC Ebola case definition, DENV, CHIKV, and MAYV could cause overlapping symptoms with the above diagnoses, supporting the need for broad-spectrum, sensitive, and economic diagnostic assays to identify infectious agents without concerns regarding cross-reaction of closely related agents. This does not imply that an Ebola outbreak is a given in Brazil but exemplifies a complication that could arise in such a scenario.

The emergence of a novel pathogen should always be preconsidered. Moreover, if it has the potential for establishing itself within a human transmission network, having diagnostic tools to confirm its presence and prompt notification of national health authorities should be priorities. Depending on the pathogen's transmissibility, contact tracing and quarantine should be established rapidly. Although EBOV is not novel in the sense of being unknown, it was novel in the locality of its zoonotic emergence in West Africa. Once a country undergoes a zoonotic-related emergence event leading to morbidity and mortality, the aftermath leads to preemptive measures to ensure that the scale of the next outbreak is significantly diminished. Continuous sentinel and active surveillance are costly and time consuming from a resource perspective. However, considering resource limitations, early diagnostic efforts and surveillance of emerging pathogens are the best ways to reduce mortality, morbidity, and the enormous health care costs associated with epidemics[9].

Brazil has thousands of communities surrounded by tropical ecosystems teeming with potentially pathogenic agents. Are there countries with limited resources whose healthcare system have successfuly prevented large outbreaks efficiently and strategically, from which we can learn? Uganda could be a good model; similar in respect to climate and ecological diversity, also limited in resources disposed towards surveillance and epidemic management. Uganda presented a model system of emerging infectious disease control, demonstrating from previous EBOV outbreaks that continuous surveillance, that dedicated teams, devoted to contact tracing, case isolation, and cooperating with local communities and bordering countries, successfully limits case counts and spread[10]. Brazil and other nations could learn from Uganda's experience, applying the approach to CHIKV, MAYV, DENV, and other pathogens. Although our ability to diagnose, treat, trace, and contain pathogen outbreaks has improved over the last decade, it has not been sufficient to prevent large outbreaks of new serotypes of dengue, and there is no indication that CHIKV will be halted. In any case, one has to conceive that what is ongoing in West Africa could happen in Brazil. *Where there is no overlap, there is a gap;* therefore, we should establish much-needed constructive interactions between governments, academia, and the private sector to continue searching for novel methods of diagnostics, improve surveillance, strategize, and maximize resources to prepare for any future emergence. The question is not one of whether an emergence will occur but when and whether we will be ready.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

1. Gatherer D. The 2014 Ebola virus disease outbreak in west Africa. J Gen Virol 2014; 95 (Pt 8):1619-1624.
2. Althaus CL. Estimating the Reproduction Number of Ebola Virus (EBOV) During the 2014 Outbreak in West Africa. PLoS Currents Outbreaks 2014; doi: 10.1371/currents.outbreaks.91afb5e0f279e7f29e7056095255b288. Abstract
3. Figueiredo MLG, Figueiredo LTM. Emerging alphaviruses in the Americas: Chikungunya and Mayaro. Rev Soc Bras Med Trop 2014; 47: 677-683.
4. Figueiredo LTM. Saint Louis encephalitis virus and other arboviruses in the differential diagnosis for dengue. Rev Soc Bras Med Trop 2014; 47:541-542.
5. West TE, von Saint André-von Arnim A. Clinical Presentation and Management of Severe Ebola Virus Disease. Ann Am Thorac Soc 2014; 11:1341-1350.
6. Messina JP, Brady OJ, Scott TW, Zou C, Pigott DM, Duda KA, et al. Global spread of dengue virus types: mapping the 70 year history. Trends Microbiol 2014; 22:138-146.
7. Teixeira MG, Siqueira JB, Ferreira GLC, Bricks L, Joint G. Epidemiological trends of dengue disease in Brazil (2000-2010): a systematic literature search and analysis. PLoS Negl Trop Dis 2013; 7:e2520.
8. Pan American Health Organization. Number of Reported Cases of Dengue and Andean Figures for 2014 (to week noted by each country). [Internet]. [Updated 2014 November 2]; [Cited 2014 December 3]. Available at: http://www.paho.org/.
9. Daszak P, Lipkin WI. The search for meaning in virus discovery. Curr Opin Virol 2011; 1: 620-623.
10. Mbonye AK, Wamala JF, Nanyunja M, Opio A, Makumbi I, Aceng JR. Ebola viral hemorrhagic disease outbreak in west Africa-lessons from Uganda. Afr Health Sci 2014; 14:495-501.

PLOS | NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

Biological and phylogenetic characteristics of West African lineages of West Nile virus

Gamou Fall[1☯], Nicholas Di Paola[2☯], Martin Faye[1], Moussa Dia[1], Caio César de Melo Freire[3], Cheikh Loucoubar[4], Paolo Marinho de Andrade Zanotto[2]*, Ousmane Faye[1], Amadou Alpha Sall[1]

1 Pôle de Virologie, Unité des Arbovirus et virus des fièvres hémorragiques, Institut Pasteur de Dakar, Dakar, Sénégal, **2** Laboratory of Molecular Evolution and Bioinformatics, Department of Microbiology, Biomedical Sciences Institute, University of Sao Paulo, Sao Paulo, Brazil, **3** Department of Genetics and Evolution, Federal University of Sao Carlos, Sao Carlos, Brazil, **4** Groupe à 4 ans de Biostatistiques, Bioinformatique et modélisation, Institut Pasteur de Dakar, Dakar, Sénégal

☯ These authors contributed equally to this work.
* pzanotto@usp.br

Citation: Fall G, Di Paola N, Faye M, Dia M, Freire CCdM, Loucoubar C, et al. (2017) Biological and phylogenetic characteristics of West African lineages of West Nile virus. PLoS Negl Trop Dis 11 (11): e0006078. https://doi.org/10.1371/journal.pntd.0006078

Editor: David W.C. Beasley, University of Texas Medical Branch, UNITED STATES

Received: March 16, 2017

Accepted: October 27, 2017

Published: November 8, 2017

Data Availability Statement: All relevant data are within the paper and its Supporting Information files. All new sequences are available from the Genbank nucleotide database (accession numbers KY703854-KY703856). Sequence alignments can be found at https://github.com/caiofreire. All Genbank accession numbers used in this study are listed in the Materials and Methods section.

Funding: This work has been financially supported by Institut Pasteur de Dakar, and funded by EU grant HEALTH.2010.2.3.3-3-261391 EuroWestNile

Abstract

The West Nile virus (WNV), isolated in 1937, is an arbovirus (arthropod-borne virus) that infects thousands of people each year. Despite its burden on global health, little is known about the virus' biological and evolutionary dynamics. As several lineages are endemic in West Africa, we obtained the complete polyprotein sequence from three isolates from the early 1990s, each representing a different lineage. We then investigated differences in growth behavior and pathogenicity for four distinct West African lineages in arthropod (Ap61) and primate (Vero) cell lines, and in mice. We found that genetic differences, as well as viral-host interactions, could play a role in the biological properties in different WNV isolates *in vitro*, such as: (*i*) genome replication, (*ii*) protein translation, (*iii*) particle release, and (*iv*) virulence. Our findings demonstrate the endemic diversity of West African WNV strains and support future investigations into (*i*) the nature of WNV emergence, (*ii*) neurological tropism, and (*iii*) host adaptation.

Author summary

The West Nile virus (WNV) can cause severe neurological diseases including meningitis, encephalitis, and acute flaccid paralysis. Differences in WNV genetics could play a role in the frequency of neurological symptoms from an infection. For the first time, we observed how geographically similar but genetically distinct lineages grow in cellular environments that agree with the transmission chain of West Nile virus—vertebrate-arthropod-vertebrate. We were able to connect our *in vitro* and *in vivo* results with relevant epidemiological and molecular data. Our findings highlight the existence of West African lineages with higher virulence and replicative efficiency *in vitro* and *in vivo* compared to lineages similar to circulating strains in the United States and Europe. Our investigation of four West African lineages of West Nile virus will help us better understand the biology of the virus and assess future epidemiological threats.

and the Conselho Nacional de Pesquisa (CNPq) project 441105/2016-5. We also thank the Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) for the funding (project #2013/22136-1 and #2014/17766-9). 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.

Introduction

West Nile virus (WNV) is a member of the Japanese Encephalitis virus (JEV) serocomplex and is a part of the genus *Flavivirus* of the family *Flaviviridae*. The WNV is a single-stranded, positive-sense RNA virus. The genomic RNA is about 11 kilobases (kb), containing one long open reading frame (ORF) flanked by 2 non-coding regions. This ORF encodes for a polyprotein, which is processed into three individual structural (Capsid, pre-Membrane, Envelope), and seven non-structural (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) proteins [1–4].

West Nile fever disease (WN fever) is caused by the WNV. WN fever in humans can range from asymptomatic infections or mild acute febrile illness, to neurological diseases including meningitis, encephalitis, and acute flaccid paralysis [5–7]. WNV's host range is extensive: it has been detected in over 65 species of mosquitoes and ticks, 225 species of birds, and 29 different animals [8,9]. A human vaccine or specific antiviral treatment for WN fever is currently unavailable.

WNV was first discovered and isolated from the blood of a woman suffering from febrile illness in 1937 in Uganda [10]. Cases of WN fever were documented in Israel and Egypt in the early 1950s, France in the 1960s, and South Africa in the 1970s [11]. The global awareness of WN fever increased in the 1990s, as sporadic and major outbreaks occurred, primarily in the Mediterranean Basin and occasionally in Europe [5]. In 1999, WNV unexpectedly emerged in New York City, signifying the first confirmed incidence of WNV in the Western Hemisphere. Since then, WNV has spread throughout the Americas, causing over 20,265 cases of neurological disease and 1,783 case fatalities in humans and even higher rates of mortality among birds in the United States [12–16].

Meanwhile, WNV continued to spread and cause WN disease and encephalitis in Europe, Asia, and Oceania [17]. In the 1990s, the largest outbreaks occurred in Romania in 1996 [18], and Russia in 1999 [19], with 17 and 40 human fatalities, respectively. In the 21st century, emergences of WN fever and encephalitis have been reported in Europe [20], with a hallmark human neurological outbreak in Greece in 2010 [21], and several noteworthy outbreaks in Italy [22–24], Hungary [25] and Serbia [26].

WNV is biologically diverse; up to nine lineages have been proposed [27–30]. However, most human outbreaks of WN encephalitis have been attributed to lineages 1 and 2. Lineage 1 is globally spread and exists in distinct clades. Clade 1a comprises of strains isolated from Europe, Africa, and the Americas. Clade 1b, also referred to as Kunjin virus, has been restricted to Oceania. Major outbreaks in Europe, Africa, and the Americas with neurological diseases are caused by strains belonging to lineage 1, with an exception to clade 1b where neurological disease is rarely reported [12,31,32].

Lineage 2 was exclusively reported in Africa up until 2004, until it was isolated from humans and bird populations in Hungary, Greece, and Italy [4,21,23,33]. Lineage 2 was also considered to be less pathogenic than lineage 1, until it caused severe disease in South Africa and encephalitis among birds and humans in Europe [4,21,23,33,34]. Both lineages include strains with varying degrees of neuroinvasiveness in humans [35].

Besides lineages 1 and 2, there are lineages that are less widespread. Lineage 3, also referred to as Rabensburg virus, was repeatedly isolated in the Czech Republic [36–38]. Lineage 4 has been isolated and reported from Russia [39]. The 5th lineage was isolated from India, and is often identified as a distinct clade of lineage 1 (clade 1c) [40]. A putative 6th lineage, based on a small gene fragment, has been described from Spain [27,41].

Koutango virus (lineage 7) was initially classified as a different virus, but is now a distinct lineage of WN virus [31,42]. Lineage 7 strains were isolated from ticks (this study) and rodents, a rare feature among WN virus lineages [4]. The Koutango strain virus has also been shown to

have a higher virulence than the lineage 1a strain "NY99" in mice [43,44]. Although there was a report of an accident where a Senegalese lab worker was symptomatically infected with the Koutango strain, a natural human infection has yet to be confirmed [45]. Additionally, a new lineage (putative lineage 8) of WNV was isolated from *Culex perfuscus* in Kedougou, Senegal in 1992 [4]. Finally, a putative 9th lineage, or sublineage of lineage 4, was isolated from *Uranotaenia unguiculata* mosquitoes in Austria [27].

Despite the presence of lineages 1, 2, 7 (Koutango) and a putative 8th lineage circulating in Africa [4,46,47], WNV has had minor impact on human health. Sporadic outbreaks were observed in several African counties [48–50], with lower frequencies of neurological disease than that reported from outbreaks in the USA [51,52]. For example, Senegal has never had a major outbreak of WN fever, but was the source of several endemic genotypes that were identified and sequenced. Moreover, in Senegal, WNV antibody seroprevalence has been around 80% in sampled humans, horses, and birds [53–57].

A recent study on the vector competence of African WNV lineages demonstrated that local mosquito populations lack efficient transmission of WNV [4]. Besides vector competence—*i.e.* intrinsic genetic variations among lineages—host adaptation, movement of host populations, climate and ecological factors could play a role in viral replication, virulence, and the outcome of infection. The N-linked glycosylation site of the envelope protein may be associated with differences observed in: (*i*) WNV neuroinvasiveness in mice, (*ii*) viral replication, and (*iii*) transmission of WNV in mosquitoes [4,58–60]. In this regard, Senegal has been a focal point in the studies of WNV virus, where multiple lineages of WNV are co-circulating endemically, but whose biology remains poorly understood.

