World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

Check for updates

The algorithmic cartography of influenza's genomic plasticity: Integrated genomics, high-throughput sequencing and actionable public health surveillance

Ayanfeoluwa Alabetutu [1, 2, *] and Adekunle O. Adeluwoye [2, 3]

[1] *Research Institute of Virology, Federal Research Center for Fundamental and Translational Medicine, Novosibirsk, Russia.*
[2] *Bioinformatics & Big Data Analysis Programs and Trainings, UREKA Biotec/OmicsLogic Africa, OmicsLogic Inc. USA.*
[3] *Department of Medical Laboratory Science, KolaDaisi University, Ibadan, Oyo State, Nigeria.*

World Journal of Advanced Research and Reviews, 2026, 30(01), 395-405

Publication history: Received on 20 February 2026; revised on 1 April 2026; accepted on 4 April 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.30.1.0766

Abstract

Background: Influenza viruses, with their segmented RNA genomes, exhibit remarkable genomic plasticity through antigenic drift and shift, enabling rapid evolution and persistent zoonotic threats. Traditional surveillance methods lack the resolution to monitor viral quasispecies and emerging zoonotic threats effectively.

Aim/Objectives: This comprehensive review synthesizes advancements in integrated phylogenomics and phylodynamics with computational methods for influenza surveillance that enhance the resolution, speed, and utility of influenza surveillance for public health action.

Materials and Methods: We evaluated the comparative strengths of short-read (Illumina) and long-read (Oxford Nanopore) sequencing platforms for Whole-Genome Sequencing (WGS) of influenza. We also assessed bioinformatics pipelines from raw data quality control to sophisticated algorithmic analysis including variant calling (GATK, LoFreq), reassortment mapping (RDP4, GiRaF), and computational tracking of antiviral resistance markers.

Results: Integrated genomic and phylogenetic analysis provides high-resolution characterization of viral diversity, enabling precise identification of emerging clades and adaptive mutations. NGS platforms offer distinct advantages: Illumina provides exceptional accuracy for population-level surveillance, while Oxford Nanopore enables real-time sequencing for outbreak response and long-read capability for resolving reassortment events. Advanced computational tools successfully decode viral quasispecies, track antigenic drift and shift, and identify antiviral resistance markers, transforming raw sequence data into actionable insights.

Conclusions: The integrated framework transforms sequence data into actionable public health intelligence, enabling real-time evolutionary monitoring and predictive modeling essential for effective public health preparedness.

Implication of the Study: This multi-layered surveillance paradigm guides the transition toward One Health surveillance, integrating human, animal, and environmental data for comprehensive global health security.

Keywords: Influenza genomics; Next-generation sequencing; Phylodynamics; Antiviral resistance; Reassortment; Surveillance; Whole-genome sequencing

* Corresponding author: Ayanfeoluwa Alabetutu

1. Introduction

Influenza viruses, members of the *Orthomyxoviridae* family, present a formidable and continually evolving challenge to global public health. The virus's segmented, negative-sense RNA genome, typically composed of eight distinct segments, is the biological engine driving its remarkable capacity for change.[1] This inherent genomic architecture is the substrate for two principal evolutionary mechanisms: antigenic drift, the gradual accumulation of point mutations (single nucleotide polymorphisms, or SNPs) primarily in the Hemagglutinin (HA) and Neuraminidase (NA) genes, which allows the virus to evade host immunity; and the more dramatic antigenic shift, catalyzed by the exchange of entire gene segments between co-circulating strains—a process known as reassortment.[2] It is this latter mechanism that historically underpins the emergence of pandemic strains, capable of widespread transmission in human populations lacking pre-existing immunity.

These two evolutionary mechanisms—the gradual drift and the abrupt shift—create a spectrum of genomic plasticity that continuously challenges immune recognition and vaccine efficacy. To conceptualize this dynamic, Figure 1 provides a visual summary, contrasting the incremental mutational landscape of antigenic drift with the dramatic genomic reassortment that drives antigenic shift. Understanding this foundational dichotomy is critical for appreciating the resolution required by modern genomic surveillance methods, which we detail in the following sections.

