

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,300

Open access books available

116,000

International authors and editors

130M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Mosquito-Borne Viral Diseases: Control and Prevention in the Genomics Era

*Vagner Fonseca, Joilson Xavier, San Emmanuel James, Tulio de Oliveira, Ana Maria Bispo de Filippis, Luiz Carlos Junior Alcantara and Marta Giovanetti*

## Abstract

Mosquito-borne viral diseases are infections transmitted by the bite of infected mosquitoes. The burden of these diseases is highest in tropical and subtropical areas and they disproportionately affect the poorest populations. Since 2014, major outbreaks of dengue, chikungunya, yellow fever and Zika have afflicted populations and overwhelmed health systems in many countries. Distribution of mosquito-borne diseases is determined by complex demographic, environmental and social factors, causing diseases to emerge in countries where they were previously unknown. Coupling genomic diagnostics and epidemiology to innovative digital disease detection platforms raises the possibility of an open, global, digital pathogen surveillance system. Considering pathogen surveillance in mind, real-time sequencing, bioinformatics tools and the combination of genomic and epidemiological data from viral infections can give essential information for understanding the past and the future of an epidemic, making possible to establish an effective surveillance framework on tracking the spread of infections to other geographic regions.

**Keywords:** mosquito-borne viral diseases, arboviral infections, genomics epidemiology, next-generation sequencing, genomic surveillance, viral pathogens

## 1. Introduction

Mosquito-borne viral diseases have lately integrated worldwide headlines since the emergence of arbovirus outbreaks in big urban areas. According to the World Health Organization, more than 17% of all infectious diseases registered worldwide are represented by vector-borne diseases, and they account for more than 700,000 deaths annually [1]. Due to this scenario of increasing cases number and expansion to new areas, the spread of infectious diseases was listed second in the top 10 risks in term of impact according to the Global Risks 2015 report [2].

Mosquitos of the genus *Aedes* have been responsible for the emergence and re-emergence of many arboviral diseases worldwide [3]. The species *Aedes aegypti* is the main vector species responsible for the major arbovirus epidemics recorded in recent years [4]. The species *A. aegypti* and *A. albopictus* are possibly suitable to survive and establish in 215 countries/territories, and their expanding range

is underlined by the increasing number of countries reporting transmission of mosquito-borne viruses. Transmissions of arboviruses, such as Zika, dengue, chikungunya, yellow fever, and Rift Valley fever, have been reported in 85, 111, 106, 43, and 39 countries, respectively [5]. Projections indicated that 3.83 billion people are living in areas prone to transmission of dengue and it is predicted that by 2050 large increases in dengue suitability will be seen in southern Africa and in the Sahel in West Africa [14]. Bhatt et al. projected the global burden of dengue around the world whose estimate indicated that 96 million dengue infections occur per year worldwide and this number represents infections that manifest at any level of the disease severity [6]. The Americas, comprising North and South America, registered more than 2 million dengue cases in 2016, and more than 1.4 million cases in 2019 [7]. For chikungunya fever, the Americas registered more than 94,000 cases in 2018, and in that same region, Zika fever accounted for more than 650,000 cases in 2016 [8, 9]. High number of cases of arboviral diseases was also registered in other regions in recent years, such as in the western pacific region where more than 375,000 suspected dengue cases were reported in 2016 [10]. In Africa, the government of Congo reported 6149 suspected cases of chikungunya until April 2019, and more than 13,000 chikungunya cases were reported in Sudan until October 2018 [11, 12]. The increasing in frequency and distribution of arboviral diseases in recent years represents a worrying burden not only for the public health system, but also for the economic sector [3]. Some estimates of the economic costs of arboviral infections have been made and for the case of dengue infections, it has been estimated that the median cost of all reported dengue hospital admissions registered in a municipality from Brazil was US\$ 259.9 per hospitalization [13, 17]. Also, in Maldives, in the Indian Ocean, dengue fever represented a total cost of \$3 million in 2015 [14]. Another estimate indicated that West Nile fever hospitalized cases in US represented a total cumulative cost of \$778 million between 1999 and 2012 [15].

