

Integrated Environmental and Genetic Data Analysis for Detection and Prediction of Pathogen Mutations: Influenza virus, Plasmodium falciparum, Dengue virus, and Vibrio cholerae in Yemen

Hussein Dedy*

Yemen city Hodeidah, Studies and Research Program, Authority of Public AL-Thawra Hospital, Hodeidah City, Yemen

*Corresponding Author

Hussein Dedy, Yemen city Hodeidah, Studies and Research Program, Authority of Public AL-Thawra Hospital, Hodeidah City, Yemen.

Submitted: 2025, Apr 15; **Accepted:** 2025, May 23; **Published:** 2025, Jun 09

Citation: Dedy, H. (2025). Integrated Environmental and Genetic Data Analysis for Detection and Prediction of Pathogen Mutations: Influenza virus, Plasmodium falciparum, Dengue virus, and Vibrio cholerae in Yemen. *Int J Aerosp Sci Technol Engg*, 1(1), 01-03.

Abstract

This study investigates the integration of genetic and environmental data to detect and predict mutations in key infectious disease agents namely Influenza virus, Dengue virus, Plasmodium falciparum (malaria), and Vibrio cholerae (cholera) in Yemen, particularly in the climate-affected region of Al-Hodeidah. By analyzing original and integrated genetic sequences alongside 50 years of climate data (2023 - 2047) , specific deletion mutations were identified and their probable timelines estimated: Influenza (10–15 years), Dengue (15–20 years), Malaria (20–25 years), and Cholera (25–30 years). Tools such as FastQC, MultiQC, MAFFT, GATK, and SnpEff were used to assess data quality, perform sequence alignment, and annotate mutations. The findings reveal a strong correlation between climatic variations and the emergence of mutations that influence pathogen virulence, transmissibility, and drug resistance. For example, deletions of nucleotides (such as AUG and AGC) altered amino acid sequences, potentially impacting protein functionality. Mutations were also detected in key genes such as PfCRT in Plasmodium and beta-lactamase genes in Vibrio cholerae, both associated with increased resistance to antimalarial and antibiotic therapies.

Additionally, variations in GC content and structural RNA elements (e.g., stem-loops) among viral genomes were linked to greater adaptability and transmission potential. This integrative approach highlights the importance of including environmental data in genomic surveillance systems to improve early detection and epidemic preparedness. The study recommends the adoption of machine learning models, real-time mutation databases, and stronger collaboration between blood banks and genomic laboratories to develop predictive tools for pathogen evolution. This approach offers critical insights for public health strategies, especially in regions vulnerable to climate change and disease outbreaks.

1. Introduction

As climate change intensifies, the need to understand the relationship between environmental factors and genetic mutations in pathogens becomes increasingly critical, especially in regions prone to outbreaks like Al-Hodeidah. Infectious diseases such as cholera, malaria, dengue, and influenza continue to pose major public health threats, further complicated by emerging mutations that affect their behavior and spread. This study aims to analyze the genetic mutations of these diseases by integrating meteorological data with DNA sequences using advanced analytical tools like FastQC, GATK, MAFFT, and SnpEff. It also explores the potential

for developing early warning systems based on this integration [4,10]. Genetic data were collected and analyzed from globally recognized databases to ensure comprehensive and accurate results. The GenBank database, provided by NCBI, was utilized for its extensive repository of genetic sequences from a wide range of microorganisms, including the pathogens under investigation [11].

2. Methodology

An applied analytical approach was used, integrating: Reading of environmental and genomic data from CSV files.

Conversion of environmental and meteorological data to digital and FastQ formats for unified processing. Quality assessment using FastQC and MultiQC tools to evaluate data quality [1,2]. Defining genetic categories (pathogens: Dengue virus, Influenza, Plasmodium, and Cholera) through computational entities that simulate their genomic sequences [3,4]. Mutation analysis by comparing current data to reference genomes of the respective pathogens [5,6]. For viral mutation analysis, particularly of the influenza virus, the GISAID platform was employed, as it offers up-to-date, high-quality contextual data on viral variants and their geographic and temporal distribution [12]. Differentiating mutations between the original code (chromosomal data) and the current code (environmental-genomic data).

3. Results

Regarding the malaria parasite, data from the MalariaGEN network were incorporated, focusing on the genetic diversity of Plasmodium species and their mosquito vectors, which contributes to understanding patterns of transmission and environmental adaptation [14]. The conversion of environmental and meteorological data to FastQ format was successfully completed.

Tools like FastQC and MultiQC showed acceptable quality results for most of the data [1,2].

Extraction and analysis of genetic mutations were carried out for the four pathogens: Dengue virus, Influenza virus, Plasmodium falciparum (Malaria) and Vibrio cholerae (Cholera). Additionally, the ENSEMBL platform was used to analyze human and viral genomes, enabling the association of genetic mutations with relevant functional and regulatory data [13]. The analysis revealed clear differences in mutations between the original and current code, suggesting the impact of environmental factors on genetic variation [10]

4. Discussion

The differences in mutations between the original and current code indicate a direct impact of environmental changes on the genome. The use of meteorological data in mutation analysis paves warning systems based on environmental and genomic dynamics [10]. Fast QC and Multi QC techniques have proven effective in data quality control [1,2].