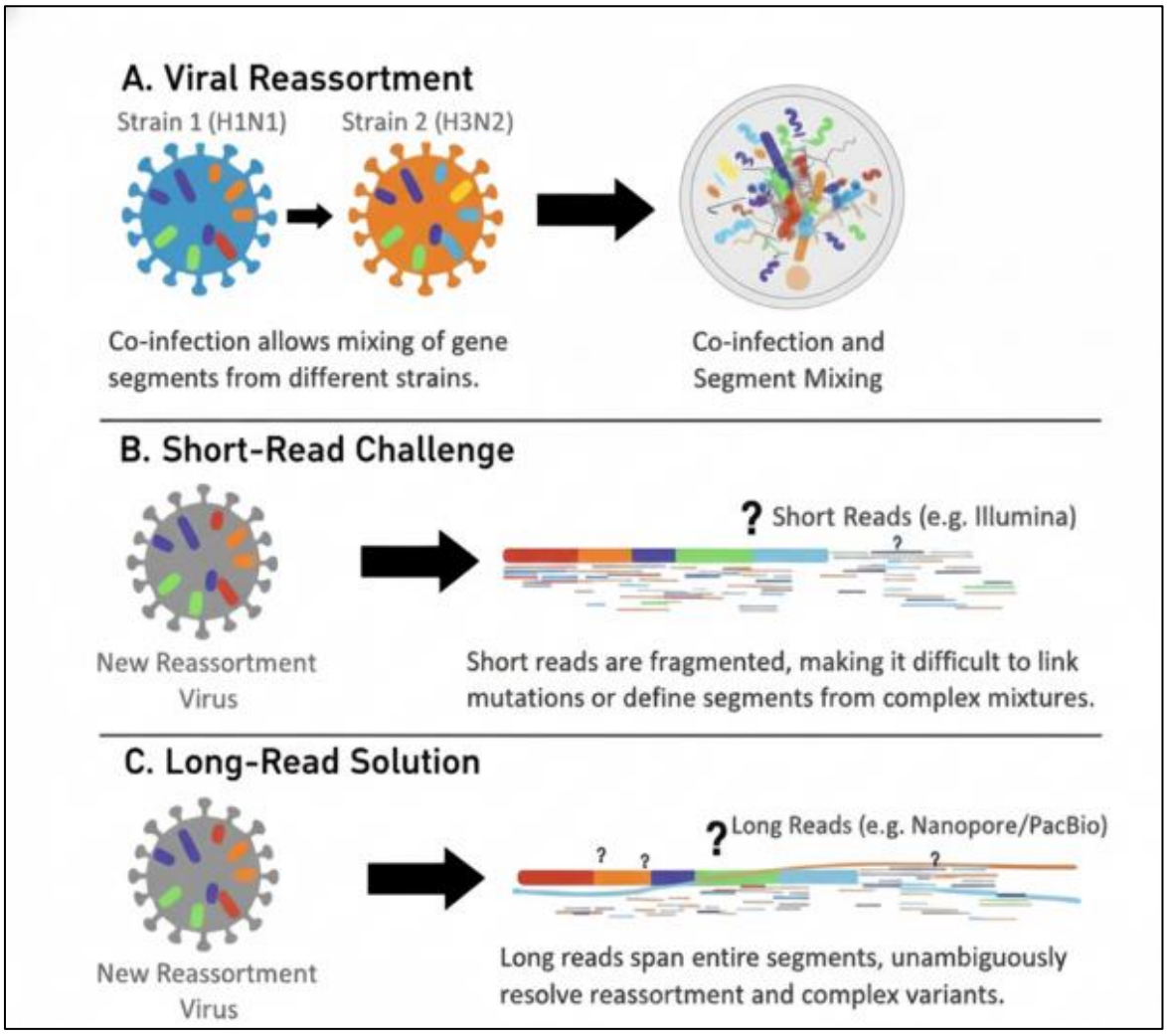

Figure 1 Illustration of Reassortment, the Short-Read Challenge, and the Role of Long Reads. A conceptual diagram illustrating viral reassortment through co-infection of two strains (H1N1 and H3N2) resulting in mixed gene segments (Panel A); the challenge of resolving reassortment in the segmented influenza genome using short reads, which are fragmented and unable to definitively link mutations across complex mixtures (Panel B); and the solution provided by long reads that span entire segments, enabling unambiguous resolution of reassortment events and complex variants (Panel C)

The public health and economic burden imposed by influenza is substantial and unrelenting. Beyond the episodic nature of pandemics, annual seasonal epidemics are estimated to cause between 290,000 and 650,000 respiratory-related deaths worldwide, alongside millions of hospitalizations and significant economic costs due to lost productivity.[3] Furthermore, the threat is not confined to seasonal human strains; the constant circulation of avian and swine influenza viruses creates a persistent risk of zoonotic spillover and subsequent adaptation to human hosts, exemplified by the ongoing concern over highly pathogenic avian influenza (HPAI) H5N1.[6] Monitoring the genetic markers of host adaptation and virulence in these animal reservoirs is a critical, yet often resource-intensive, component of global preparedness.