Dengue and chikungunya are two arboviral diseases present in the list of neglected tropical diseases from the World Health Organization. Neglected tropical diseases are a group of diseases that have received insufficient public attention, thrive in tropical and subtropical areas, and strongly affect populations living in poverty [12]. It is argued that arboviruses can be considered a group of neglected tropical diseases, since they can have a long-lasting impact in the health and economic life of affected populations [16]. Some studies have argued that socioeconomic factors and land-use changes associated with the effects of climate change and global travel, and trade modulate the dynamics of expansion of emerging and re-emerging mosquito-borne diseases [17–20]. Movement of people between neighboring countries has been considered a good predictor for chikungunya spread in the Caribbean and Indian Ocean [14]. The expansion of the geographic distribution of arbovirus has significant negative impact on public health in many regions of the world. As measures to reduce such impacts, it has been argued about the relevance to public health of the implementation of a surveillance system that monitors virus diffusion and the appearance of new genetic variants [21]. In this sense, the use of genomic sequencing data and bioinformatics has been employed in the study of virus evolution, aiming to elucidate phylogenetic relationships and patterns of virus spread during an epidemic [22].

## **2. Genomic surveillance**

Infectious diseases continue to be one of the leading causes of death worldwide [23] and pathogens such as viruses can evolve and spread rapidly, leading to the

emergence of newly-mutated human pathogens, more virulent strains, as well as antibiotic and drug resistant organisms [24, 25]. In this context, genomic surveillance aims are to: (i) to perform global surveillance of pathogens using whole genome sequencing and (ii) to understand drug resistance, emergence and spread of viral pathogens. Several approaches have been developed and are widely used for the quick detection and identification of viral pathogens (i.e., diagnostics). Some of them are based on different serological and molecular strategies including, for example, assays based on real-time polymerase chain reaction [26]. Even though these kinds of approaches present high sensitivity and specificity for their purpose, they are more suitable for diagnostics only and cannot provide detailed genomic information [27].

Bearing these limitations in mind, the main point of developing new genomic surveillance tools is to answer the following inquiry: what sort of questions is important for genomic surveillance that cannot be addressed by conventional RT-qPCR or serology? (i) RT-qPCR assays do not allow genotype classification, neither does it help identify particular and/or characteristic transmission routes; (ii) RT-qPCR assays also do not allow to determine how fast a viral pathogen is being transmitted and in what direction it is spreading; (iii) serological and molecular assays also cannot help identify epidemiologically linked individuals, neither predict future outbreaks; and (iv) finally, serological and some molecular approaches cannot help to identify novel pathogenic agents and are, therefore, unsuitable for pathogen discovery [27].

Next generation sequencing (NGS) technologies produce significantly more raw data than other molecular diagnostic assays, including Sanger sequencing, and are also capable of informing not just pathogen diagnostics but also epidemiology [28]. This is why whole genome sequencing of viral genomes by using new technologies plays an important role in the fight against emerging and re-emerging epidemics [29, 30]. The availability of high-throughput sequencing has also provided immense insights into the ecology of health care-associated pathogens [31]. Therefore, real-time sequencing of entire pathogen genomes has become a standard and indispensable research tool for the critical role of genomic surveillance in the prevention and control of emerging infectious diseases [32], which justifies why NGS can be considered a powerful strategy that also allows the discovery of novel potential viral pathogens [33, 34].

Considering pathogen surveillance in mind, bioinformatics tools and the combination of genomic and epidemiological data from viral infections can give essential information for understanding the past and the future of an epidemic, because genomic data generated by real-time sequencing can provide important information on how and when viruses were introduced in a particular site, their pattern and determinants of dissemination in neighboring locations and the extent of genetic diversity, i.e., its dynamics, making it possible to establish an effective surveillance framework on tracking the spread of infections to other geographic regions [21, 22, 34]. In this context, recently established international networks for real-time, portable genomic sequencing, genomic surveillance and data analysis made it possible to monitor the evolution of viral genomes, to understand the origins of outbreaks and epidemics, to predict future outbreaks and to assist in the maintenance of updated diagnostic methods [33–35]. Additionally, genomic surveillance framework allows to determine, through genome sequencing, the real-time molecular epidemiology of viruses circulating and co-circulating in different regions in a specific area, and also to detect and characterize the early emergence of new pathogens in large urban centers, generating data that can inform outbreak control responses [27, 34]. Generated data regarding the molecular, epidemiological, phylogenetic and geographical aspects of circulating viral pathogens in a specific setting contribute to a better understanding of those