To address these questions, we analyzed complete coding regions (polyproteins) of four different lineages circulating in Senegal and West Africa. Using additional WNV sequences from Genbank, we performed a phylogenetic analysis using the complete polyprotein sequences of the viruses and investigated sites for positive selection. We also analyzed the biological properties of these 4 WNV lineages using *in vitro* and *in vivo* models. Ultimately, understanding the relationships among ecological and genetic differences will ameliorate our understanding of WNV emergence, epidemiology, and its maintenance in nature.

Results

Full-length polyprotein sequencing

In this study, three complete polyprotein genes from Senegal isolates were sequenced: ArD76986, ArD96655, and ArD94343 (Table 1). These novel sequences are representative of lineages 1, 7 (Koutango) and 8 (putative), respectively. The lineage 1 and lineage 8 strains were isolated from *Culex* mosquito species, while the lineage 7 strain was isolated from a tick species.

Acknowledging previous works that have reconstructed the evolutionary history and those that have characterized novel isolates and lineages of WNV, we included seven additional

Table 1. Strains of West Nile virus used in this study.

Strain	Lineage	Place of isolation	Year of isolation	Isolation source	Number of passages	Passage history[a]	Accession number
ArD76986	1	Senegal	1990	*Culex poicilipes*	10	Ap4NBM3Ap3	KY703854
B956	2	Uganda	1937	Human	11	Ap4NBM3Ap4	AY532665
ArD96655	7/ Koutango	Senegal	1993	*Rhipicephalus guihoni*	8	Ap4NBM3Ap1	KY703855
ArD94343	8	Senegal	1992	*Culex perfuscus*	12	Ap4NBM3Ap5	KY703856

[a] Ap4NBM2Ap4 is equivalent to 4 serial passages in Ap61 (Ap) followed by 2 passages in newborn mice (NBM) followed by 4 serial passages in Ap61.

https://doi.org/10.1371/journal.pntd.0006078.t001

complete ORF sequences to compare differences at the gene and protein level (Fig 1). Among representative sequences, the average nucleotide pairwise identity is 77.6% (s.d. = 4.1%) and the amino acid average pairwise identity is 90.1% (s.d. = 3.3%). When comparing individual sequences, the NY99 strain (Accession number: AF196835, lineage 1a, United States 1999) shared a 99.5% pairwise identity to ArD76986 (Accession number: KY703854, lineage 1a, Senegal 1990) at the amino-acid level (Fig 1A). The sequence diversity of endemic WNV lineages in Senegal (SN) is notable, as the lineage 1 strain (ArD76986) was 88.9% and 90.9% identical at the amino-acid level to the ArD96655 (Accession number: KY703855, lineage 7, SN 1993) and the ArD94343 (Accession number: KY703856, lineage 8, SN 1992) strains respectively.

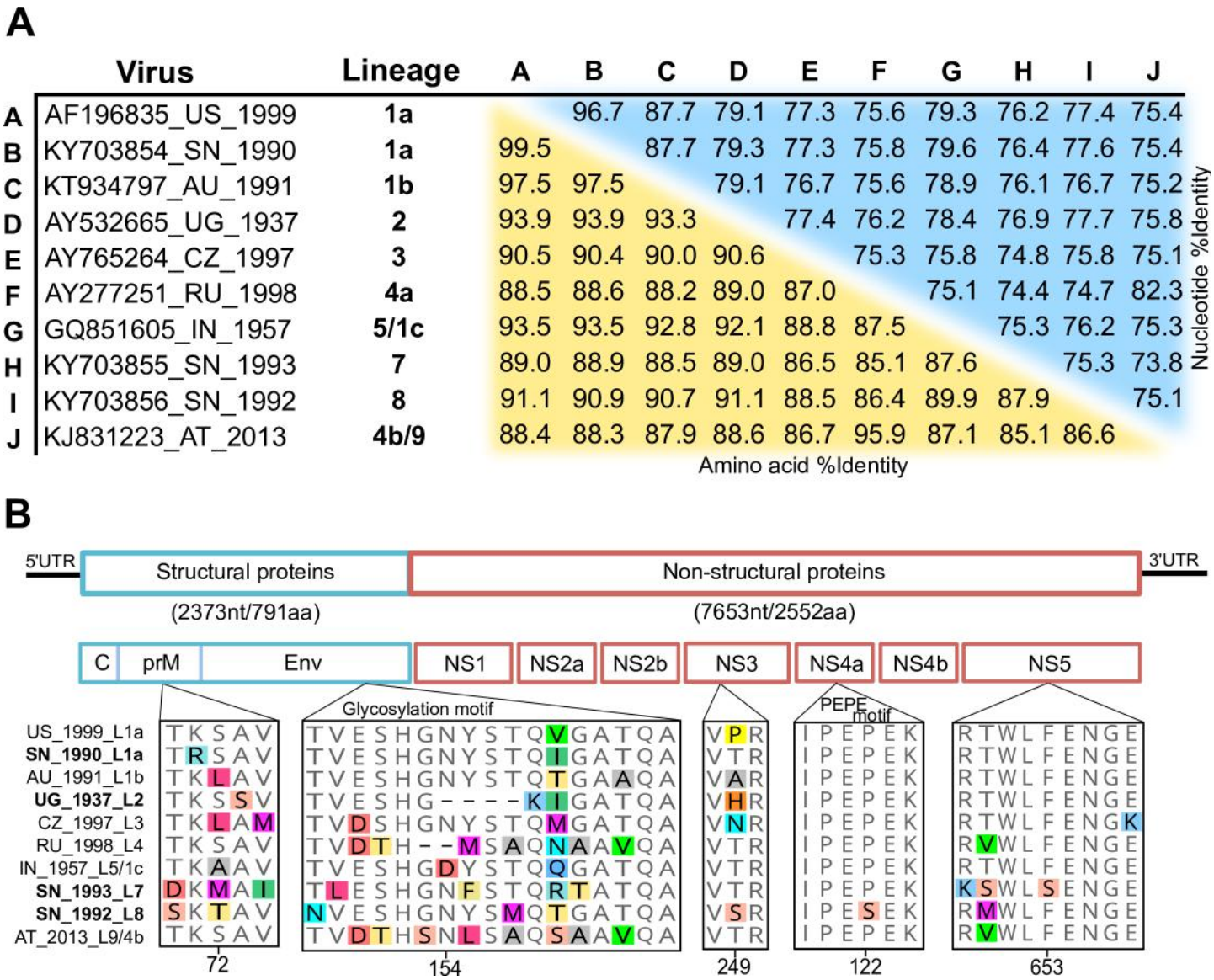


Fig 1. The genetic diversity of the West Nile virus lineages. A) Pairwise percent identity between nucleotide (blue) and amino acid (orange) sequences of the polyprotein. Sequences are labeled in the following format: accession number, 2-letter country code, and year of isolation. B) Genomic structure of West Nile virus with genes labeled. Alignments of known virulence motifs are shown. Codons of special interest are labeled by their position at each individual protein sequence and not by their position in the polyprotein. Sequences are labeled by country, year of isolation and phylogenetic lineage.

https://doi.org/10.1371/journal.pntd.0006078.g001

Between the lineage 7 and lineage 8 strains, the amino-acid pairwise identity was 87.9% (Fig 1A).

The 1937 WNV isolate of strain B956 (Accession number: AY532665, lineage 2, Uganda 1937) is of particular interest, as it is the oldest clinical isolate available with a complete ORF sequenced. Amino acid pairwise identity was 93.9% to the NY99 strain sequence and to the lineage 1a (SN) sequence, 89% to the lineage 7 strain sequence and 91.1% to the lineage 8 strain sequence. At the nucleotide level, the lineage 2 strain (UG) had a pairwise identity of 79.1% to the NY99 sequence, 79.3% to the lineage 1a strain (SN) sequence, 76.9% to the lineage 7 strain sequence and 77.7% to the lineage 8 strain sequence. Additionally, B956 contains a 12 base pair deletion at nucleotide position 1,331, corresponding to the WNV envelope glycosylation site.

We compared published sequences and published works to identify whether mutations that have been shown to influence WNV virulence and replication were present in the newly sequenced open-reading frames. For each strain, we discovered amino acid changes that were associated to a phenotypical change and many additional mutations with unknown consequences (Fig 1B). For example, the 22^{nd} and 72^{nd} codon sites of the pre-membrane protein (prM) have been shown to play a role in enhancing the virulence and particle secretion in WNV [61]. At this site, we found alterations in the lineage 7 strain (SN_1993_L7), and the lineage 8 strain (SN_1992_L8).

Another example is the glycosylation site found in the $154\text{-}156^{th}$ positions of the envelope (Env) protein, which is considered a virulence factor [62]. We found that the lineage 1 strain from Senegal (SN_1990_L1a), the Kunjin strain (AU_1991_L1b), the NY99 strain (US_1999_L1a) and the lineage 8 strain (SN_1992_L8) harbored the NYS motif while other strains had variations or deletions in this locus. Next, the 249^{th} codon position of the NS3 protein [63], the helicase protein, was found to increase viremia and virulence in birds, and could play a role in other hosts. We observed several variations in our data at the 249^{th} codon position (Fig 1B).

Additionally, changes in the highly conserved ^{120}P-E-P-E^{123} region of the NS4A protein can attenuate or even impair virion replication and release [64], which we found present in the lineage 8 strain. Finally, a mutation in the NS5 protein, serine (S) to phenylalanine (F) at the 653^{rd} position in the NS5 protein, is associated with an increased resistance to interferon [65], a mutation that is shared by the lineage 7 strain (SN_1993_L7) (Fig 1B). We also found several synonymous changes in positions corresponding to known virulence motifs, such as variability in the third codon site position (the wobble base) during the translation of serine (S) at the 156^{th} codon site. We also investigated sites within the NS2A [66], NS4B [67], and additional sites within the NS5 region that are known to impact on infectivity and virulence [65], but no mutations were present in our sequences.

Phylogeny of West Nile virus

The phylogenetic analysis revealed a similar topology to the ones obtained from previous maximum likelihood trees [27,40,42,68,69]. Currently, up to 9 distinct lineages have been suggested.

A total of 95 sequences, including 3 novel polyprotein sequences from Senegalese isolates (Table 1), were used to estimate a maximum-likelihood tree with FastTree (S1 Fig) and a very similar relaxed clock Bayesian maximum-clade credibility (MCC) tree (Fig 2), summarizing the MCMC runs with BEAST. The MCC tree was scaled to time (years) and branch tip-nodes were colored to identify previously classified lineages [27]. Here, the time to the most recent common ancestor (tMRCA) with its corresponding 95% highest posterior density (HPD)

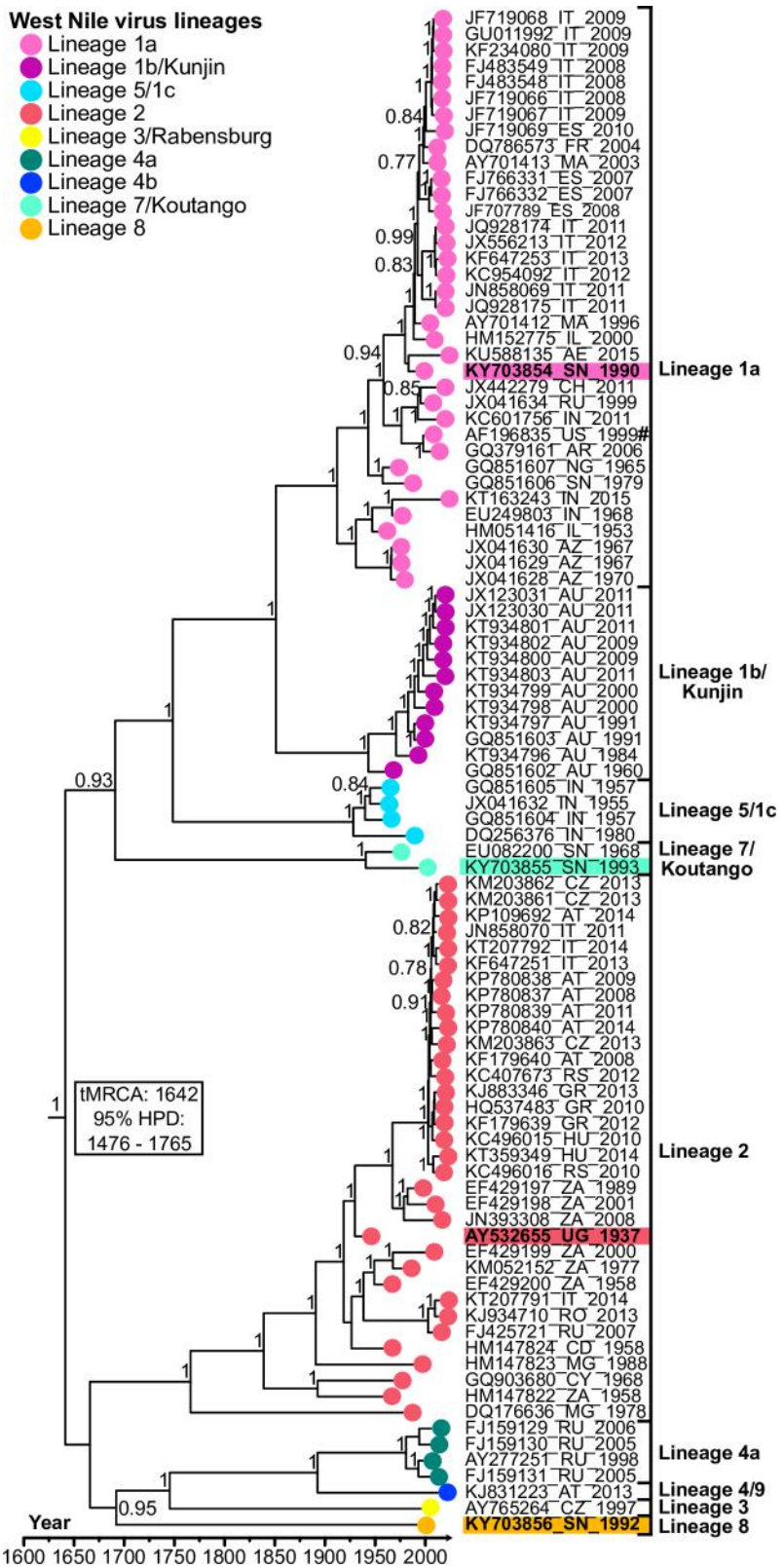


Fig 2. Bayesian maximum clade credibility tree estimating the phylogenetic relationships of West Nile virus. Tree nodes with a posterior probability greater than 0.7 are displayed. Tree tip nodes are colored by proposed lineage and for visual clarity. For each sequence, the two-letter code representing a country of isolation is included in the sequence label. Branches are scaled in years before 2015. # NY99 strain.

https://doi.org/10.1371/journal.pntd.0006078.g002

interval for WNV was estimated in the unit of years. The tMRCA of WNV is predicted to have originated in the late 16th/early 17th century (95%HPD: 1476–1765), a major split that diverges lineages 1, 5 and 7 from lineages 2, 3, 4, 8, and 9. Both lineage 1 and 2 show multiple introductions into Europe and other *New World* countries. Additionally, we see that lineages 1, 2, 7, and 8 have been isolated in West Africa, yet only lineages 1 and 2 have emigrated.