Crucially, the high error rate of the viral RNA-dependent RNA polymerase (RdRp) results in the existence of influenza within an infected host not as a single sequence, but as a diverse population of closely related variants – a quasispecies**.[**6] This intra-host diversity is the raw material for rapid adaptation, including the swift selection of drug-resistant or vaccine-escape mutants. Traditional surveillance methods, relying on low-throughput sequencing of consensus sequences or serological assays, inherently average out this critical quasispecies-level information.[5] While foundational, these approaches often yield insufficient granularity and speed to capture the full complexity and rapid kinetics of modern viral evolution, particularly in a world characterized by unprecedented global travel and interconnectedness.

The sheer volume and velocity of genomic data generated by contemporary Next-Generation Sequencing (NGS) technologies have exposed this analytical gap, demanding equally sophisticated computational and analytical solutions. This review argues that a surveillance paradigm rooted in the seamless integration of genomics and phylogenetics – supported by advanced computational cartography – is indispensable for effective public health preparedness. We aim to synthesize the technological advancements in WGS with the algorithmic strategies that translate raw sequence data into actionable intelligence, providing a cohesive perspective on how these methods collectively illuminate the intricate evolutionary pathways of this persistent pathogen. The subsequent sections will detail the technological platforms, the algorithmic tools for decoding plasticity, and the direct public health applications of this integrated framework.

2. The Foundational Role of Integrated Genomics and Phylogenetics

Genomics and phylogenetics are two complementary disciplines that, when integrated, provide a synergistic analytical depth for influenza surveillance. Genomics focuses on the comprehensive analysis of the viral genome, providing granular detail on genetic diversity, the presence of specific mutations, and markers associated with drug resistance.[4] Phylogenetics, conversely, employs methods to reconstruct the evolutionary history and relationships among viral strains, illuminating transmission patterns, geographic spread, and temporal evolutionary dynamics.[9]

The true transformative potential, however, is realized through their convergence, giving rise to the field of phylodynamics. Integrated genomic and phylogenetic analysis allows for the high-resolution characterization of viral diversity, facilitating the precise identification of emerging clades and adaptive mutations.[9] This combined approach is not merely academic; it translates directly into actionable intelligence by providing a temporal and spatial context for viral evolution. For example, the tracing of the 2009 H1N1 pandemic virus, which emerged from a complex reassortment of gene segments originating in swine, avian, and human influenza lineages, was a seminal demonstration of the power of this integrated approach.[4] By embedding genomic data within phylogenetic frameworks, researchers can map the spread of influenza across populations and identify specific transmission chains, effectively reconstructing the epidemiological history of an outbreak. This information is critical for informing public health interventions, such as targeted vaccination campaigns and resource allocation.[6,10] The utility of such integrated platforms has been demonstrated in contexts ranging from monitoring the 2009 H1N1 pandemic to tracking the emergence of highly pathogenic avian influenza strains like H7N9, where genomic surveillance revealed reassortment events and mutations linked to increased human virulence.[7] Furthermore, phylodynamics enables the estimation of key epidemiological parameters—such as the basic reproduction number (R_0) and the time of the most recent common ancestor (TMRCA)—directly from sequence data, offering a powerful, independent means of validating and augmenting traditional epidemiological surveillance data.[8] This convergence of evolutionary and epidemiological modeling provides a crucial advantage in anticipating the trajectory of emerging strains.

3. High-Throughput Sequencing: Generating the Genomic Map

The revolution in influenza surveillance is inextricably linked to the advent of NGS technologies, which have made WGS of viral isolates a timely and cost-effective reality.[5] NGS platforms operate on the principle of massively parallel sequencing, generating millions of reads simultaneously.

3.1. Comparative Analysis of Sequencing Platforms

The choice of sequencing platform is a critical decision, with the two dominant technologies—Illumina (short-read) and Oxford Nanopore Technologies (ONT) (long-read)—presenting distinct trade-offs (Table 1). This table compares the two main platforms discussed (Illumina and Oxford Nanopore) based on key metrics relevant to viral genomics.