viral infections in a national and international context, assuming an important role in solving issues relevance to Public Health [35]. As a result, studies involving more in-depth molecular and dispersion analysis of circulating pathogens may help the World Health Organization appropriately adopt measures to control epidemics and to monitor the dynamics and spreading of new viral strains. However, even though NGS has advantages over diagnostics routine, all of the different strategies and technologies, developed by Illumina, Thermo Scientific, Oxford Nanopore and others, are not yet considered a panacea. Remaining challenges include dealing with high data throughput, which requires sophisticated computational processing as well as the annotation of large amounts of sequencing data, high DNA or RNA input sample requirements (in some cases hundreds of nanograms), which often raises the need for previous PCR-based amplification approaches. On top of all this, there are relatively few researchers in the area with sufficient bioinformatics expertise and who are able to engage in near-patient or disease surveillance activities [35].

### **3. Bioinformatics tools and phylogenetic tools**

The advent of next generation sequence (NGS) and advancements in bioinformatics present an opportunity to tap into new insights that are crucial to the establishment of an open, global digital surveillance system. NGS technologies have enabled the production and deposit of vast amounts of whole genomes into public repositories [36–38] ushering the field of genomics into era of big data. This has in turn increased the scale of genomic studies from the analysis of single or few genomes to an ever-increasing large number of genomes [39, 40].

Toward the development of global surveillance system, bioinformatics provides the tools to answer pertinent questions including the identification of organisms responsible for an outbreak, the source of an outbreak and evolutionary information of pathogens crucial for understanding the unique phenotypes such as drug resistance, virulence and disease outcome.

Several bioinformatics tools and pipelines have been developed to facilitate the processing, analysis and visualization of these data in order to derive useful information from it [41]. The major fields of interest addressed by these tools include comparative genomics which involves comparing the genetic content of one organism against that of another; prediction of the function of genes and sequences of the coding regions; identification of evolutionary events and inference of phylogenetic relationships. These fields of study play a critical role in elucidating pathogen evolution, niche adaptation, population structure and host-pathogen interaction. Furthermore, these findings inform vaccine and drug design, as well as the identification of virulence genes.

### **4. Bioinformatics pipelines and workflows**

Bioinformatics pipelines and workflows comprise of a series of third-party executable command line software assembled to perform a specific task or analysis. A complete pipeline will, therefore, be able to support the end of analysis of a given field of study such as phylogenetics or variant detection. Pipelines can thus be broken down into two major components i.e. the data processing component and the analytical component that performs the core analysis of the pipeline. Below, we review some of the prominent bioinformatics pipelines and workflows that support the processing and analysis of NGS data to provide insights on relevant global surveillance of arboviral outbreaks.

## 5. Virus discovery and identification tools

Viral discovery and identification from isolates and metagenomic samples present major challenges to bioinformatics in general. This is because viral genomes are prone to very high variability and deviation from reference genomes [42], continuous emergence of new viruses with no available references, high intrapopulation diversity, and the relative rareness of viral DNA fragments in metagenomic samples [43]. These challenges have largely been addressed through the following pipelines.

### 5.1 Genome detective

Genome detective (<http://www.genomedetective.com/app/>) is an easy to use web-based software application that assembles the genomes of viruses quickly and accurately, designed to generate and analyze whole or partial viral genomes directly from NGS reads within minutes [44]. The application gains accuracy by using a novel alignment method that uses a combination of both amino acids and nucleotide scores to construct genomes by reference-based linking of de novo contigs. Speed and accuracy are also gained by using DIAMOND [45] with a UniProt90 reference dataset to sort viral taxonomy units. The use of DIAMOND and UniRef90 allowed genome detective to identify viral short reads at least 1000 times faster than when Blastn and the viral nucleotide database of NCBI were used. The software was optimized using synthetic datasets to represent the great diversity of virus genomes. The application was then validated with next-generation sequencing data of hundreds of viruses.