Growth kinetics

Infection, viral proliferation, and virulence in each cell type were measured by 4 different tests over a 146 hours post-infection period: quantitative reverse transcriptase PCR (qRT-PCR) of the lysed cell fracture to measure genome replication (Fig 3A and 3B), qRT-PCR of the supernatant fraction to detect genome replication dynamics (*i.e.*, total number of particle release) (Fig 3C and 3D), immunofluorescence staining of the cells to visualize the infectivity of cells and estimate protein translation efficiency (Fig 3E and 3F), and plaque assays to determine the amount of infectious viral particles (PFU/ml) from the supernatant fraction (Fig 3G and 3H). Using Ap61 and Vero cells, our goal was to replicate the biology of WNV in a mosquito vector and its vertebrate host.

We found that African lineages have different growth dynamics in mosquito and mammalian cell lines. In *Aedes pseudoscutellaris* cells, growth dynamics were similar for all lineages, (Fig 3, left column) where all lineages exhibited successful replication and generation of infectious particles. In Vero cells (Fig 3, right column), lineages 1, 2, and 7 showed exceptional growth, with lineage 2 strain exhibiting the highest replication and particle release capabilities, and lineage 7 strain having exceptional translational dynamics and highest PFU/ml during the infection interval.

We observed cell-specific growth differences among different WNV strains. For example, Fig 3A and 3B showed differences in genome replication dynamics in the cells with respect to host cells. Interestingly, lineage 1 strain had higher genome replication in Ap61 cells (p-value ranging from $2.22x10^{-16}$ to 0.002) while the lineage 2 strain had higher genome replication in Vero cells (statistically comparable). Lineage 8 showed a lower significant replication profile in Vero cells (p-value ranging from $8.81x10^{-13}$ to 0.031). Furthermore, differences in growth at T_0 further supports that WNV lineages could have a preference to a specific cellular environment. The rate of viral attachment, entry and replication initiation can all depend on the genetics of the infecting strain [70].

We estimated the total number of released particles at different times post infection by measuring the WNV RNA copy number in the cell supernatant. All tested lineages had comparable genome copy numbers in Ap61 supernatants (Fig 3C). However, we found a significantly higher copy number of total particles released for the lineage 2 strain at 22, 28, and 50 hours post-infection (hpi) in both *in vitro* models (p-value ranging from $2.22x10^{-16}$ to 0.023). Lineage 8 strain showed significantly lower genome copy numbers in Vero supernatants (p-value ranging from $8.81x10^{-13}$ to 0.031).

Next, we approached differences in protein translation efficiency between lineages by detecting viral proteins using an immunofluorescence assay (IFA) (Fig 3E and 3F). The lineage 7 strain displayed more efficient protein translation in both cells (p-value ranging from $3.98x10^{-13}$ to 0.011), while lineage 8 strain had significantly lower levels of protein translation in Vero cells. Nevertheless, the translation rate in lineage 8 increased significantly from $T_{124-146}$ in Ap61 cells. We also noticed a delay on translation detection in both cells, with no detectable protein production until T_{99} hours (Fig 3E) and T_{50} hours (Fig 3F) respectively.

To quantify the infectious particles of different WNV strains, we used plaque assays to estimate the amount of infectious viral particles (PFU/ml) in the supernatant fractions. In Ap61

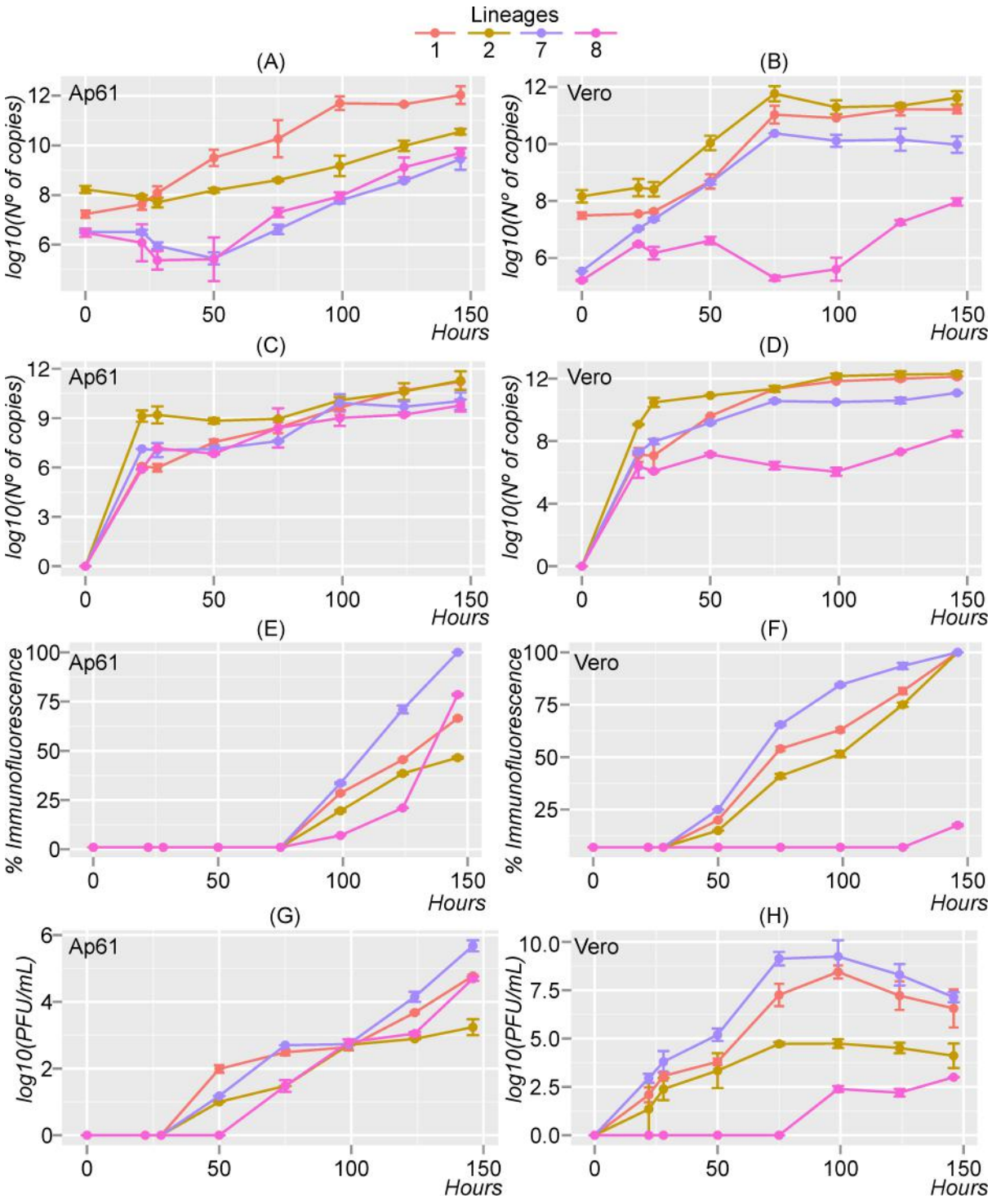


Fig 3. Growth kinetics of West African West Nile virus strains. The strain lineage label is in reference to the strains in Table 1. Figs A-D show the amount of viral RNA equivalents isolated from cells (A and B) and supernatant (C and D) (log_{10} of RNA copy number), the percent (% immunofluorescence) of cells infected (E and F) and the number of infectious viral particles (G and H) (log_{10}PFU/ml) over a 146-hour post-infection time period. The experiments were performed with Ap61 cells (left column) and Vero cells (right column). The error bars indicate the range in values of two independent experiments.

https://doi.org/10.1371/journal.pntd.0006078.g003

cells, we found a similar profile of infectious particles production for all lineages, with significant higher rates at 124 hpi and 146 hpi for the lineage 7 strain (p-value ranging from $2.22x10^{-16}$ to 0.028) (Fig 3G). In Vero cells, lineage 1 and lineage 7 strains had higher number of PFU/ml, while lineage 2 had an intermediate profile and lineage 8 had the lowest amount of infectious viral particles, with significant differences from 28 to 124 hpi (p-value ranging from $2.22x10^{-16}$ to 0.049) (Fig 3H).

Finally, we approximated the replication efficiency by finding the ratio of the number of virions released in the supernatant–particles that completed the infectious cycle–divided by the number of plaque forming units (PFU) [71–73]. We estimated the ratio for each strain in each cell and found significant differences in replication efficiency (p-value ranging from $5.49x10^{-16}$ to 0.0223) (Fig 4). There are some consistencies with Fig 3, where the lineage 7 strain was the most efficient and the B956 strain was the least efficient *in vitro*. Lineage 1 and lineage 7 strains seem to be more cell-specific; both replicated less efficiently in Ap61 cells.

Virulence and survival of West Nile virus *in vivo*

To determine the virulence of WNV strains (Table 1), we challenged five- to six-week-old mice with three different viral doses and observed their overall survival for 21 days. Depending on the strain and dose used, several mice developed clinical disease and died (Table 2). Clinical signs included tremors, reduced activity and reluctance to move, hind leg paralysis and closed eyes. The PBS-inoculated control groups exhibited no signs of disease throughout the experiment.

The lineage 7 strain was the most virulent of the strains at all administered doses (Wilcoxon rank sum test, p-values < 0.05). In fact, the lineage 7 strain induced the shortest survival time compared to the other strains and always resulted in 100% mortality in every experiment (Fig 5 and Table 2). Interestingly, in most cases, mice inoculated with the lineage 7 strain died without showing any clinical signs.

Comparatively, mice inoculated with the lineage 1 and lineage 2 strains usually showed signs of disease at least 1 day before dying. However, lineage 8 showed no virulence (100% survival) at 100 and 1000 PFU doses (Fig 5B and 5C). In fact, only one mouse mortality was observed at 10000 PFU (Fig 5A and Table 2).

Selection

To determine the evolutionary pressures acting on the WNV ORF, we estimated the ratio of nonsynonymous (dN) to synonymous (dS) substitutions per codon site (where dN—dS > 0, signifies positive selection) using 95 sequences, which represent all investigated WNV lineages. Our investigation on selection regimens acting on all WNV complete ORF sequences—with the FUBAR method—revealed 3313 well supported (posterior probability ≥ 0.9 and Bayes Factor < 3.0) sites under purifying selection (S1 Table and S2 Fig). However, we found 95 statistically significant sites (p-value ≤ 0.1) under diversifying episodic selection (S2 Table and S3 Fig), using MEME method.

Discussion

Despite the presence of at least four different lineages in West Africa, there has never been a major outbreak, nor a large frequency of encephalitic cases connected with WNV. The lack of a WN disease "burden" within Senegal could suggest that WNV is endemic, which could explain the high seroprevalence, and therefore, few susceptible hosts [54]. However, the threat of WNV emerging to places where the population's seroprevalence is much lower or even naive is a serious concern. Avian migratory routes could have played a role in the emigration

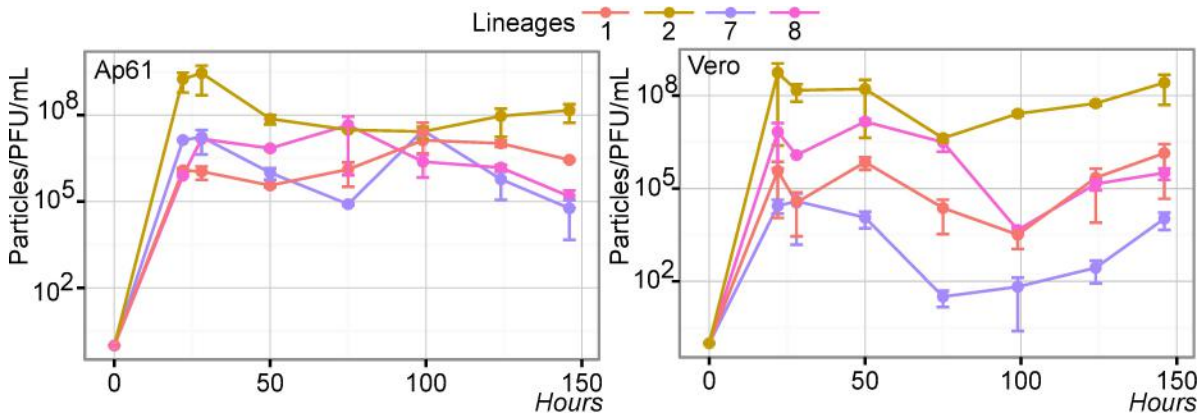


Fig 4. Replication Efficiency of West Nile virus *in vitro*. Replication efficiency of West African strains in Ap61 and Vero cell lines over 146 hour post-infection period. The error bars indicate the range in values of two independent experiments.

https://doi.org/10.1371/journal.pntd.0006078.g004

of WNV strains from Africa [53], and for other African-borne arboviruses such as Usutu virus [74]. The extensive genetic diversity (Fig 1A) and broad host range of WNV [8,9] could have also contributed to its global dissemination (Fig 2), as certain mutations have been previously prosecuted with lineage 1's entrance in the United States [32] and lineage 2's emergence in Europe [75]. As a consequence, several groups have investigated how specific genetic changes and selective pressures within the WNV ORF can affect the phenotypical behavior of a WNV strain.