Table 1 Comparison of Next-Generation Sequencing Platforms for Influenza WGS

Feature	**Illumina (e.g., MiSeq/NextSeq)**	**Oxford Nanopore Technologies (ONT) (e.g., MinION)**
Read Length	Short (typically 150-300 bp)	Ultra-long (up to >1 Mbp)
Throughput	Very High (Millions of reads/run)	Variable (Medium to High)
Accuracy (Raw)	Very High (typically >99.9%)	Lower (typically 90-99%)
Time to Result	Days (Batch processing)	Real-time (Minutes to Hours)
Cost per Base	Very Low	Low to Moderate
Key Advantage for Influenza	High accuracy for SNP/minor variant detection, High throughput for large-scale surveillance.	Real-time sequencing for outbreak response, Long reads for resolving reassortment/structural variation.
Key Limitation	Short reads complicate *de novo* assembly of segmented genomes and structural variant detection.	Higher raw error rate requires deeper coverage and robust QC/polishing pipelines.

Illumina platforms are the workhorse of large-scale, population-level surveillance, offering exceptional accuracy for detecting low-frequency variants (quasispecies) and single nucleotide precision.[5] However, their short read lengths can complicate the *de novo* assembly of the segmented influenza genome, particularly in samples containing mixed infections or novel reassortants. Conversely, ONT platforms offer the distinct advantage of ultra-long reads and real-time data acquisition. The ability to generate reads spanning the entire length of an influenza segment is invaluable for unambiguously resolving complex genomic rearrangements and reassortment events.[10] While the raw accuracy of ONT is lower, its rapid, portable nature has proven its clinical utility for integrated outbreak management and timely surveillance, especially in resource-limited settings.[10]

3.2. Sample Preparation and Upstream Quality Control

Regardless of the chosen sequencing platform, the ultimate reliability and interpretability of WGS data are fundamentally dependent upon the meticulous execution of upstream sample preparation. This process, which bridges the clinical sample with the sequencing instrument, is often the most critical bottleneck in high-throughput surveillance. It typically involves three sequential, yet intricately linked, steps: viral RNA extraction, reverse transcription into complementary DNA (cDNA), and the targeted or multiplex amplification of the eight viral segments.

The initial challenge resides in obtaining high-quality viral RNA from clinical specimens, which frequently present with low viral titers or high concentrations of inhibitory substances. The efficiency of the RNA extraction protocol – often leveraging magnetic bead-based or column-based methods – directly dictates the final yield and purity. A low viral load necessitates an aggressive amplification strategy, which in turn elevates the risk of introducing artifacts or amplifying non-target sequences, such as host or bacterial contaminants.

Following extraction, the conversion of RNA to cDNA via reverse transcription is a crucial enzymatic step. The choice of primers for this step is a delicate balance: while random hexamers offer the broadest coverage, they also increase the sequencing of host and non-viral RNA. Conversely, targeted primers (either universal influenza primers or subtype-specific primers) enhance viral yield but risk introducing a sequence bias, potentially obscuring novel variants or reassortment partners.[5] The subsequent amplification step, often a multiplex PCR, must be carefully optimized to ensure all eight segments are amplified with relatively equal efficiency. Imbalanced amplification can lead to highly uneven coverage across the genome, resulting in gaps or regions of low confidence in the final consensus sequence.

The continued refinement of these methodologies, such as the development of optimized high-throughput workflows and the integration of automated liquid handling systems, is key to maximizing the utility of NGS in influenza

research.[11] Furthermore, the implementation of pre-sequencing quality control – including fluorometric or qPCR-based quantification and size-distribution analysis of the prepared libraries – is essential to predict and preemptively mitigate sequencing failures, thereby ensuring that only high-quality, normalized libraries proceed to the expensive sequencing stage. The inherent complexity of the influenza genome, particularly its segmented nature and the prevalence of co-infection, demands that these upstream QC steps are not merely procedural, but are treated as integral components of the overall bioinformatics pipeline.

4. Algorithmic Dissection: Computational Methods for Decoding Plasticity

The journey from raw sequence data to a final, annotated genome sequence is mediated by a series of sophisticated bioinformatics steps. The sheer volume of data necessitates automated, robust, and reproducible pipelines, often leveraging cloud-based resources for scalability.[12] The reliability of the final sequence, and consequently the validity of any derived conclusions, hinges on rigorous quality control (QC) and the application of specialized algorithms.