### 5.2 VirusTAP: viral genome-targeted assembly pipeline

One of the major difficulties in this process is the correct de novo assembly of viral genomes from crude metagenomic deep sequencing reads, including large amounts of bacteria and human related sequencing reads. Such read contaminations often force the server to overload during de novo assembly and might cause misassembly of the resultant contigs. Pre-filtering by host-mapping subtraction could lead to efficient de novo assembly, allowing the rapid and accurate procurement of a complete viral genome sequence. In addition to the accuracy of de novo assembly, the exclusion of human-related sequences can circumvent conflicting ethical issues by avoiding analyzing the personal genetic information of patients [46, 47].

VirusTAP is web-based, integrated NGS analysis tool designed to facilitate rapid and accurate viral genome assembly from raw reads by just clicking on several selections. Like genome detective, it ensures that non-viral reads are eliminated prior to de novo assembly in order to ensure performance is not compromised.

### 5.3 Virus identification pipeline (VIP)

VIP (<https://github.com/keylabivdc/VIP>) is a web-based virus discovery and identification tool [46]. With a single click, it will filter out background-related reads, classify reads on basis of nucleotide and remote amino acid homology, and perform phylogenetic analysis to provide evolutionary insights.

### 5.4 TAR-VIR: a pipeline for TARgeted VIRal strain reconstruction from metagenomic data

TAR-VIR is a non-reference based NGS analysis tool for the reconstruction of viral strains from metagenomic samples [46, 47]. It was developed to classify RNA

viral reads from viral metagenomic data and also to produce the assembled viral strains (i.e. haplotypes) from classified reads. It mainly has two components: (1) viral read classification using partial or remotely related reference genomes and (2) de novo assembly of viral haplotypes from recruited reads with PEHaplo [47, 48], which is a haplotype reconstruction tool. As TAR-VIR has a modular structure, the users have options to use other assembly tools after read classification in step (1).

## **6. Genotyping tools**

While variant discovery and identification tools play a critical role in determining the pathogen responsible for the infection, they are unable to determine the subtype or quasispecies that is responsible for the outbreak. Arboviruses exist as a mixed population of genomic variants due to rapid replication and the error prone nature of viral RNA-dependent RNA polymerase (RdRp) [47]. Monitoring virus genotype diversity is therefore crucial to understand the emergence and spread of outbreaks. Genotyping tools provide an efficient workflow to enable researchers and public health practitioners to determine the strain that is responsible for the outbreak.

Most free-access bioinformatics programs used to classify the genetic profile of subtypes, genotypes, subgroups or groups of viruses are based on the use of similarity search tools to determine the genotype of a new sequence. These genotyping tools use a set of reference sequence genomes, carefully selected for the purpose of representing each individual genotype. The use of a number of reference sequences representing the genotype of a given group increases the consistency and reproducibility of the data, thus ensuring a higher speed in the search for the data and offering greater and more complete information while ensuring that the results are not limited to an inadequate set of reference sequences that do not represent the information needed to identify the virus.

The similarity-based methods are useful for identifying recombination patterns in viral sequences, but they need further confirmation of their own phylogenetic methods and have no statistical support for their results.

Recently [49], four viral genotyping tools for yellow fever (YFV) (<https://www.genomedetective.com/app/typingtool/yellowfevervirus/>), dengue (DENV) (<https://www.genomedetective.com/app/typingtool/dengue/>), Chikungunya (CHIKV) (<https://www.genomedetective.com/app/typingtool/chikungunya/>) and Zika (ZIKV) (<https://www.genomedetective.com/app/typingtool/zika/>) were developed and linked to genome detective to enable phylogenetic classification below species level [50, 51].

### **6.1 Castor**

The classification and annotation of virus genomes constitute important assets in the discovery of genomic variability, taxonomic characteristics and disease mechanisms. Existing classification methods are often designed for specific well-studied families of viruses [43]. Thus, the viral comparative genomic studies could benefit from more generic, fast and accurate tools for classifying and typing newly sequenced strains of diverse virus families.

CASTOR is a virus classification platform based on machine learning methods, inspired by a well-known technique in molecular biology: restriction fragment length polymorphism [52]. It simulates, *in silico*, the restriction digestion of genomic material by different enzymes into fragments. It uses two metrics to construct feature vectors for machine learning algorithms in the classification step. The performance of CASTOR, its genericity and robustness could permit the conduct of

novel and accurate large-scale virus studies. The CASTOR web platform provides an open access, collaborative and reproducible machine learning classifiers. CASTOR can be accessed at (<http://castor.bioinfo.uqam.ca>).