In our study, the growth kinetics of the different West African WNV lineages were explored in *Aedes pseudoscutellaris* (Ap61) and African green monkey kidney cells (Vero) to reflect infection dynamics in two common classes of WNV hosts (insect vector and primate) (Figs 3 and 4). The virulence of these lineages in mice was also analyzed (Fig 5 and Table 2). We found that these 4 West African lineages have significant differences in their ability to proliferate in our tested cell lines and their degree of virulence in mice (Figs 3, 4 and 5). We also explored how our *in vitro* and *in vivo* results could be explained by their evolutionary (Fig 2) and individual genetic variations (Fig 1).

Table 2. Mice mortality and virulence of West Nile virus *in vivo* of 5- to 6-week-old Swiss mice observed for 21 days.

Strain	Lineage	Viral dose (PFU)	dead/total (mice)	%Mortality	AST[a](days)
ArD76986	1	100	0/12	0	-
		1000	6/12	50	16.4
		10000	5/12	42	16.7
B956	2	100	1/12	8	19.7
		1000	7/12	58	14.3
		10000	3/12	25	18.7
ArD96655	7	100	8/8	100	5.5
		1000	12/12	100	10.8
		10000	12/12	100	9.2
ArD94343	8	100	0/12	0	-
		1000	0/12	0	-
		10000	1/12	8	19.7

[a] AST (Average survival time)

https://doi.org/10.1371/journal.pntd.0006078.t002

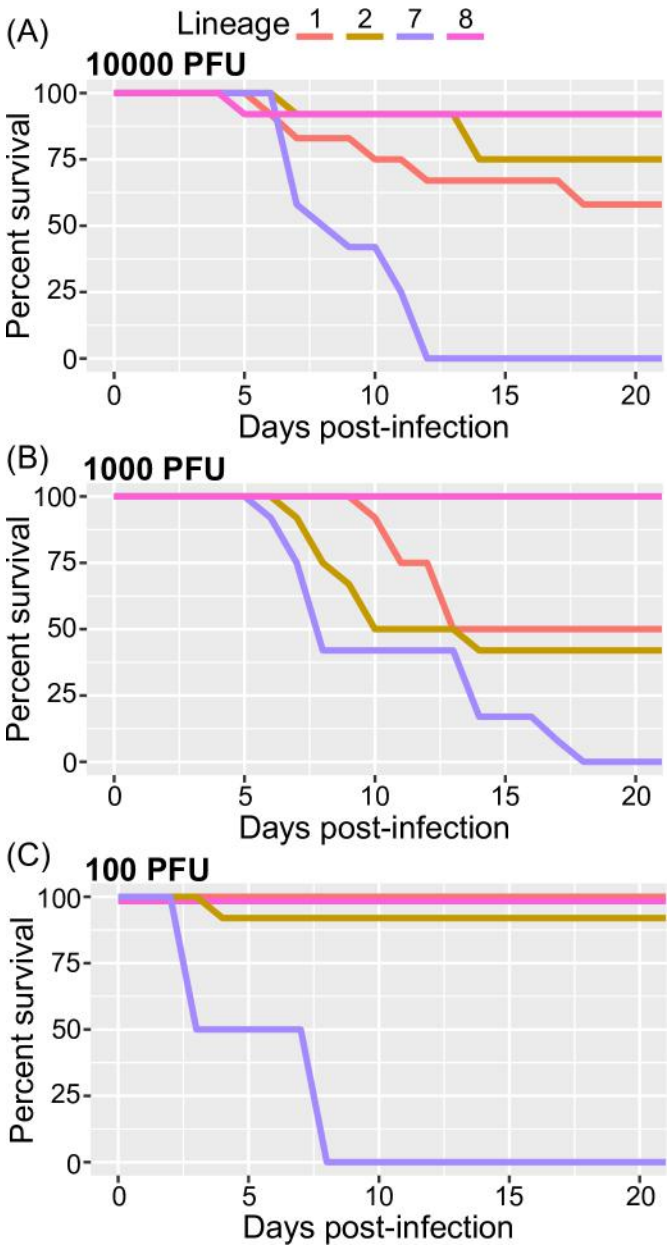


Fig 5. Survival curves of 5- to 6-week-old mice following intraperitoneal infection with (A) 10,000, (B) 1,000 and (C) 100 PFU. Mice were monitored daily for 21 days. Except in (C) when comparing lineage 8 and lineage 1, all survival curves were significantly different (Wilcoxon rank sum test, *p*-values < 0.05).

https://doi.org/10.1371/journal.pntd.0006078.g005

In agreement with other viruses that use alternate hosts (vertebrate-arthropod-vertebrate) and cause acute infections, we found that the majority of WNV codon sites are undergoing purifying selection [76]. Nevertheless, some of the significant episodic diversifying sites that we found are related to virulence, like the 444th and 446th codons in polyprotein gene (154-156th positions of the Env protein) that encode the N-glycosylation motif (NYS). This site is present in many lineage 1 strains, some neuroinvasive lineage 2 strains [34,77], and the lineage 1 and 8 strains from this study (Fig 1B). We found significant diversifying selection acting on these codon sites in 3% and 24% in all of the WNV lineages, respectively (S1 Table). These

episodic non-conservative changes could have resulted in the loss of the N-linked glycosylated site motif, which is related to less efficient replication in *Culex* cells [59] and better replication in *Aedes albopictus*cells [60]. This N-linked site is also associated with neuroinvasiveness in mice [62]. Second, we found that 1754th codon in polyprotein (249th in NS3) was under diversifying selection (ω = 33.1) 13% of the time and under purifying selection (ω = -0.91) 87% of the time (S1 Table) in our WNV dataset. As this site was discovered to increase viremia and virulence in birds [63], further experiments in avian cell lines should be explored to see if our discovered substitutions effect replication in avian hosts and transmission dynamics.

Although no significant diversifying selection was observed on the cleavage site in the NS4A protein (^{120}PEPE^{123} motif), we did discover the P122S substitution in the lineage 8 strain (Fig 1B). Crucially, induced mutations in this motif are related to low rates of replication and protein production in Vero cells [64], which we expected and observed for the lineage 8 strain *in vitro* (Fig 3B, 3D and 3F), and could help explain its low virulence *in vivo* (Fig 5). In general, we observed little change in viral replication and protein production between West African strains in Ap61 cells (Fig 3A, 3C and 3E). This could suggest that the conservation of PEPE motif may have a lesser role in replication in mosquito cells (lineage 8) or that the strains may have been "pre-adapted" prior to our experiment.

The lineage 7 strain has the S653F NS5 mutation that is associated with an increased resistance to interferon [65], which could help explain its phenotypical virulence *in vitro* and *in vivo* (Figs 1B, 3, 4 and 5). However, because Vero cells are known to be interferon-deficient, we could not associate this mutation to our *in vitro* results for lineage 7 in Vero cells (Fig 3, right column). Nevertheless, all three West African strains contained a non-synonymous change in a locus that was previously explored by site-directed mutagenesis experiments (Fig 1B). Interestingly, we also detected synonymous changes in the "wobble" base position of the codons in "sites of interest". However, our knowledge of how synonymous changes impact infectivity, virulence, and replication of WNV is still limited.

As previously described, lineages 1 and 2 originated in Africa and emerged as a *New World* pathogen over the last 60 years [68]. Lineage 7 could be following a similar path; besides Senegal, it has been detected in Somalia, Gabon and possibly in Italy [78–80]. Our study supports that there are other lineages besides 1 and 2—such as lineage 7—that can exhibit high virulence in mice and efficient replication in mammalian cells (Figs 3, 4 and 5). This high virulence of lineage 7/Koutango strains in mice has been explored in two other studies, where the high virulence is suggested to be a result of delayed viral clearance and a weak neutralizing antibody response [43,44]. All three doses tested resulted in 100% mice mortality (Table 2), which agreed with previous results. The differences in average survival time and mortality rates compared to previous studies, could be explained by differences in the passage history of viral strains, and the age of the infected mice [81].

The Senegalese lineage 1 strain exhibited moderate virulence in mice (Fig 5) and caused comparably less mortality than the NY99 strain when compared to similar studies [44]. Differences in neuroinvasive potential and virulence among lineage 1 strains has been reported and could be explained by genetic differences [35]. Alternatively, lineage 8 showed poor growth capabilities in Vero cells (Figs 3 and 4) and almost no virulence in mice (Fig 5 and Table 2), suggesting that it may be restricted to vertical transmission or is species restrictive. Lineage 8 was described to have a similar phenotype to Rabensburg virus (lineage 3, Czech Republic 1997). Moreover, growth kinetics and vector competence studies revealed poor growth of the Rabensburg virus in mammalian cell lines and low virulence in mice [36,82]. This similarity could indicate that both lineage 8 and the Rabensburg strain may be restricted in host range and are also maintained in nature through vertical transmission. Investigating the vector competence of lineage 8 in different arthropod species (*i.e. Culex*, *Aedes*, and tick species) could

lead to a better understand the transmission dynamics and maintenance cycles of WNV in nature. The low virulence phenotype of the lineage 8 strain could also be a factor for its consideration as a potential vaccine candidate for West Nile fever.

Further studies could complement our analysis, particularly, on other factors that could explain differences in WNV host and disease dynamics. Exploring variations in codon usage bias could also help explain biological differences [83], as distinct lineages have shown different degrees of natural selection and mutational bias. Site-directed mutagenesis studies may also help explain how strain-specific mutations, both synonymous and non-synonymous, could explain deviations in replication efficiency and virulence for our *in vitro* and *in vivo* results. For example, future studies in cell lines with interferon may help clarify the impact of the S653F NS5 mutation for the lineage 7 strain. Additionally, the flavivirus 5' and 3' untranslated regions (UTR) can affect replication and translation; certain mutations in these regions can cause complete viral attenuation [84–87]. Unfortunately, we could not investigate their impact in this study, as the majority of WNV UTR's were publically unavailable.

Taking everything into account, especially differences in sequences, growth dynamics and virulence *in vivo*, the West Nile virus is a pathogen with the capability to cause severe epidemics anywhere in the globe. As complete genome sequences including the 5' and 3' UTR regions are currently being generated, this could lead to future studies focused on *in vivo* transmission and growth dynamics. As additional strains of WNV are characterized, monitoring the global diversity and distribution will aid in threat assessment and epidemiological modeling if future outbreaks are to occur.

Materials and methods

Cell lines

Two cell lines have been used for virus cultivation and growth kinetics. Ap61 cells (*Aedes pseudocutellaris*) were grown in L15 (Leibovitz's 15) medium (10% heat-inactivated fetal bovine serum [FBS], 1% penicillin-streptomycin, 0.05% amphotericin B [Fungizone] (GIBCO by life technologies; USA) and 10% tryptose phosphate (Becton, Dickinson and Company Sparks, USA) and incubated at 28˚C without CO_2. Vero cells (African green monkey kidney epithelial cells; *Cercopithecus aethiops*) (obtained from Sigma Aldrich, France) were grown using the same medium without tryptose phosphate and CO_2. Furthermore, PS (Porcine Stable kidney cell line, American type Culture Collection, Manassas, USA) cells were grown in same conditions than Vero cells and have been used for plaque assay.

Virus strains

The virus strains used in this study corresponding to lineages 1, 2, Koutango (lineage 7) and 8 were described in Table 1. The virus stocks were prepared by inoculating *Aedes pseudoscutellaris* (Ap61) continuous cells lines for 4 days. The infection status was tested by immunofluorescence assay (IFA), real-time RT-PCR (Reverse Transcriptase-Polymerase Chain Reaction) and plaque assay. The supernatant of infected cells were aliquoted, frozen at -80˚C, and used as viral stocks for growth kinetics.

Phylogenetic analyses

A total of 862 complete WNV polyprotein gene sequences with country and year of isolation data were available and initially downloaded from Genbank for this study. A large number of sequences were from the Americas and formed a monophyletic group of lineage 1a comprising 770 sequences. To reduce computer-processing requirements while maintaining the authenticity

of our results, we removed all lineage 1a sequences except for a single representative sequence denoted "NY99" (accession number: AF196835). With the addition of 3 new sequences, a total of 95 sequences were aligned using Muscle v3.8.31 [88] and manually curated using Se-Al v2 [89]. For Fig 1, the available complete polyprotein sequences representative of WNV diversity (excluding lineage 6, which there is only a partial sequence available) were included to compare genetic percent identities.

Likelihood mapping analyses for estimation of data quality were performed using Tree-Puzzle (Quartets ranged between 10,000 and 40,000) [90,91]. For each alignment we performed recombination screening (RDP, GeneConv, Chimaera, MaxChi, BootScan and SiScan) in RDP4.61 [92].