The overall process can be conceptualized as a flow from sample collection to public health action, as illustrated in Figure 2.

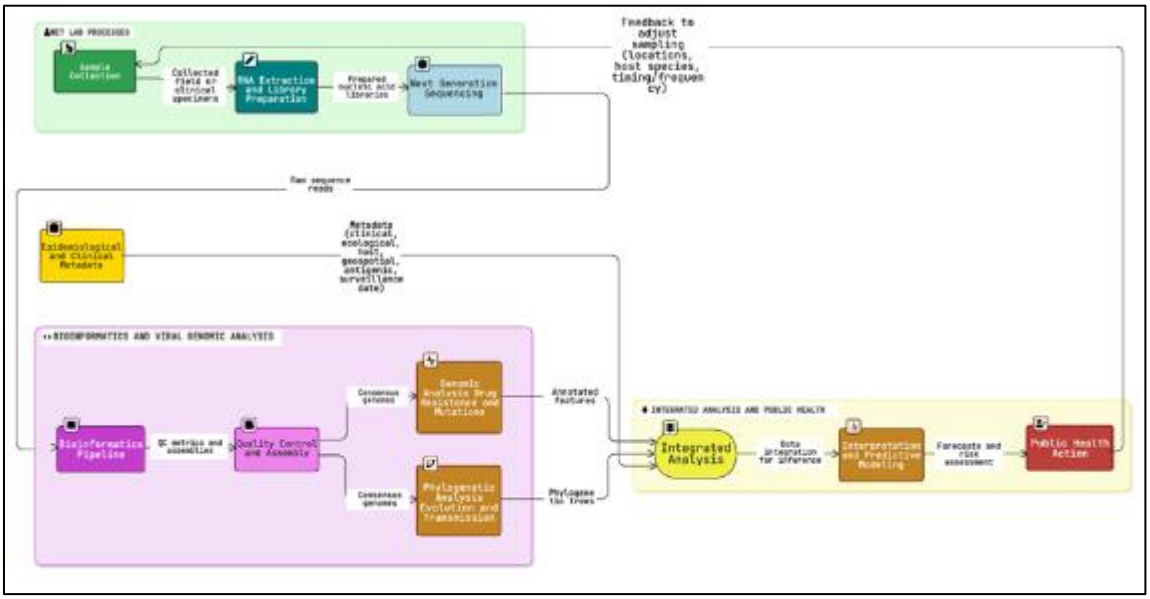

Figure 2 A conceptual flowchart illustrating the process from Sample Collection through Next-Generation Sequencing (NGS), the Bioinformatics Pipeline, Integrated Analysis (Genomics and Phylogenetics), Interpretation, and culminating in Public Health Action

The core steps of the bioinformatics workflow, with their associated quality metrics, are detailed in Table 2. This table details the bioinformatics process, focusing on the quality control steps mentioned in the draft.

Table 2 Key Steps and Quality Metrics in an Influenza WGS Bioinformatics Pipeline

Pipeline Step	Description	Key Quality Metric/Output
Raw Data Quality Control	Filtering out low-quality reads, adapter trimming, and removal of host/primer sequences.	Mean Phred Score (Q-score), Percentage of Reads Filtered, Read Length Distribution.
Genome Assembly/Mapping	Assembling reads *de novo* or mapping them to a reference genome (e.g., a closely related strain).	Coverage Depth (X-fold coverage), Breadth of Coverage (Genome % covered), Number of Contigs (for *de novo*).
Consensus Sequence Generation	Deriving the final, single sequence for each of the 8 segments, often involving variant calling and polishing.	Segment Completeness (e.g., 100% of segment length), Ambiguity Codes (e.g., N's), Minor Variant Frequency.

Annotation & Typing	Identifying open reading frames (ORFs), assigning subtype (H/N), and identifying key mutations (e.g., drug resistance).	Subtype (H1, H3, N1, N2 etc.), Presence of Drug Resistance Markers (e.g., S31N in M2), Clade/Lineage Assignment.
Data Submission	Uploading final sequences and metadata to public repositories (e.g., GISAID, NCBI).	Accession Number, Metadata Completeness (Date, Location, Host).