## 7. Phylogenetic and phylodynamic tools

Phylogenetic tools are an extremely important resource used in the field of virology to study viral evolution, trace the origin of epidemics, establish the mode of transmission, investigate the occurrence of drug resistance or determine the origin of the virus in different body compartments. Thus, the tools developed by bioinformatics are fundamental to monitor the evolution of viral diversity, supporting studies of genomic sequence analysis, crucial for the surveillance of viral polymorphism, the development of new therapeutic strategies, the development of vaccine products or the appropriate choice products. Toward the development of a global surveillance outbreak surveillance system, the advances below have been made.

### 7.1 Nextstrain (<https://nextstrain.org/>)

Nextstrain is a real-time pathogen evolution tracking platform that implements cutting-edge analysis and visualization of pathogen genome data [53]. It provides evolutionary information in the form of interactive visualizations to virologists, epidemiologists, public health officials and citizen scientists. It has been used to track various arboviral epidemics globally including West Nile Virus (WNV) in the Americas, Zika virus in 33 countries and Dengue virus outbreaks in 64 countries. The platform is continually updated with publicly available datasets to provide new insights into viral epidemic outbreaks globally in an intuitive and visually esthetic manner.

## 8. Functional prediction tools

In disease surveillance, understanding the effect of mutations detected in the viral genomes through the methods identified above is invaluable in the development of relevant controls and interventions [47]. Many of these mutations serve as drug targets as well as provide insights into the response mechanism of the pathogens to existing interventions. A global surveillance system would therefore be incomplete without the capability to provide insights to the function of discovered mutations. Below we explore some of the tools that have been applied to understand the functional relevance of mutations found in arboviruses.

### 8.1 The SIFT (sorting intolerant from tolerant)

The SIFT algorithm predicts the effect of coding variants on protein function [54, 55]. Since its introduction in 2001, SIFT has become one of the standard tools for characterizing missense variation. It has a corresponding website that provides users with predictions on their variants.

## 9. Conclusion

Augmenting epidemiological data with insights from genomic data provides a powerful tool for surveillance and control of disease outbreaks. Advances in

bioinformatics particularly leverage large genomic datasets to determine pathogenic organisms responsible for the outbreak, the origin of the infection and mutations responsible unique phenotypic traits. This information is crucial for effective planning interventions and combating outbreaks. An area of research interest that remains to be explored is the development of online platforms to perform functional analyses of statistically significant mutations in arboviruses. This information is invaluable in the development of vaccines and identification of drug targets.

## Acknowledgements

This work was supported by the ZiBRA2 project supported by the Brazilian Ministry of Health (SVS-MS) and the Pan American Organization (OPAS) and founded by Decit/SCTIE/MoH and CNPq (440685/2016-8 and 440856/2016-7); by CAPES (88887.130716/2016-2100, 88881.130825/2016-2100 and 88887.130823/2016-2100). MG is supported by Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro—FAPERJ.

## Conflict of interest

The authors declare no conflict of interest.

## Appendices and nomenclature

RT-qPCR	real time quantitative polymerase chain reaction
NGS	next generation sequencing
DNA	deoxyribonucleic acid
RNA	ribonucleic acid
VIP	virus identification pipeline
TAR-VIR	targeted viral
RdRp	RNA-dependent RNA polymerase
YFV	yellow fever virus
DENV	dengue virus
CHIKV	Chikungunya virus
ZIKV	Zika virus
WNV	West Nile virus
SIFT	sorting intolerant from tolerant

IntechOpen

## Author details

Vagner Fonseca<sup>1,2</sup>, Joilson Xavier<sup>2</sup>, San Emmanuel James<sup>1</sup>, Tulio de Oliveira<sup>1</sup>,  
Ana Maria Bispo de Filippis<sup>3</sup>, Luiz Carlos Junior Alcantara<sup>2,3</sup>  
and Marta Giovanetti<sup>2,3\*</sup>