The Bayesian phylogenetic analysis was performed using Bayesian Inference (BI) using a general time-reversible with gamma-distributed rate variation and invariant sites model (GTR +Γ+I), as selected by Akaike's information criterion (AICc) in jModelTest 0.1 [93]. The evolutionary analysis was conducted assuming a relaxed Gamma clock and GMRF Bayesian Skyride coalescent tree prior. We then employed a Bayesian MCMC approach using BEAST v1.8.4 and performed five independent MCMC runs with up to 100 million generations to ensure the convergence of estimates. Trees were summarized in a maximum clade-credibility tree after a 10% burn-in [94] and used Tracer (http://beast.bio.ed.ac.uk/Tracer) to ensure convergence during MCMC by reaching effective sample sizes greater than 100.

To reduce the number of sequences from the original 862 downloaded from Genbank, a maximum likelihood tree was estimated using FastTree v2.1.7 [95] after identical alignment and curating methods. FastTree was run using GTR+Γ+I nucleotide model with 2000 Γ-rate categories, exhaustive search settings, with 5000 bootstrap replications using the Shimodaira-Hasegawa (SH) test. The analysis was repeated for the dataset of 95 sequences to compare tree topologies inferred by the Bayesian approach (S1 Fig). All alignments referred to in this manuscript can be found at https://github.com/caiofreire.

Growth kinetics

To perform this study and make it comparable with other studies [60,96], viral stocks were standardized in number of plaque forming units per milliliter (PFU/mL) for cell infections rather than copy numbers of genome. The growth kinetics assays were performed in 12-well plates using one plate per virus strain with one uninfected well as a negative control. Each well was seeded with $2.4x10^5$ Ap61 or Vero cells in a volume of 400 μl of appropriate medium and infected with $2.4x10^3$ PFU (plaque-forming unit) of virus in 400 μl of medium, resulting in a multiplicity of infection (MOI) of 0.01. After an incubation time of 4 hours, the medium was removed and replaced with 2 ml of new medium to set a zero point for the growth curves (T_0). The harvesting of one well occurred at 22, 28, 50, 75, 99, 124, and 146 h post infection. Each harvest was performed as follows. Supernatants were removed and frozen at -80˚C in small aliquots. Cells were washed once with phosphate-buffered saline (PBS) and then removed in 500 μl PBS. A volume of 20 μl of cell suspension was dried on a glass slide for a subsequent immunofluorescence assay as previously described [97] to measure viral proteins production. The remaining cell suspensions were frozen at -80˚C.

RNA was extracted from cell suspensions and supernatants and copy numbers of genome were quantified by real time RT-PCR as previously described [98]. Infectious viral particles were measured in supernatants by plaque assay also as previously described [99]. This study was performed two times on each cell type. The initial titers of lineages 1, 2, 7 and 8 were respectively $3x10^8$, $5x10^4$, $7.5x10^6$ and 10^{10} PFU/ml. For each lineage, $2.4x10^3$ PFU were used for kinetics in mosquito and mammal cells. The ratio of particles per infectious unit in the

initial viral stocks ranged from 8 to 600 [98]. Our viral stocks had a similar ratio of particles per infectious unit as that seen produced by fully infectious extracellular WN virus particles [100] and mosquito-derived replicon WN virus particles [101]. Variances in replication efficiency between studies observed during *in vitro* infection could be explained to differences in the viral strain and to the infection conditions *i.e.* very low MOIs (0.01), and distinct cell lines.

Mouse infection and survival studies

Mice were produced in the Institut Pasteur de Dakar farm, located in Mbao, approximately 15 kilometers from Dakar, Senegal. After one week of acclimatization, five-to-six-week-old Swiss mice were challenged by intraperitoneal (IP) injection with 100, 1000 and 10000 PFU of WNV lineages diluted in phosphate buffer saline + 0.2% endotoxin-free serum albumin (BSA). For each lineage and dose, two independent experiments of infection were made. Each individual experiment had 4 to 8 mice. A group of mice inoculated in parallel with an equivalent volume of phosphate buffer saline + 0.2% endotoxin-free serum albumin (BSA) was maintained as a control. Mice were kept on clean bedding and given food and water *ad libitum*. Infected animals were monitored daily for first signs of encephalitis (hunching, lethargy, eye closure, or hind legs paralysis) and death throughout the 21 days after infection. All statistical inferences were calculated using the Wilcoxon rank sum test.

Selection

To evaluate selection patterns on the complete coding sequences, we estimated the ratio of substitution rates (ω) per non-synonymous site (*d*N) over synonymous substitutions per synonymous site (*d*S) per codon sites. Briefly, sites with $\omega>1$ are assumed to be under positive (diversifying) selection, and sites where $\omega<1$ are undergoing negative (purifying) selection. When $\omega = 0$, the site is undergoing neutral selection. To estimate ω, we applied three maximum likelihood methods: single likelihood ancestor counting (SLAC), fixed-effects likelihood (FEL), and internal fixed-effects likelihood (IFEL). We also investigated the presence of transient (episodic) selective pressures, using the mixed-effects model of evolution (MEME) [102] and fast, unconstrained Bayesian approximation (FUBAR) [103] approaches. For FEL, SLAC, IFEL, and MEME analyses, sites were identified as undergoing significant positive selection when *p*-value $\leq$0.10. For FUBAR, sites were identified as undergoing positive selection when there was a posterior probability $\geq$0.90. All estimations were implemented using HyPhy v2.11 [104].

RNA extraction and quantitative real-time (qRT-PCR)

Extraction of viral RNA from supernatants was performed with the QIAamp viral RNA mini kit (Qiagen, Heiden, Germany) according to manufacturer's instructions. For cell fractions, prior to RNA extraction, cells were lysed by serial cycles of freeze/thaw. For the detection and quantification of viral RNA, a consensus WNV real-time RT-PCR assay and corresponding RNA standard were used as previously described [98]. The real-time PCR assays were performed using the Quantitect Probe RT-PCR Kit (Qiagen, Heiden, Germany) in a 96-well plate under the following conditions: 50˚C for 10 min, 95˚C for 15 min followed by 40 cycles of 95˚C for 15 s and 60˚C for 1 min. Copy numbers of genome were calculated using Ct (Cycle threshold) and corresponding RNA standard.

Complete polyprotein sequencing

Overlapping RT-PCRs were done to recover the complete genome. All primer sequences can be found in the S3 Table. The NS5, envelope and NS5-partial 3'UTR regions were first

amplified using flavivirus consensus or West Nile specific primers [1,105,106], followed by amplification of NS3 region using designed WNV primers. The 5' non-coding region of the genome was obtained using the 5'RACE kit (Invitrogen, Carlsbad, USA) and a designed consensus primer in the capsid protein for reverse primer. Finally, specific primers were designed according to the first sequences obtained and a second step of RT-PCR was done to obtain the complete genome.

The PCR fragments were obtained using AMV reverse transcription kit (Promega, Madison, USA) for reverse transcription and Go-Taq PCR kit (Promega, Madison, USA) for amplification. The RT conditions were set according to the manufacturer's instructions, and the PCR conditions were as follows: 5 minutes at 95˚C, 40 cycles of 1 minute at 95˚C, 1 minute at 53˚C, 1 to 4 minutes (according the size of the PCR product) at 72˚C, and 10 minutes at 72˚C. The PCR products were purified from the agarose gel using the Gel extraction kit (Qiagen) and sequenced by Cogenics (Beckman Coulter Genomics, Essex, UK).

Immunofluorescence assay (IFA)

Infected cells at different time points were dissolved in PBS and dropped on a glass slide. After complete drying, cells were fixed for at least 20 min in cold acetone, dried again, and then stored at -20˚C until staining. Staining was done with a WNV-polyclonal mouse immune ascit diluted in PBS and incubated for 30 minutes at 37˚C. After washing three times with PBS, cells were incubated with the second antibody (goat anti-mouse IgG, fluorescein isothiocyanate [FITC] conjugated Biorad), diluted 1:40 and blue Evans 1/100 in PBS, for 30 minutes at 37˚C in the dark. The cells were washed again three times with PBS, dried, and covered with 50% glycerol in PBS. After dehydration, examination was done by fluorescence microscopy.

Genbank accession numbers

KJ831223, FJ159131, AY277251, FJ159130, FJ159129, AY765264, KY703856, DQ176636, HM147823, FJ425721, KT207791, KJ934710, KP780840, KP780839, KT359349, KC496016, KC407673, KF179639, KJ883346, KC496015, HQ537483, KF647251, KT207792, JN858070, KP109692, KF179640, KM203863, KP780838, KP780837, KM203861, KM203862, JN393308, EF429197, EF429198, HM147824, EF429199, KM052152, EF429200, GQ903680, HM147822, GQ851605, DQ256376, GQ851604, JX041632, GQ851602, KT934796, KT934801, JX123031, JX123030, KT934800, KT934802, KT934803, GQ851603, KT934797, KT934799, KT934798, JX041628, JX041629, JX041630, HM051416, KT163243, EU249803, KC601756, JX442279, JX041634, KU588135, JQ928175, JN858069, KF647253, KC954092, JQ928174, JX556213, JF707789, FJ766331, FJ766332, JF719069, FJ483549, FJ483548, JF719066, JF719067, KF234080, GU011992, JF719068, DQ786573, AY701413, HM152775, AY701412, GQ851606, GQ851607, GQ379161, AF196835, KY703855, EU082200, KY703854, AY532665.

Supporting information

S1 Fig. Phylogenetic inference of West Nile virus using a maximum-likelihood tree. The Shimodaira-Hasegawa values greater than 70% are shown at respective nodes. Tip labels are colored by proposed lineage. Sequences from Table 1 are labeled.
(TIF)

S2 Fig. Selection regimens acting on codons of West Nile Virus polyprotein via FUBAR method. The dashed line marks neutral selection (dN-dS = 0), points above the line (dN>dS) are under diversifying selection and below (dN<dS) are under purifying selection. The intensity of the point color is proportional to the posterior probability to observe that codon under

the selection regimen, calculated with Fubar method.
(TIF)

S3 Fig. Selection regimens acting on codons of West Nile Virus polyprotein via MEME method. Using 95 WNV sequences, A) diversifying selection (dN>dS) and B) purifying selection (dN<dS) were estimated. The intensity of the point color is proportional to the posterior probability to observe that codon under the selection regimen, calculated with MEME method. Significant positively selected sites detected by other methods were also included in A).
(TIF)

S1 Table. Raw data for FUBAR analysis.
(DOCX)

S2 Table. Raw data for MEME analysis.
(CSV)

S3 Table. List of primers used for sequencing. The NS5, envelope and NS5-partial 3'UTR regions were first amplified using flavivirus consensus or West Nile specific primers. This was followed by amplification of NS3 region using designed WNV primers. Finally, specific primers were designed according to the first sequences obtained and a second step of RT-PCR was done to obtain the complete genome.
(DOC)

S1 Dataset. Raw growth kinetics data. Raw data for Fig 3.
(XLSX)

S2 Dataset. Raw mice survival data. Raw data for Fig 5.
(XLSX)

Acknowledgments

We would like to thank Mireille Mondo, Arame Ba, and Khardiata Mbaye for their excellent technical support. We also would like to thank Erica Mendes, John Almberg, and Elissa Rotwein for their helpful comments and critiques in the development and improvement of this work.

Author Contributions

Conceptualization: Gamou Fall, Martin Faye, Ousmane Faye, Amadou Alpha Sall.

Data curation: Gamou Fall, Nicholas Di Paola, Martin Faye, Moussa Dia, Caio César de Melo Freire.

Formal analysis: Gamou Fall, Nicholas Di Paola, Martin Faye, Caio César de Melo Freire, Cheikh Loucoubar.

Funding acquisition: Paolo Marinho de Andrade Zanotto, Ousmane Faye, Amadou Alpha Sall.

Investigation: Gamou Fall, Nicholas Di Paola, Martin Faye, Moussa Dia.

Methodology: Gamou Fall, Nicholas Di Paola, Martin Faye, Moussa Dia, Caio César de Melo Freire.

Project administration: Paolo Marinho de Andrade Zanotto, Ousmane Faye, Amadou Alpha Sall.

Resources: Paolo Marinho de Andrade Zanotto, Ousmane Faye, Amadou Alpha Sall.

Software: Nicholas Di Paola, Caio César de Melo Freire, Cheikh Loucoubar.

Supervision: Gamou Fall, Caio César de Melo Freire, Cheikh Loucoubar, Paolo Marinho de Andrade Zanotto, Ousmane Faye, Amadou Alpha Sall.

Validation: Gamou Fall, Nicholas Di Paola, Martin Faye, Caio César de Melo Freire.

Visualization: Nicholas Di Paola, Caio César de Melo Freire.

Writing – original draft: Gamou Fall, Nicholas Di Paola, Paolo Marinho de Andrade Zanotto.

Writing – review & editing: Gamou Fall, Nicholas Di Paola, Martin Faye, Caio César de Melo Freire, Paolo Marinho de Andrade Zanotto, Amadou Alpha Sall.