4.1. High-Resolution Variant Calling: Decoding Antigenic Drift

The identification of single nucleotide polymorphisms (SNPs) and small insertions and deletions (indels) is foundational to quantifying viral diversity and the rate of antigenic drift. The challenge is not merely to detect these variations, but to distinguish true biological variation, especially the low-frequency variants that constitute the viral quasispecies, from systematic sequencing artifacts. The ability to accurately profile this intra-host diversity is critical, as a minor variant carrying a drug resistance mutation or an immune-escape epitope can rapidly dominate the population under selective pressure.

The Genome Analysis Toolkit (GATK), initially developed for human genomics, has been widely adapted for viral sequencing data, providing a robust, multi-step workflow.[13] GATK's core philosophy centers on rigorous calibration, such as Base Quality Score Recalibration (BQSR), and probabilistic modeling. For viral genomics, GATK often employs local *de novo* assembly approaches like the HaplotypeCaller to account for the possibility of multiple haplotypes in a population—a feature particularly relevant to the quasispecies nature of RNA viruses. However, GATK's default settings, optimized for diploid organisms, must be carefully tuned to avoid over-filtering true low-frequency variants in a highly diverse viral population.

In contrast, LoFreq is specifically engineered for ultra-sensitive variant detection, a necessity in viral populations where minority variants can pre-exist and rapidly emerge under selective pressure.[14] LoFreq utilizes a unique statistical model that explicitly accounts for sequencing errors, allowing it to detect variants present at frequencies potentially below 1%. This capacity for deep quasispecies analysis provides the high-resolution map necessary to monitor the evolutionary "testing ground" within the host. The output from these variant callers is typically managed and manipulated using tools like BCFtools, which facilitates the filtering, annotation, and visualization of variants stored in the Variant Call Format (VCF), bridging the gap between raw variant calls and biological interpretation.[15] The VCF format, a standardized text file, allows for the seamless integration of variant data into downstream phylogenetic and epidemiological models. The continuous development of these tools reflects the increasing demand for computational methods that can accurately and efficiently navigate the complexity of viral population dynamics.

4.2. Mapping the Topology of Reassortment: Decoding Antigenic Shift

The segmented genome of influenza allows for reassortment, the primary engine of antigenic shift. Detecting these events requires moving beyond simple sequence alignment to analyzing the topological relationships between different genomic segments. A reassortment event is inferred when the phylogenetic tree for one segment exhibits a topology that is significantly incongruent with the topologies of the other segments.

To systematically and robustly identify these topological inconsistencies, specialized computational methods have been developed:

Recombination Detection Program (RDP4): This program implements an extensive array of methods for detecting and visualizing recombination and genomic reassortment.[16] It can differentiate between recombination (within-segment exchange) and segment-level reassortment, providing comprehensive reports on potential events, including statistical support.

GiRaF (Graph-incompatibility-based Reassortment Finder): This tool offers a novel, non-phylogenetic approach by constructing graphs to represent the relationships among different viral strains based on their genetic sequences.[17] By analyzing the graph structure for incompatibilities, GiRaF can robustly identify reassortment patterns, offering a visualization that is particularly insightful for complex evolutionary histories.

Bootscanning: This technique provides statistical support for reassortment by employing a sliding window across the genome. By comparing the phylogenetic relationships across these windows, researchers can detect segments that exhibit significant incongruence, thereby pinpointing the likely location of the segment exchange.

The confluence of these phylogenetic, graph-based, and statistical methods allows for a comprehensive and triangulated analysis of reassortment. However, the reliance on short-read data (e.g., from Illumina platforms) introduces a significant and often overlooked challenge: the inability to definitively link the entire sequence of a segment to a single viral lineage in cases of co-infection or mixed samples. Short reads spanning the segment junctions may not be long enough to unambiguously assign the segment to a specific viral background, leading to potential misclassification or the inability to resolve complex reassortment patterns. This is where the long-read capabilities of platforms like ONT become increasingly critical. The ability to generate a single read that spans the entire length of an influenza segment (up to ~2.3 kb) allows for the unambiguous resolution of these events, directly linking the segment to a specific lineage and thus providing a definitive, high-confidence map of the reassortment event.[10] This technological advance is essential for moving beyond inferential reassortment detection to direct, high-resolution mapping (Figure 1).

4.3. Computational Tracking of Antiviral Resistance

The continuous use of antiviral agents exerts a powerful selective pressure, leading to the emergence of resistant strains. The ability to rapidly and accurately detect these resistance markers is paramount for clinical decision-making and public health policy.