1 KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), College of Health Sciences, University of KwaZuluNatal, Durban, South Africa

2 Laboratório de Genética Celular e Molecular, ICB, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

3 Laboratório de Flavivírus, Instituto Oswaldo Cruz Fiocruz, Rio de Janeiro, Brazil

\*Address all correspondence to: [giovanetti.marta@gmail.com](mailto:giovanetti.marta@gmail.com)

## IntechOpen

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

## References

- [1] WHO. Vector-Borne Diseases. 2017. Available from: <https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases>
- [2] World Economic Forum. Global Risks 2015. World Economic Forum. Insight Report. 10th ed 2015. Available from: [http://www3.weforum.org/docs/WEF\\_Global\\_Risks\\_2015\\_Report15.Pdf](http://www3.weforum.org/docs/WEF_Global_Risks_2015_Report15.Pdf)
- [3] LaBeaud AD. Why arboviruses can be neglected tropical diseases. *PLoS Neglected Tropical Diseases*. 2008;**25**:6-247. DOI: 10.1371/journal.pntd.0000247
- [4] Powell JR. Mosquito-borne human viral diseases: Why *Aedes aegypti*? *The American Journal of Tropical Medicine and Hygiene*. 2018;**98**:1563-1565. DOI: 10.4269/ajtmh.17-0866
- [5] Leta S, Beyene TJ, De Clercq EM, Amenu K, Kraemer MU, Revie CW. Global risk mapping for major diseases transmitted by *Aedes aegypti* and *Aedes albopictus*. *International Journal of Infectious Diseases*. 2018;**67**:25-35. DOI: 10.1016/j.ijid.2017.11.026
- [6] Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. *Nature*. 2013;**496**:504-544. DOI: 10.1038/nature12060.
- [7] PAHO. Dengue and Severe Dengue, Cases and Deaths for Subregions of the Americas. 2019. Available from: <http://www.paho.org/data/index.php/en/mnu-topics/indicadores-dengue-en/dengue-regional-en/261-dengue-reg-ano-en.html>
- [8] PAHO. 2019. Chikungunya Total Cases. Available from: <http://www.paho.org/data/index.php/en/mnu-topics/chikv-en/551-chikv-subregions-en.html>
- [9] PAHO. Zika Total Cases. 2019. Available from: <http://www.paho.org/data/index.php/en/mnu-topics/zika.html>
- [10] WHO. Neglected Tropical Diseases in the Eastern Mediterranean Region. 2019. Available from: [https://apps.who.int/iris/bitstream/handle/10665/275463/Fact\\_Sheet\\_CDT\\_2018\\_EN\\_20491.pdf?ua=1](https://apps.who.int/iris/bitstream/handle/10665/275463/Fact_Sheet_CDT_2018_EN_20491.pdf?ua=1)
- [11] WHO. Dengue and Severe Dengue. 2019. Available from: <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>
- [12] WHO. Emergencies Preparedness, Response. Chikungunya. 2018. Available from: <https://www.who.int/csr/don/archive/disease/chikungunya/en/>
- [13] Schar DL, Yamey GM, Machalaba CC, Karesh WB. A framework for stimulating economic investments to prevent emerging diseases. *Bulletin of the World Health Organization*. 2018;**96**:138. DOI: 10.2471/BLT.17.199547
- [14] Messina JP, Brady OJ, Golding N, Kraemer MU, Wint GW, Ray SE, et al. The current and future global distribution and population at risk of dengue. *Nature Microbiology*. 2019;**1**:10-11. DOI: 10.1038/s41564-019-0476-8
- [15] Franklinos LH, Jones KE, Redding DW, Abubakar I. The effect of global change on mosquito-borne disease. *The Lancet Infectious Diseases*. 2019;**01**:18-124. DOI: 10.1525/abt.2017.79.3.169
- [16] Zanotto PMA, Leite LCC. The challenges imposed by dengue, Zika, and chikungunya to Brazil. *Frontiers in Immunology*. 2018;**9**:1960-1964. DOI: 10.3389/fimmu.2018.01964
- [17] Machado AA, Estevan AO, Sales A, da Silva Brabes KC, Croda J, Negrão FJ.