References

1. Kuno G, Chang GJ, Tsuchiya KR, Karabatsos N, Cropp CB. Phylogeny of the genus Flavivirus. J Virol. 1998; 72: 73–83. Available: http://jvi.asm.org/content/72/1/73.short PMID: 9420202
2. Chambers TJ, Hahn CS, Galler R, Rice CM. Flavivirus genome organization, expression, and replication. Annu Rev Microbiol. 1990; 44: 649–88. https://doi.org/10.1146/annurev.mi.44.100190.003245 PMID: 2174669
3. Gould EA, Zanotto PM de A, Holmes EC. The genetic evolution of flaviviruses. 1997.
4. Fall G, Diallo M, Loucoubar C, Faye O, Sall AA. Vector competence of Culex neavei and Culex quinquefasciatus (Diptera: Culicidae) from Senegal for lineages 1, 2, Koutango and a putative new lineage of West Nile virus. Am J Trop Med Hyg. 2014; 90: 747–754. https://doi.org/10.4269/ajtmh.13-0405 PMID: 24567319
5. Murgue B, Zeller H, Deubel V. The Ecology and Epidemiology of West Nile Virus in Africa, Europe and Asia. In: Mackenzie JS, Barrett ADT, Deubel V, editors. Berlin, Heidelberg: Springer Berlin Heidelberg; 2002. pp. 195–221. https://doi.org/10.1007/978-3-642-59403-8_10
6. Hayes EB, Sejvar JJ, Zaki SR, Lanciotti RS, Bode A V., Campbell GL. Virology, pathology, and clinical manifestations of West Nile virus disease. Emerg Infect Dis. 2005; 11: 1174–1179. https://doi.org/10.3201/eid1108.050289b PMID: 16102303
7. Kramer LD, Li J, Shi P-Y. West Nile virus. Lancet Neurol. 2007; 6: 171–181. https://doi.org/10.1016/S1474-4422(07)70030-3 PMID: 17239804
8. Marra PP, Griffing SM, McLean RG. West Nile Virus and Wildlife. Emerg Infect Dis. 2003; 9: 7–9.
9. McLean RG. West Nile Virus: emerging threat to public health and animal health. J Vet Med Educ. 2003; 30: 143–4. Available: http://www.ncbi.nlm.nih.gov/pubmed/12970859 PMID: 12970859
10. Smithburn KC, Hughes TP, Burke a W, Paul JH, African A. A neurotropic virus isolated from the blood of a native of uganda. Am J Trop Med Hyg. 1940;s1–20: 471–492.
11. Murgue B, Murri S, Triki H, Deubel V, Zeller HG. West Nile in the Mediterranean basin: 1950–2000. Ann N Y Acad Sci. 2001; 951: 117–26. https://doi.org/10.1111/j.1749-6632.2001.tb02690.x PMID: 11797769
12. Murray KO, Mertens E, Despres P. West Nile virus and its emergence in the United States of America. Vet Res. 2010; 41: 67. https://doi.org/10.1051/vetres/2010039 PMID: 21188801
13. Hubálek Z, Halouzka J. West Nile fever—a reemerging mosquito-borne viral disease in Europe. Emerg Infect Dis. 1999; 5: 643–650. https://doi.org/10.3201/eid0505.990505 PMID: 10511520
14. ArboNET Arboviral Diseases Branch. Final Maps and Data for 1999–2015. In: Centers for Disease Control and Prevention.
15. Asnis DS, Conetta R, Waldman G, Teixera AA. The West Nile Virus Encephalitis Outbreak in the United States (1999–2000). Ann N Y Acad Sci. Blackwell Publishing Ltd; 2006; 951: 161–171. https://doi.org/10.1111/j.1749-6632.2001.tb02694.x
16. Nash D, Mostashari F, Fine A, Miller J, O'Leary D, Murray K, et al. The Outbreak of West Nile Virus Infection in the New York City Area in 1999. N Engl J Med. 2001; 344: 1807–1814. https://doi.org/10.1056/NEJM200106143442401 PMID: 11407341
17. Rossi SL, Ross TM, Evans JD. West Nile Virus. Clin Lab Med. 2010; 30: 47–65. https://doi.org/10.1016/j.cll.2009.10.006 PMID: 20513541

18. Tsai T, Popovici F, Cernescu C, Campbell G, Nedelcu N. West Nile encephalitis epidemic in southeastern Romania. Lancet. 1998; 352: 767–771. https://doi.org/10.1016/S0140-6736(98)03538-7 PMID: 9737281

19. Platonov AE, Shipulin GA, Shipulina OY, Tyutyunnik EN, Frolochkina TI, Lanciotti RS, et al. Outbreak of West Nile virus infection, Volgograd Region, Russia, 1999. Emerg Infect Dis. 2001; 7: 128–32. https://doi.org/10.3201/eid0701.700128 PMID: 11266303

20. Sambri V, Capobianchi M, Charrel R, Fyodorova M, Gaibani P, Gould E, et al. West Nile virus in Europe: emergence, epidemiology, diagnosis, treatment, and prevention. Clin Microbiol Infect. 2013; 19: 699–704. https://doi.org/10.1111/1469-0691.12211 PMID: 23594175

21. Papa A, Bakonyi TT, Xanthopoulou K, Vázquez A, Tenorio A, Nowotny N, et al. Genetic Characterization of West Nile Virus Lineage 2, Greece, 2010. Emerg Infect Dis. 2011; 17: 920–922. https://doi.org/10.3201/eid1705.101759 PMID: 21529413

22. Rossini G, Carletti F, Bordi L, Cavrini F, Gaibani P, Landini MP, et al. Phylogenetic analysis of west nile virus isolates, Italy, 2008–2009. Emerg Infect Dis. 2011; 17: 903–906. https://doi.org/10.3201/eid1705.101569 PMID: 21529408

23. Bagnarelli P, Marinelli K, Trotta D, Monachetti A, Tavio M, Gobbo R Del, et al. Human case of autochthonous West Nile virus lineage 2 infection in Italy, September 2011. Eurosurveillance. 2011; 16: 1–4.

24. Barzon L, Papa A, Lavezzo E, Franchin E, Pacenti M, Sinigaglia A, et al. Phylogenetic characterization of Central/Southern European lineage 2 West Nile virus: analysis of human outbreaks in Italy and Greece, 2013–2014. Clin Microbiol Infect. Elsevier Ltd; 2015; 21: 1122.e1–1122.e10. https://doi.org/10.1016/j.cmi.2015.07.018 PMID: 26235197

25. Bakonyi T, Ferenczi E, Erdélyi K, Kutasi O, Csörgő T, Seidel B, et al. Explosive spread of a neuroinvasive lineage 2 West Nile virus in Central Europe, 2008/2009. Vet Microbiol. 2013; 165: 61–70. https://doi.org/10.1016/j.vetmic.2013.03.005 PMID: 23570864

26. Popović N, Milošević B, Urošević A, Poluga J, Lavadinović L, Nedelijković J, et al. Outbreak of West Nile virus infection among humans in Serbia, August to October 2012. Euro Surveill. 2013; 18: 20613. Available: http://www.eurosurveillance.org/images/dynamic/EE/V18N43/art20613.pdf PMID: 24176618

27. Pachler K, Lebl K, Berer D, Rudolf I, Hubalek Z, Nowotny N, et al. Putative New West Nile Lineage in Uranotaenia unguiculata Mosquitoes, Austria, 2013. Emerg Infect Dis. 2014; 20: 2119–2122. https://doi.org/10.3201/eid2012.140921 PMID: 25418009

28. MacKenzie JS, Williams DT. The zoonotic flaviviruses of southern, south-eastern and eastern Asia, and australasia: The potential for emergent viruses. Zoonoses Public Health. 2009; 56: 338–356. https://doi.org/10.1111/j.1863-2378.2008.01208.x PMID: 19486319

29. David S, Abraham AM. Epidemiological and clinical aspects on West Nile virus, a globally emerging pathogen. Infect Dis (Auckl). 2016; 48: 571–586. https://doi.org/10.3109/23744235.2016.1164890 PMID: 27207312

30. Rizzoli A, Jiménez-Clavero M, Barzon L, Cordioli P, Figuerola J, Koraka P, et al. The challenge of West Nile virus in Europe: knowledge gaps and research priorities. Eurosurveillance. 2015; 20: 21135. https://doi.org/10.2807/1560-7917.ES2015.20.20.21135 PMID: 26027485

31. Hall R a, Scherret JH, Mackenzie JS. Kunjin virus: an Australian variant of West Nile? Ann N Y Acad Sci. 2001; 951: 153–60. https://doi.org/10.1111/j.1749-6632.2001.tb02693.x PMID: 11797773

32. Lanciotti RS, Roehrig JT, Deubel V, Smith RJ, Parkes M, Steele K, et al. Origin of the West Nile Virus Responsible for an Outbreak of Encephalitis in the Northeastern United States. Science (80-). 1999; 286: 2333–2337. https://doi.org/10.1126/science.286.5448.2333

33. Bakonyi TT, Ivanics E, Erdleyi K, Ursu K, Ferenczi EE, Weissenbock H, et al. Lineage 1 and 2 strains of encephalitic West Nile virus, Central Europe. Emerg Infect Dis. 2006; 12: 618–623. https://doi.org/10.3201/eid1204.051379 PMID: 16704810

34. Botha EM, Markotter W, Wolfaardt M, Paweska JT, Swanepoel R, Palacios G, et al. Genetic determinants of virulence in pathogenic lineage 2 West Nile virus strains. Emerg Infect Dis. 2008; 14: 222–230. https://doi.org/10.3201/eid1402.070457 PMID: 18258114

35. Beasley DWC, Li L, Suderman MT, Barrett ADT. Mouse Neuroinvasive Phenotype of West Nile Virus Strains Varies Depending upon Virus Genotype. Virology. Elsevier; 2002; 296: 17–23. https://doi.org/10.1006/viro.2002.1372 PMID: 12036314

36. Hubalek Z, Halouzka J, Juricova Z, Sebesta O. First isolation of mosquito-borne west nile virus in the Czech Republic [1]. Acta Virol. 1998; 42: 119–120. PMID: 9770080

37. Bakonyi T, Hubálek Z, Rudolf I, Nowotny N. Novel Flavivirus or New Lineage of West Nile Virus, Central Europe. Emerg Infect Dis. 2005; 11: 225–231. https://doi.org/10.3201/eid1102.041028 PMID: 15752439

38. Hubálek Z, Rudolf I, Bakonyi T, Kazdová K, Halouzka J, Šebesta O, et al. Mosquito (Diptera: Culicidae) Surveillance for Arboviruses in an Area Endemic for West Nile (Lineage Rabensburg) and Ťahyňa Viruses in Central Europe. J Med Entomol. 2010; 47: 466–472. https://doi.org/10.1093/jmedent/47.3.466 PMID: 20496595

39. Lvov DK, Butenko AM, Gromashevsky VL, Kovtunov AI, Prilipov AG, Kinney R, et al. West Nile virus and other zoonotic viruses in Russia: examples of emerging-reemerging situations. In: Calisher CH, Griffin DE, editors. Emergence and Control of Zoonotic Viral Encephalitides. Vienna: Springer Vienna; 2004. pp. 85–96. https://doi.org/10.1007/978-3-7091-0572-6_7

40. Lanciotti RS, Ebel GD, Deubel V, Kerst AJ, Murri S, Meyer R, et al. Complete Genome Sequences and Phylogenetic Analysis of West Nile Virus Strains Isolated from the United States, Europe, and the Middle East. Virology. 2002; 298: 96–105. https://doi.org/10.1006/viro.2002.1449 PMID: 12093177

41. Vázquez A, Sánchez-Seco MP, Ruiz S, Molero F, Hernández L, Moreno J, et al. Putative new lineage of West Nile virus, Spain. Emerg Infect Dis. 2010; 16: 549–552. https://doi.org/10.3201/eid1603.091033 PMID: 20202444

42. Charrel RN, Brault AC, Gallian P, Lemasson JJ, Murgue B, Murri S, et al. Evolutionary relationship between Old World West Nile virus strains: Evidence for viral gene flow between Africa, the Middle East, and Europe. Virology. 2003; 315: 381–388. https://doi.org/10.1016/S0042-6822(03)00536-1 PMID: 14585341

43. Prow N a, Setoh YX, Biron RM, Sester DP, Kim KS, Hobson-Peters J, et al. The West Nile-like flavivirus Koutango is highly virulent in mice due to delayed viral clearance and the induction of a poor neutralizing antibody response. J Virol. 2014; 88: 9947–9962. https://doi.org/10.1128/JVI.01304-14 PMID: 24942584

44. Pérez-Ramírez E, Llorente F, del Amo J, Fall G, Lubisi A, Lecollinet S, et al. Pathogenicity evaluation of twelve West Nile virus strains belonging to four lineages from five continents in a mouse model: discrimination between three pathogenicity categories. J Gen Virol. Microbiology Society; 2017; 98: 662–670. https://doi.org/10.1099/jgv.0.000743 PMID: 28475031

45. Shope RE. Epidemiology of Other Arthropod-Borne Flaviviruses Infecting Humans. Adv Virus Res. 2003; 61: 373–391. https://doi.org/10.1016/S0065-3527(03)61009-2 PMID: 14714437

46. Venter M, Human S, Van Niekerk S, Williams J, van Eeden C, Freeman F. Fatal neurologic disease and abortion in mare infected with lineage 1 West Nile virus, South Africa. Emerg Infect Dis. 2011; 17: 1534–1536. https://doi.org/10.3201/eid1708.101794 PMID: 21801644

47. Virus d'Afrique (base de donnees) [Internet]. [cited 5 Oct 2016]. Available: http://www.pasteur.fr/recherche/banques/CRORA/

48. Solomon T. West Nile encephalitis. BMJ. 2003; 326: 865–869. https://doi.org/10.1136/bmj.326.7394.865 PMID: 12702624

49. Depoortere E, Kavle J, Keus K, Zeller H, Murri S, Legros D. Outbreak of West Nile virus causing severe neurological involvement in children, Nuba Mountains, Sudan, 2002. Trop Med Int Heal. 2004; 9: 730–736. https://doi.org/10.1111/j.1365-3156.2004.01253.x PMID: 15189465

50. Hachfi W, Bougmiza I, Bellazreg F, Bahri O, Kaabia N, Bahri F, et al. Une deuxième épidémie de méningo-encéphalite à virus West Nile en Tunisie. Médecine Mal Infect. 2010; 40: 456–461. http://dx.doi.org/10.1016/j.medmal.2009.12.005

51. Mostashari F, Bunning ML, Kitsutani PT, Singer DA, Nash D, Cooper MJ, et al. Epidemic West Nile encephalitis, New York, 1999: results of a household-based seroepidemiological survey. Lancet. Elsevier; 2016; 358: 261–264. https://doi.org/10.1016/S0140-6736(01)05480-0

52. Wang W, Sarkodie F, Danso K, Addo-Yobo E, Owusu-Ofori S, Allain J- P, et al. Seroprevalence of west nile virus in ghana. Viral Immunol. 2009; 22: 17–22. https://doi.org/10.1089/vim.2008.0066 PMID: 19210224

53. Rappole JH, Derrickson SR, Hubálek Z, Hubálek Z. Migratory birds and spread of West Nile virus in the Western Hemisphere. Emerg Infect Dis. 2000; 6: 319–328. https://doi.org/10.3201/eid0604.000401 PMID: 10905964

54. Renaudet J, Jan C, Ridet J, Adam C, Robin Y. [A serological survey of arboviruses in the human population of Senegal]. Bull Soc Pathol Exot Filiales. 1977; 71: 131–140.