Antiviral resistance in influenza primarily targets two viral proteins: the Neuraminidase (NA) and the M2 ion channel.[2] The key mechanisms and mutations are summarized in Table 3.

Table 3 Key Mechanisms and Mutations in Influenza Antiviral Resistance

Target Protein	Antiviral Drug Class	Mechanism	Key Mutation	Resistance Impact
Neuraminidase (NA)	Neuraminidase Inhibitors (e.g., Oseltamivir, Zanamivir)	Cleaves sialic acid to release new virions. Inhibitors block this release.	H274Y (N1 numbering)	Reduces drug affinity for the active site, conferring resistance to Oseltamivir.[18]
M2 Ion Channel	Adamantanes (e.g., Amantadine, Rimantadine)	Forms a proton channel critical for viral uncoating. Inhibitors block the channel.	S31N	Changes the pore structure, preventing drug binding and conferring resistance to Adamantanes.[19]

NGS-based methods, coupled with computational analysis, offer a superior approach to traditional Sanger sequencing by enabling the simultaneous detection of all potential resistance mutations, including those that are novel or unexpected. Crucially, NGS can detect minority variants – low-frequency resistant strains that may be missed by Sanger sequencing – which are critical indicators of emerging resistance.[18]

The interpretation of this genomic data is heavily reliant on specialized databases and computational resources:

NCBI Influenza Virus Resource: This comprehensive database provides access to a wealth of annotated influenza genomic sequences, essential for comparing newly identified mutations against a large repository of known resistance markers.[20]

Public Drug Resistance Databases: Resources like the Stanford Drug Resistance Database provide curated information on drug resistance across various viruses. A public database is required to represent, store, and analyze the diverse forms of data underlying our knowledge of drug resistance, facilitating comparisons between genotype and phenotype.[21]

4.4. An Integrated Bioinformatics Toolkit for Public Health Genomics

The computational methods discussed – from variant calling to reassortment mapping and resistance screening – form a cohesive toolkit for decoding influenza's genomic plasticity. To provide a practical reference for researchers and public health agencies, Table 4 synthesizes these specialized software tools, databases, and integrated platforms. This curated toolbox underscores the maturity of the field and provides a starting point for laboratories aiming to implement or refine their genomic surveillance pipelines. The seamless operation of these tools, often via cloud-based platforms, is what transforms raw sequencing data into the actionable intelligence we discuss next.

Table 4 A Curated Toolbox for Influenza Genomic Surveillance

Category	Tool/resource	Primary function	Use case in influenza surveillance
Variant Calling	LoFreq	Ultra-sensitive SNV/indel calling	Detecting low-frequency drug-resistance mutations in quasispecies
	GATK	Robust variant discovery, haplotype-aware	Comprehensive variant analysis in diverse viral populations
Reassortment Detection	RDP4	Suite of methods for recombination/reassortment detection	Identifying segment-swapping events between co-circulating strains
	GiRaF	Graph-based reassortment finder	Visualizing and detecting complex reassortment patterns
Phylogenetics	BEAST	Bayesian evolutionary analysis	Estimating evolutionary rates, TMRCA, and phylodynamics
	Nextstrain	Real-time tracking of pathogen evolution	Visualizing global transmission patterns and clade emergence
Integrated Platforms	INSaFLU	Web-based suite for end-to-end WGS analysis	Accessible, standardized pipeline for public health labs
	ViReflow	Scalable, user-friendly viral sequence analysis pipeline	Large-scale genomic epidemiology studies
Databases	GISAID	Global repository for influenza sequences	Primary platform for data sharing and vaccine strain selection
	NCBI Flu Resource	Comprehensive database with analysis tools	Querying sequences and known genetic markers (e.g., resistance)

5. Public Health Action and Future Directions: From Predictive Modeling to Global Health Security

The synergistic integration of genomic, phylogenetic, and computational data has fundamentally reshaped our capacity for influenza surveillance, moving the field toward a more proactive, predictive model rather than a purely reactive one. This sophisticated framework is not merely an academic exercise; it allows for the early detection of variants exhibiting increased transmissibility, altered pathogenicity, or immune escape properties, providing the crucial lead time necessary for optimizing antiviral treatment strategies and implementing targeted public health measures.[9]