- Direct costs of dengue hospitalization in Brazil: Public and private health care systems and use of WHO guidelines. *PLoS Neglected Tropical Diseases*. 2014;**4**:8-104. DOI: 10.1371/journal.pntd.0003104
- [18] Bangert M, Latheef AT, Pant SD, Ahmed IN, Saleem S, Rafeeq FN, et al. Economic analysis of dengue prevention and case management in the Maldives. *PLoS Neglected Tropical Diseases*. 2018;**27**:12-96. DOI: 10.1371/journal.pntd.0006796
- [19] Staples JE, Shankar MB, Sejvar JJ, Meltzer MI, Fischer M. Initial and long-term costs of patients hospitalized with West Nile virus disease. *The American Journal of Tropical Medicine and Hygiene*. 2014;**3**:402-409. DOI: 10.4269/ajtmh.13-0206
- [20] Rossi G, Karki S, Smith RL, Brown WM, Ruiz MO. The spread of mosquito-borne viruses in modern times: A spatio-temporal analysis of dengue and chikungunya. *Spatial and Spatio-temporal Epidemiology*. 2018;**26**:113-125. DOI: 10.1016/j.sste.2018.06.002
- [21] Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nature Reviews Genetics*. 2017;**1**:2-256. DOI: 10.1038/nrg.2017.88
- [22] Grubaugh ND. Tracking virus outbreaks in the twenty-first century. *Nature Microbiology*. 2019;**4**:10-19. DOI: 10.1038/s41564-018-0296-2
- [23] Morens DM, Folkers GK, Fauci AS. The challenge of emerging and re-emerging infectious diseases. *Nature*. 2004;**430**:242-249. DOI: 10.1038/nature02759
- [24] Daszak P, Cunningham AA, Hyatt AD. Emerging infectious diseases of wildlife—Threats to biodiversity and human health. *Science*. 2000;**287**:443-449. DOI: 10.1126/science.287.5452.443
- [25] Morse SS. Factors in the emergence of infectious diseases. *Emerging Infectious Diseases*. 1995;**1**:7-15. DOI: 10.3201/eid0101.950102
- [26] Versalovic J, Lupski JR. Molecular detection and genotyping of pathogens: More accurate and rapid answers. *Trends in Microbiology*. 2002;**10**:15, 12377563-21
- [27] Sabat AJ, Budimir A, Nashev D, Sá-Leão R, van Dijk JM, Laurent F. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Euro Surveillance*. 2013;**18**:20-380. DOI: 10.2807/ese.18.04.20380-en
- [28] Shendure J, Ji H. Next-generation DNA sequencing. *Nature Biotechnology*. 2008;**26**:1135-1145. DOI: 10.1038/nbt1486
- [29] Haagmans BL, Andeweg AC, Osterhaus ADME. The application of genomics to emerging zoonotic viral diseases. *PLoS Pathogens*. 2009;**5**:100-557. DOI: 10.1371/journal.ppat.1000557
- [30] McHardy AC, Adams B. The role of genomics in tracking the evolution of influenza A virus. *PLoS Pathogens*. 2009;**5**:10-56. DOI: 10.1371/journal.ppat.1000566
- [31] Tang P, Gardy JL. Stopping outbreaks with real-time genomic epidemiology. *Genome Medicine*. 2014;**6**:1-104. DOI: 10.1186/s13073-014-0104-4
- [32] Holmes EC. Viral evolution in the genomic age. *PLoS Biology*. 2007;**5**:2-78. DOI: 10.1371/journal.pbio.0050278
- [33] Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K. Multiplex PCR method for MinION and illumina sequencing of Zika and other