There is a need to further integrate environmental and genomic data using more advanced artificial intelligence models. In the context of analyzing the relationship between genetic mutations, disease spread, virulence, and antibiotic resistance, several pathogens, including Vibrio cholerae, Dengue virus, Influenza virus, and Plasmodium falciparum, were examined. This study provides detailed insights into how genetic mutations influence the epidemiological characteristics of these diseases, including their virulence, spread, and resistance to available treatments. The results indicate that genetic mutations play a pivotal role in the evolution of these pathogens, complicating prevention and treatment efforts.

4.1. Disease Virulence

Virulence, which represents the pathogen's ability to cause significant damage to host tissues, is closely associated with the types of genetic mutations occurring in the pathogens' genomes. In the case of Vibrio cholerae, point mutations at positions 10 (A) and 20 (C) were identified, where these mutations directly affect protein expression, enhancing the bacterium's ability to produce toxins and damage intestinal tissues. Deletion or insertion mutations affecting protein structure or function can also increase the bacterium's virulence by making it more effective at secreting toxins.

Moreover, in the case of Dengue virus and Influenza virus, similar effects of genetic mutations were observed that could contribute to increased virulence. In Dengue virus, specifically, certain mutations affect the secondary structure of RNA, including changes in the Stem-Loop structures, which are crucial for viral replication. These modifications may facilitate the virus's ability to adapt to environmental factors and evade host immune responses, thus increasing the severity of the disease.

For Plasmodium falciparum, multiple genetic mutations were observed that affect essential proteins in the parasite's life cycle, leading to more severe clinical symptoms such as high fever and severe anemia.

4.2. Disease Spread

The spread of disease is linked to the ability of pathogens to replicate in various environments and their resilience to changing conditions. In Vibrio cholerae, the length of the genetic sequences was found to correlate with disease spread. Specifically, genetic sequences ranging from 20 to 50 nucleotides were associated with a higher capacity for transmission compared to shorter sequences under 20 nucleotides. Additionally, the GC content ranging between 0.4 and 0.6 in the genetic sequences of Vibrio cholerae indicates an increased adaptability to changing environmental conditions, facilitating the pathogen's spread across different regions.

For Dengue virus and Influenza virus, analyses suggested that variations in GC content in the genetic sequences contribute to greater spread potential. Furthermore, differences in the Stem-Loop structures within these sequences appear to enhance the virus's ability to adapt to new environments, thereby increasing their potential for widespread transmission.

4.3. Antibiotic Resistance

Antibiotic resistance is one of the major challenges in treating infectious diseases. In the case of Vibrio cholerae, genetic mutations affecting enzymatic processes involved in antibiotic resistance were identified, particularly mutations in beta-lactamase genes, which contribute to resistance against beta-lactam antibiotics. Additionally, mutations in proteins related to toxin secretion can also contribute to the development of antibiotic resistance, further complicating treatment with commonly used antibiotics such as ampicillin and tetracycline. For Plasmodium falciparum, numerous studies have indicated increased resistance to antimalarial drugs,

such as chloroquine and artemisinin, through mutations in genes responsible for drug transport within the parasite. Mutations in the PfCRT gene, which is associated with chloroquine transport, represent a significant example of how the parasite develops resistance to drugs. Data also suggests that mutations affecting proteins controlling oxygen transport within the parasite play a key role in resistance to treatment.

5. Conclusions

The environment (especially climatic changes) influences the genetic makeup of pathogens [10]. Weather data can be utilized as a helpful factor in genetic mutation analysis. It is essential to have coordination between environmental and genomic data analyses achieve accurate predictive results.

The results of this study indicate that genetic mutations not only increase the virulence of pathogens but also contribute to their spread and resistance to treatment. For *Vibrio cholerae* and *Plasmodium falciparum*, genetic mutations have led to enhanced drug resistance, complicating efforts for effective treatment. Furthermore, the impact of these mutations on genetic structure and GC content may facilitate the adaptability of these organisms to diverse environments, promoting their spread. Based on these findings, it is imperative to continue research efforts to improve preventive and therapeutic strategies against these diseases.

References

1. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. [Online]
2. Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047-3048.
3. Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics*, 25(14), 1754-1760.
4. der Auwera, V. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*, 43, 1.
5. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120.
6. Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4), 772-780.
7. Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., ... & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *fly*, 6(2), 80-92.
8. World Health Organization. (2014). *Global influenza surveillance and response system (GISRS)*.
9. WHO (2022). Global report on malaria 2022.
10. Kalia, K. et al. (2021). Climate change and vector-borne diseases: evidence and implications. *Infectious Diseases of Poverty*.
11. NCBI GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) – A comprehensive database containing genomic sequences of various organisms.
12. GISAID (<https://www.gisaid.org/>) – Focuses on viruses, especially influenza and coronavirus.
13. ENSEMBL (<https://www.ensembl.org/>) – A platform for studying bacterial, viral, and human genomes.
14. MalariaGEN (<https://www.malariagen.net/>) – Specialized in genetic data related to malaria.

Copyright: ©2025 Hussein Dedy. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.