55. Traore-Lamizana M, Zeller HG, Mondo M, Hervy J- P, Adam F, Digoutte J- P. Isolations of West Nile and Bagaza Viruses from Mosquitoes (Diptera: Culicidae) in Central Senegal (Ferlo). J Med Entomol. 1994; 31: 934 LP–938. Available: http://jme.oxfordjournals.org/content/31/6/934.abstract

56. Traoré-Lamizana M, Fontenille D, Diallo M, Bâ Y, Zeller HG, Mondo 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. J Med Entomol. 2001; 38: 480 LP–492. Available: http://jme.oxfordjournals.org/content/38/4/480.abstract

57. Chevalier V, Dupressoir A, Tran A, Diop OM, Gottland C, Diallo M, et al. Environmental risk factors of West Nile virus infection of horses in the Senegal River basin. Epidemiol Infect. Cambridge, UK: Cambridge University Press; 2010; 138: 1601–1609. https://doi.org/10.1017/S095026881000035X PMID: 20175940

58. Shirato K, Miyoshi H, Goto A, Ako Y, Ueki T, Kariwa H, et al. Viral envelope protein glycosylation is a molecular determinant of the neuroinvasiveness of the New York strain of West Nile virus. J Gen Virol. 2004; 85: 3637–45. https://doi.org/10.1099/vir.0.80247-0 PMID: 15557236

59. Moudy RM, Zhang B, Shi P- Y, Kramer LD. West Nile virus envelope protein glycosylation is required for efficient viral transmission by Culex vectors. Virology. 2009; 387: 222–8. https://doi.org/10.1016/j.virol.2009.01.038 PMID: 19249803

60. Hanna SL, Pierson TC, Sanchez MD, Ahmed AA, Murtadha MM, Doms RW. N-linked glycosylation of west nile virus envelope proteins influences particle assembly and infectivity. J Virol. 2005; 79: 13262–74. https://doi.org/10.1128/JVI.79.21.13262-13274.2005 PMID: 16227249

61. Setoh YX, Prow NA, Hobson-Peters J, Lobigs M, Young PR, Khromykh AA, et al. Identification of residues in West Nile virus pre-membrane protein that influence viral particle secretion and virulence. J Gen Virol. 2012; 93: 1965–75. https://doi.org/10.1099/vir.0.044453-0 PMID: 22764317

62. Beasley DWC, Whiteman MC, Zhang S, Huang CY-H, Schneider BS, Smith DR, et al. Envelope protein glycosylation status influences mouse neuroinvasion phenotype of genetic lineage 1 West Nile virus strains. J Virol. 2005; 79: 8339–47. https://doi.org/10.1128/JVI.79.13.8339-8347.2005 PMID: 15956579

63. Brault AC, Huang CY, Langevin SA, Kinney RM, Bowen RA, Ramey WN, et al. A single positively selected West Nile viral mutation confers increased virogenesis in American crows. Nat Genet. 2007; 39: 1162–1166. https://doi.org/10.1038/ng2097 PMID: 17694056

64. Ambrose RL, Mackenzie JM. A Conserved Peptide in West Nile Virus NS4A Protein Contributes to Proteolytic Processing and Is Essential for Replication. J Virol. 2011; 85: 11274–11282. https://doi.org/10.1128/JVI.05864-11 PMID: 21880777

65. Van Slyke G a., Ciota AT, Willsey GG, Jaeger J, Shi PY, Kramer LD. Point mutations in the West Nile virus (Flaviviridae; Flavivirus) RNA-dependent RNA polymerase alter viral fitness in a host-dependent manner in vitro and in vivo. Virology. Elsevier Inc.; 2012; 427: 18–24. https://doi.org/10.1016/j.virol.2012.01.036 PMID: 22365326

66. Liu WJ, Wang XJ, Clark DC, Lobigs M, Hall R a, Khromykh A a. A Single Amino Acid Substitution in the West Nile Virus Nonstructural Protein NS2A Disables Its Ability To Inhibit Alpha / Beta Interferon Induction and Attenuates Virus Virulence in Mice A Single Amino Acid Substitution in the West Nile Virus Nonstructur. J Virol. 2006; 80: 2396–2404. https://doi.org/10.1128/JVI.80.5.2396-2404.2006 PMID: 16474146

67. Wicker JA, Whiteman MC, Beasley DWC, Davis CT, Zhang S, Schneider BS, et al. A single amino acid substitution in the central portion of the West Nile virus NS4B protein confers a highly attenuated phenotype in mice. Virology. 2006; 349: 245–253. https://doi.org/10.1016/j.virol.2006.03.007 PMID: 16624366

68. May FJ, Davis CT, Tesh RB, Barrett ADT. Phylogeography of West Nile virus: from the cradle of evolution in Africa to Eurasia, Australia, and the Americas. J Virol. 2011; 85: 2964–74. https://doi.org/10.1128/JVI.01963-10 PMID: 21159871

69. Ciccozzi M, Peletto S, Cella E, Giovanetti M, Lai A, Gabanelli E, et al. Epidemiological history and phylogeography of West Nile virus lineage 2. Infect Genet Evol. 2013; 17: 46–50. https://doi.org/10.1016/j.meegid.2013.03.034 PMID: 23542457

70. Brinton M a. Replication cycle and molecular biology of the west nile virus. Viruses. 2013; 6: 13–53. https://doi.org/10.3390/v6010013 PMID: 24378320

71. Carpenter JE, Henderson EP, Grose C. Enumeration of an extremely high particle-to-PFU ratio for Varicella-zoster virus. J Virol. 2009; 83: 6917–21. https://doi.org/10.1128/JVI.00081-09 PMID: 19369328

72. Alfson KJ, Avena LE, Beadles MW, Staples H, Nunneley JW, Ticer A, et al. Particle-to-PFU Ratio of Ebola Virus Influences Disease Course and Survival in Cynomolgus Macaques. Dermody TS, editor. J Virol. 2015; 89: 6773–6781. https://doi.org/10.1128/JVI.00649-15 PMID: 25903348

73. Hussmann KL, Vandergaast R, Zheng K, Hoover LI, Fredericksen BL. Structural proteins of West Nile virus are a major determinant of infectious particle production and fitness in astrocytes. J Gen Virol. 2014; 95: 1991–2003. https://doi.org/10.1099/vir.0.065474-0 PMID: 24920724

74. Engel D, Jöst H, Wink M, Börstler J, Bosch S, Garigliany M, et al. Reconstruction of the Evolutionary History and Dispersal of Usutu Virus, a Neglected Emerging Arbovirus in Europe and Africa. MBio. 2016; 7: 1–12. https://doi.org/10.1128/mBio.01938-15.Editor

75. McMullen AR, Albayrak H, May FJ, Davis CT, Beasley DWC, Barrett ADT. Molecular evolution of lineage 2 West Nile virus. J Gen Virol. 2013; 94: 318–25. https://doi.org/10.1099/vir.0.046888-0 PMID: 23136360

76. 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. Mol Biol Evol. SMBE; 2004; 21: 1074–1080. https://doi.org/10.1093/molbev/msh109 PMID: 15014142

77. Berthet FX, Zeller HG, Drouet MT, Rauzier J, Digoutte JP, Deubel V. Extensive nucleotide changes and delections within the envelope glycoprotein gene of Euro-African West Nile viruses. JGenVirol. 1997; 78: 2293–2297. Available: http://www.ncbi.nlm.nih.gov/pubmed/9292017

78. Butenko AM, Semashko I V, Skvortsova TM, Gromashevskiĭ VL, Kondrashina NG. Detection of the Koutango virus (Flavivirus, Togaviridae) in Somalia. Med Parazitol (Mosk). 1986; 65.

79. Jan C, Languillat G, Renaudet J, Robin Y. A serological survey of arboviruses in Gabon. Bull Soc Pathol Exot Filiales. 1978; 71: 140. PMID: 743766

80. Rizzoli A, Rosà R, Rosso F, Buckley A, Gould E. West Nile Virus Circulation Detected in Northern Italy in Sentinel Chickens. Vector-Borne Zoonotic Dis. 2007; 7: 411–417. https://doi.org/10.1089/vbz.2006.0626 PMID: 17767411

81. Dow C, Jarrett WFH. Age, strain and sex differences in susceptibility to Cysticercus fasciolaris in the mouse. Exp Parasitol. Elsevier; 1960; 10: 72–74. PMID: 13724123

82. Aliota MT, Kramer LD. Replication of West Nile virus, Rabensburg lineage in mammalian cells is restricted by temperature. Parasit Vectors. 2012; 5: 293. https://doi.org/10.1186/1756-3305-5-293 PMID: 23241081

83. Moratorio G, Iriarte A, Moreno P, Musto H, Cristina J. A detailed comparative analysis on the overall codon usage patterns in West Nile virus. Infect Genet Evol. 2013; 14: 396–400. https://doi.org/10.1016/j.meegid.2013.01.001 PMID: 23333335

84. Lo MK, Tilgner M, Bernard K a, Shi P-Y. Functional analysis of mosquito-borne flavivirus conserved sequence elements within 3' untranslated region of West Nile virus by use of a reporting replicon that differentiates between viral translation and RNA replication. J Virol. 2003; 77: 10004–10014. https://doi.org/10.1128/JVI.77.18.10004-10014.2003 PMID: 12941911

85. Brault AC, Kinney RM, Maharaj PD, Green EN, Reisen WK, Huang CY. Replication of the primary dog kidney-53 dengue 2 virus vaccine candidate in Aedes aegypti is modulated by a mutation in the 5' untranslated region and amino acid substitutions in nonstructural proteins 1 and 3. Vector Borne Zoonotic Dis. 2011; 11: 683–689. https://doi.org/10.1089/vbz.2010.0150 PMID: 21284523

86. Mandl CW, Holzmann H, Meixner T, Rauscher S, Stadler PF, Allison SL, et al. Spontaneous and engineered deletions in the 3' noncoding region of tick-borne encephalitis virus: construction of highly attenuated mutants of a flavivirus. J Virol. 1998; 72: 2132–40. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=109508&tool=pmcentrez&rendertype=abstract PMID: 9499069

87. Brinton MA, Dispoto JH. Sequence and secondary structure analysis of the 5′-terminal region of flavivirus genome RNA. Virology. 1988; 162: 290–299. https://doi.org/10.1016/0042-6822(88)90468-0 PMID: 2829420

88. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32: 1792–7. https://doi.org/10.1093/nar/gkh340 PMID: 15034147

89. Rambaut A. Se-Al: sequence alignment editor. 1996;

90. Strimmer K, von Haeseler a. Quartet puzzling—a quartet maximum-likelihood method for reconstructing tree topologies. Mol Biol Evol. 1996; 13: 964–969. https://doi.org/10.1093/oxfordjournals.molbev.a025664

91. Schmidt HA, Strimmer K, Vingron M, von Haeseler A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics. Oxford Univ Press; 2002; 18: 502–504. PMID: 11934758

92. Martin D, Rybicki E. RDP: detection of recombination amongst aligned sequences. Bioinformatics. Oxford Univ Press; 2000; 16: 562–563. PMID: 10980155

93. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2 : more models, new heuristics and parallel computing CircadiOmics : integrating circadian genomics, transcriptomics, proteomics. Nat Methods. Nature Publishing Group; 2012; 9: 772. https://doi.org/10.1038/nmeth.2109 PMID: 22847109

94. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012; 29: 1969–73. https://doi.org/10.1093/molbev/mss075 PMID: 22367748

95. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One. 2010; 5: e9490. https://doi.org/10.1371/journal.pone.0009490 PMID: 20224823

96. Stock NK, Laraway H, Faye O, Diallo M, Niedrig M, Sall A a. Biological and phylogenetic characteristics of yellow fever virus lineages from West Africa. J Virol. 2013; 87: 2895–907. https://doi.org/10.1128/JVI.01116-12 PMID: 23269797

97. 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: 417–422. Available: http://www.ncbi.nlm.nih.gov/pubmed/1297177 PMID: 1297177

98. Fall G, Faye M, Weidmann M, Kaiser M, Dupressoir A, Ndiaye EH, et al. Real-time RT-PCR assays for detection and genotyping of West Nile virus lineages circulating in Africa. Vector-Borne Zoonotic Dis. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA; 2016; 16: 781–789. https://doi.org/10.1089/vbz.2016.1967 PMID: 27710313

99. De Madrid a T, Porterfield JS. A simple micro-culture method for the study of group B arboviruses. Bull World Health Organ. 1969; 40: 113–21. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2554446&tool=pmcentrez&rendertype=abstract PMID: 4183812

100. Wengler G, Wengler G. Cell-associated West Nile flavivirus is covered with E+pre-M protein heterodimers which are destroyed and reorganized by proteolytic cleavage during virus release. J Virol. 1989; 63: 2521–2526. PMID: 2724410

101. Boylan BT, Moreira FR, Carlson TW, Bernard KA. Mosquito cell-derived West Nile virus replicon particles mimic arbovirus inoculum and have reduced spread in mice. PLoS Negl Trop Dis. 2017; 11: 1–23. https://doi.org/10.1371/journal.pntd.0005394 PMID: 28187142

102. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 2012; 8: e1002764. https://doi.org/10.1371/journal.pgen.1002764 PMID: 22807683

103. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, et al. FUBAR: A Fast, Unconstrained Bayesian Approximation for Inferring Selection. Mol Biol Evol. 2013; 30: 1196–1205. https://doi.org/10.1093/molbev/mst030 PMID: 23420840

104. Pond SLK, Frost SDW, Muse S V. HyPhy: hypothesis testing using phylogenies. Bioinformatics. Springer; 2005; 21: 676–679. https://doi.org/10.1093/bioinformatics/bti079 PMID: 15509596

105. Gaunt MW, Sall AA, Lamballerie X De, Gould E, Falconar AKI, Dzhivanian TI, et al. Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography. J Gen Virol. 2001; 82: 1867–1876. https://doi.org/10.1099/0022-1317-82-8-1867 PMID: 11457992

106. Pierre V, Drouet MT, Deubel V. Identification of mosquito-borne flavivirus sequences using universal primers and reverse transcription/polymerase chain reaction. Res Virol. Elsevier; 1994; 145: 93–104. PMID: 7520190

jga-genome/jga99917/jga6097d17z | xppws | S=3 | 11/9/17 | 11:19 | ArtID: | DOI:10.1128/genomeA.01192-17 | CE: RL

Editor: | Section: **Viruses** | Designation:

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

VIRUSES

Complete Genome Sequences of Two Human Parainfluenza Virus Type 3 Isolates Collected in Brazil

AQ: au Nicholas Di Paola,[a] Marielton dos Passos Cunha,[a] Giuliana Stravinskas Durigon,[b,c] Eitan Naaman Berezin,[b] Edison Durigon,[a] Danielle Bruna Leal de Oliveira,[a] Paolo Marinho de Andrade Zanotto[a]

AQ: aff Department of Microbiology, Institute of Biomedical Sciences, University of Sao Paulo, Sao Paulo, Brazil[a]; Department of Pediatrics, Santa Casa de Misericórdia Hospital, Sao Paulo, Brazil[b]; Infectious Disease Division, Children's Institute University of Sao Paulo, HC-FMUSP, Brazil[c]

AQ:A **ABSTRACT** Here, we present the complete genome sequences of two human parainfluenza virus type 3 (HPIV-3) isolates collected from hospitalized infants suffering from acute respiratory disease. These are the first complete genome sequences of HPIV-3 originating from Brazil.

Received 21 September 2017 **Accepted** 6 November 2017 **Published** XXX

Citation Di Paola N, Cunha MDP, Durigon GS, Berezin EN, Durigon E, de Oliveira DBL, de Andrade Zanotto PM. 2017. Complete genome sequences of two human parainfluenza virus type 3 isolates collected in Brazil. Genome Announc 5:e01192-17. https://doi.org/10.1128/genomeA.01192-17.

Address correspondence to Paolo Marinho de Andrade Zanotto, pzanotto@usp.br.

AQ: B Human parainfluenza viruses (HPIVs) are single-stranded RNA viruses and part of the *Paramyxoviridae* family. The four types of HPIV exist in two distinct genera: *Respirovirus* (HPIV-1 and HPIV-3) and *Rubulavirus* (HPIV-2 and HPIV-4). HPIV-3 has a higher prevalence than other HPIVs and is responsible for nearly 11% of all pediatric hospitalizations for acute respiratory infections (1, 2). In Brazil, HPIV-3 infections leading to pediatric hospitalization have been reported as high as 8.3% (3).

AQ: C In 2010, a prospective study of acute respiratory infection surveillance was conducted in children under 24 months old. Patients who presented at the time of their admission signs and symptoms of lower respiratory tract infection, including history of coughing and/or respiratory distress and/or those with one or more of the following clinical diagnoses, bronchiolitis, alveolar pneumonia, wheezing, bronchospasm, croup, coqueluchoide syndrome, whooping cough, cyanosis, and apnea, were included in this study. Sixteen patients tested negative for all PCR diagnostic assays for common respiratory viruses. To discover the etiological pathogen and improve surveillance, patient samples were processed for next-generation sequencing.

Viral RNA was extracted from the nasopharyngeal aspirates using the QIAamp Viral RNA minikit (Qiagen, Valencia, CA, USA) and purified with DNase I and concentrated using the RNA Clean and Concentrator TM-5 kit (Zymo Research, Irvine, CA, USA). The AQ: D paired-end RNA libraries were constructed and validated using the TruSeq Stranded Total RNA HT sample prep kit (Illumina, San Diego, CA, USA). Sequencing was done at the Core Facility for Scientific Research–University of São Paulo (CEFAP-USP/GENIAL) using the Illumina NextSeq platform. Each sample was barcoded individually, which allowed separation of reads for each patient. Short unpaired reads and bases and low-quality reads were removed using Trimmomatic version 0.36 (4). Paired-end reads AQ: E (Phred quality score >33) were assembled *de novo* with SPAdes version 3.10 using default parameters (5).

For two patients, STA762 and STA829, the largest assembled contig was identified as HPIV-3 using BLAST searches. We extracted the consensus sequences using Geneious AQ: F version 9.1.2 (6). For STA762, the sequence had a length of 15,450 with 1,415.6× average depth. Likewise, the sequence extracted from STA829 had a length of 15,422 with 20.2× average depth. When aligned with the HPIV-3 reference genome sequence

jga-genome/jga99918/jga6280d18z | xppws | S=3 | 1/12/18 | 12:21 | ArtID: | DOI:10.1128/genomeA.01609-17 | CE: AS

Editor: | Section: **Viruses** | Designation:

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

VIRUSES

Complete Genome Sequences of Five Human Respiratory Syncytial Virus Isolates Collected in Brazil

AQ: au AQ:A Nicholas Di Paola,[a] Marielton dos Passos Cunha,[a] Giuliana Stravinskas Durigon,[b,c] Eitan Naaman Berezin,[b] Edison Durigon,[a] Danielle Bruna Leal de Oliveira,[a] Paolo Marinho de Andrade Zanotto[a]

[a]Department of Microbiology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil
[b]Department of Pediatrics, Santa Casa de Misericórdia Hospital, São Paulo, Brazil
[c]Infectious Disease Division, Children's Institute University of São Paulo, HC-FMUSP, São Paulo, Brazil

ABSTRACT Here, we present the complete genome sequences of five human respiratory syncytial virus isolates collected from hospitalized infants suffering from acute respiratory disease. These are the first five complete genome sequences of human respiratory syncytial virus to originate from Brazil.

Human respiratory syncytial virus (HRSV) is a negative-sense single-stranded RNA virus and is part of the *Orthopneumovirus* genus within the *Pneumoviridae* family. HRSV is a causative agent of severe lower respiratory tract infection in infants and children (1). Nearly all children over the age of two have been infected by HRSV (2). In Brazil, HRSV was detected in 41.8% of patients under 2 years of age with lower respiratory tract infections (3).

Patients under 2 years of age were included in prospective study of acute respiratory infection surveillance. Patients with signs, symptoms, and/or a history of lower respiratory tract infections at the time of their admission were included in this study. Specifically, patients with signs or symptoms of wheezing, whooping cough, coqueluchoide (pertussis-like) syndrome, croup, cyanosis, alveolar pneumonia, apnea, bronchiolitis, and bronchospasms were considered. Nasopharyngeal aspirates from patients were first subjected to all PCR diagnostic assays for common respiratory viruses; all were found to be negative. To discover the etiological pathogen and improve surveillance, patient samples were processed for next-generation sequencing.

AQ: B

Viral RNA was extracted from the nasopharyngeal aspirates using the QIAamp viral RNA minikit (Qiagen, Valencia, CA, USA), purified with DNase I, and concentrated using the RNA Clean & Concentrator TM-5 kit (Zymo Research, Irvine, CA, USA). The paired-end RNA libraries were constructed and validated using the TruSeq stranded total RNA high-throughput (HT) sample prep kit (Illumina, San Diego, CA, USA). Sequencing was done at the Core Facility for Scientific Research–University of São Paulo (CEFAP-USP/GENIAL) using the Illumina NextSeq platform. Each sample was barcoded individually, which allowed separation of reads for each patient. Short unpaired reads and bases and low-quality reads were removed using Trimmomatic version 0.36 (4). Paired-end reads (Phred quality score >33) were assembled *de novo* with SPAdes version 3.10 using default parameters (5).

The *de novo* assemblies of five isolates constructed contigs ranging from 15,181 to 15,268 nucleotides (nt) in length. The average depths of the assemblies ranged from 12× to 475×. Using Geneious 9.1.2, we extracted the consensus sequences (6) and used BLASTn to identify HRSV and class the subtype of each HRSV isolate. STA754, STA786, and STA826 were most closely related to HRSV subtype B (HRSV-B) isolates, while STA836 and STA839 were to HRSV subtype A (HRSV-A). When our HRSV-A isolates were compared to the HRSV-A reference genome sequence (GenBank accession number NC_001803), pairwise identities were 97% (STA836) and 96% (STA839). Similarly, for

Received 29 December 2017 **Accepted** 5 January 2018 **Published** XXX

Citation Di Paola N, Cunha MdP, Durigon GS, Berezin EN, Durigon E, de Oliveira DBL, Zanotto PMdA. 2018. Complete genome sequences of five human respiratory syncytial virus isolates collected in Brazil. Genome Announc 6:e01609-17. https://doi.org/10.1128/genomeA.01609-17.

Address correspondence to Paolo Marinho de Andrade Zanotto, pzanotto@usp.br.

jga-genome/jga99918/jga6285d18z | xppws | S=3 | 1/10/18 | 7:47 | ArtID: | DOI:10.1128/genomeA.01597-17 | CE: RL

Editor: | Section: **Viruses** | Designation:

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

VIRUSES

Complete Genome Sequence of a Human Metapneumovirus Isolate Collected in Brazil

AQ: au AQ:A

Nicholas Di Paola,[a] Marielton dos Passos Cunha,[a] Giuliana Stravinskas Durigon,[b,c] Eitan Naaman Berezin,[b] Edison Durigon,[a] Danielle Bruna Leal de Oliveira,[a] Paolo Marinho de Andrade Zanotto[a]

[a]Department of Microbiology, Institute of Biomedical Sciences, University of Sao Paulo, Sao Paulo, Brazil
[b]Department of Pediatrics, Santa Casa de Misericórdia Hospital, Sao Paulo, Brazil
[c]Infectious Disease Division, Children's Institute University of Sao Paulo, HC-FMUSP, Sao Paulo, Brazil

ABSTRACT Here, we present the complete genome sequence of a human metapneumovirus isolate collected from a hospitalized infant suffering from acute respiratory disease. This is the first complete genome sequence of human metapneumovirus originating from Brazil.

Human metapneumovirus (HMPV) is a negative-sense single-stranded RNA virus and is part of the *Metapneumovirus* genus of the family *Pneumoviridae*. Originally discovered in 2001 in the Netherlands, it is a causal agent of acute respiratory infection (ARI) in primarily children (1, 2). The incidence of HMPV in ARI cases has varied by continent; reports have generally ranged from 4% to 16% (3–5). In southeastern Brazil, HMPV has been detected in 11.4% of children suffering from ARI (6).

Patients under 2 years of age were included in a prospective study of ARI surveillance. Patients with signs, symptoms, and/or a history of lower respiratory tract infections at the time of their admission were included in this study. This included any of the following clinical diagnoses: alveolar pneumonia, apnea, bronchiolitis, bronchospasm, coqueluchoide syndrome, croup, cyanosis, wheezing, and whooping cough. Several patients tested negative for all PCR diagnostic assays for common respiratory viruses. To discover the etiological pathogen and improve surveillance, patient samples were processed for next-generation sequencing.

AQ: B

Viral RNA was extracted from the nasopharyngeal aspirates using the QIAamp Viral RNA minikit (Qiagen, Valencia, CA, USA) and purified with DNase I and concentrated using the RNA Clean and Concentrator TM-5 kit (Zymo Research, Irvine, CA, USA). The paired-end RNA libraries were constructed and validated using the TruSeq Stranded Total RNA HT sample prep kit (Illumina, San Diego, CA, USA). Sequencing was done at the Core Facility for Scientific Research–University of São Paulo (CEFAP-USP/GENIAL) using the Illumina NextSeq platform. Each sample was barcoded individually, which allowed separation of reads for each patient. Short unpaired reads and bases and low-quality reads were removed using Trimmomatic version 0.36 (7). Paired-end reads (Phred quality score >33) were assembled *de novo* with SPAdes version 3.10 using default parameters (8).

The *de novo* assembly of isolate STA755 constructed a single contig of 13,243 nucleotides (nt) that was identified as an HMPV using BLASTn analysis. The HMPV-assembled consensus sequence had a 1,069× average depth. Using Geneious version 9.1.2, we extracted the consensus sequence of STA755 (9). When aligned with the HMPV reference genome sequence (GenBank accession number NC_004148.2), the sequence exceeded a 99.3% breadth of coverage with missing nucleotides at the extremes of the 5′ and 3′ untranslated regions.

A maximum-likelihood tree was estimated using FastTree version 2.1 (10) and included publicly available HMPV complete genomes from GenBank. We found that

Received 24 December 2017 **Accepted** 5 January 2018 **Published** XXX

Citation Di Paola N, Cunha MDP, Durigon GS, Berezin EN, Durigon E, de Oliveira DBL, Zanotto PMDA. 2018. Complete genome sequence of a human metapneumovirus isolate collected in Brazil. Genome Announc 6:e01597-17. https://doi.org/10.1128/genomeA.01597-17.

Address correspondence to Paolo Marinho de Andrade Zanotto, pzanotto@usp.br.