- virus genomes directly from clinical samples. *Nature Protocols*. 2017;**12**: 12-61. DOI: 10.1038/nprot.2017.066
- [34] Thézé J, Li T, du Plessis L, Bouquet J, Kraemer MU, Somasekar S. Genomic epidemiology reconstructs the introduction and spread of Zika virus in Central America and Mexico. *Cell Host & Microbe*. 2018;**23**:855-864. DOI: 10.1016/j.chom.2018.04.017
- [35] Loman NJ, Constantinidou C, Chan JZM, Halachev M, Sergeant M, Penn CW. High-throughput bacterial genome sequencing: An embarrassment of choice, a world of opportunity. *Nature Reviews Microbiology*. 2012;**10**:599-606. DOI: 10.1038/nrmicro2850
- [36] AL-Dewik NI, Qoronfleh MW, et al. *Advances in Public Health*. 2019;**2**:44-76. DOI: 10.1155/2019/3807032
- [37] Zhang J, Chiodini R, Badr A, Zhang G. The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics*. 2011;**38**:95-109. DOI: 10.1016/j.jgg.2011.02.003
- [38] Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The next-generation sequencing revolution and its impact on genomics. *Cell*. 2013;**155**:27-38. DOI: 10.1016/j.cell.2013.09.006
- [39] Elliott LT, Sharp K, Alfaro-Almagro F, Shi S, Miller KL, Douaud G, et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature*. 2018;**562**:210-216. DOI: 10.1038/s41586-018-0571-7
- [40] Hamid JS, Hu P, Roslin NM, Ling V, Greenwood CM, Beyene J. Data integration in genetics and genomics: Methods and challenges. *Human Genomics and Proteomics*. 2009;**86**: 90-93. DOI: 10.4061/2009/869093
- [41] Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Experimental & Molecular Medicine*. 2018;**50**:8-96. DOI: 10.1038/s12276-018-0071-8
- [42] Manso CF, Bibby DF, Mbisa JL. Efficient and unbiased metagenomic recovery of RNA virus genomes from human plasma samples. *Scientific Reports*. 2017;**7**:41-73. DOI: 10.1038/s41598-017-02239-5
- [43] Rose R, Constantinides B, Tapinos A, et al. Challenges in the analysis of viral metagenomes. *Virus Evolution*. 2016;**2**:01-22. DOI: 10.1093/ve/vew022
- [44] Vilsker M, Moosa Y, Nooij S, Fonseca V, Ghysens Y, Dumon K, et al. Genome detective: An automated system for virus identification from high-throughput sequencing data. *Bioinformatics*. 2018;**2**:23-98. DOI: 10.1093/bioinformatics/bty695
- [45] Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nature Methods*. 2014;**12**:20-59. DOI: 10.1038/nmeth.3176
- [46] Remita AM, Halioui A, Diouara AAM, Daigle B, Kiani G, Diallo AB. A machine learning approach for viral genome classification. *BMC Bioinformatics*. 2017;**18**:2-08. DOI: 10.1186/s12859-017-1602-3
- [47] Chen J, Huang J, Sun Y. TAR-VIR: A pipeline for TARgeted VIRal strain reconstruction from metagenomic data. *BMC Bioinformatics*. 2019;**20**:3-05. DOI: 10.1186/s12859-019-2878-2
- [48] Chen J, Zhao Y, Sun Y. De novo haplotype reconstruction in viral quasispecies using paired-end read guided path finding. *Bioinformatics*. 2018;**34**:2927-2935. DOI: 10.1093/bioinformatics/bty202

[49] Faria NR. Genomic and epidemiological monitoring of yellow fever virus transmission potential. *Science*. 2018;**36**:894-899. DOI: 10.1126/science.aat7115

[50] Fonseca V, Libin PJK, Theys K. A computational method for the identification of dengue, Zika and Chikungunya virus species and genotypes. *PLoS Neglected Tropical Diseases*. 2019;**13**:7-231. DOI: 10.1371/journal.pntd.0007231

[51] Remita MA, Halioui A, Malick Diouara AA, Daigle B, Kiani G, Diallo AB. A machine learning approach for viral genome classification. *BMC Bioinformatics*. 2017;**18**:208

[52] Hadfield J, Megill C, Bell SM. Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics*. 2018;**34**:4121-4123. DOI: 10.1093/bioinformatics/bty407

[53] Alexander TC, Laura DK. Insights into arbovirus evolution and adaptation from experimental studies. *Viruses*. 2010;**12**:2594-2617. DOI: 10.3390/v2122594

[54] Li Y. VIP: An integrated pipeline for metagenomics of virus identification and discovery. *Scientific Reports*. 2016;**6**:23-774. DOI: 10.1038/srep23774

[55] Sim N-L, Kumar P, Hu J. SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*. 2012;**40**:452-457. DOI: 10.1093/nar/gks539