5.1. Translating Genomic Data into Actionable Intelligence

The overall process, from sample collection to public health outcome, is best conceptualized as a continuous, feedback-driven loop (Figure 2). Within this loop, the genomic data serves as the engine for predictive modeling. By combining phylogenetic estimates of evolutionary rate with epidemiological data (e.g., incidence, hospitalization rates), researchers can forecast which emerging clades are most likely to dominate future seasons. This capability is paramount for the annual vaccine strain selection process, where a lead time of several months is required to manufacture and distribute vaccines. Genomic surveillance identifies the specific amino acid changes that confer antigenic drift, allowing for a more informed and precise selection of the vaccine candidate, thereby maximizing vaccine effectiveness.[9] Furthermore, the rapid detection of drug-resistant mutations (Section 4.3) enables public health authorities to issue timely clinical alerts and adjust antiviral stockpiling and treatment guidelines, ensuring that the limited arsenal of drugs remains effective.

5.2. The Role of Cloud Computing and Democratization of Analysis

The explosive growth in WGS data volume presents a significant challenge to local computational infrastructure. The future of influenza surveillance is therefore inextricably linked to the increasing accessibility and scalability of cloud-based bioinformatics platforms. These platforms, which offer on-demand computing resources and pre-configured, validated pipelines (such as INSaFLU [11] and ViReflow [12]), effectively democratize WGS analysis. By removing the

need for local high-performance computing clusters and specialized IT expertise, they empower researchers and public health laboratories in resource-limited settings to contribute to and benefit from this integrated surveillance network. This shift is vital for achieving truly global coverage, as influenza evolution is not confined to well-resourced regions. The concept of repurposing and integrating these surveillance platforms has been successfully demonstrated in global health initiatives, showcasing the adaptability of these systems to monitor other emerging pathogens, such as SARS-CoV-2, by leveraging existing influenza infrastructure [22].

5.3. Future Directions: The "One Health" Paradigm and Beyond

Looking forward, the evolution of influenza surveillance will be defined by three key trends:

- Embracing the "One Health" Paradigm: The most significant future direction involves a deeper integration of human, animal (e.g., poultry, swine), and environmental surveillance data. Since the majority of pandemic threats originate from zoonotic spillover, a robust "One Health" approach, leveraging WGS to track interspecies transmission, is crucial.[4] This requires harmonizing sample collection and sequencing protocols across veterinary, agricultural, and public health sectors, a complex logistical and political challenge that genomic data is uniquely positioned to bridge.
- Advanced Computational Phylodynamics: The field is moving beyond simple phylogenetic tree construction to phylodynamics, which integrates evolutionary history with epidemiological processes. Future computational models will incorporate complex factors such as host population structure, spatial movement, and immunity levels to provide highly granular, real-time estimates of transmission rates and viral fitness, offering a more complete picture of epidemic spread than currently possible.[8]
- Real-Time and Decentralized Sequencing: The long-read capabilities of platforms like ONT, particularly their portability, will facilitate decentralized, near-patient sequencing in clinics and remote field sites. This real-time data stream, when coupled with automated cloud-based analysis, promises to reduce the time lag between sample collection and actionable genomic insight from days to mere hours, a critical factor in managing rapidly evolving outbreaks.[10]

As computational virology continues its rapid evolution, the seamless integration of high-throughput genomics and phylogenetics will remain the cornerstone of influenza research, driving innovation and ultimately enhancing global health security. The collaborative effort between molecular virology, bioinformatics, and public health remains the most potent defense against the ceaseless evolutionary ingenuity of the influenza virus.

6. Conclusion

This review underscores that integrating whole-genome sequencing platforms (Illumina and Oxford Nanopore) with robust bioinformatics pipelines — spanning variant calling, reassortment detection, and phylodynamic modeling — enables high-resolution, real-time influenza surveillance far beyond the capacity of traditional methods. Together, these tools decode viral quasispecies dynamics, antigenic evolution, and antiviral resistance, translating raw sequence data into direct public health action including vaccine strain selection and outbreak response. Advancing this integrated framework within a One Health paradigm will be critical for strengthening global preparedness against both seasonal and pandemic influenza, and serves as a replicable model for genomic surveillance of future emerging infectious diseases.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Van Poelvoorde L, Quoilin S, De Keersmaecker SCJ, Devleesschauwer B, et al. Whole-Genome Sequence Approach and Phylogenomic Stratification Improve the Association Analysis of Mutations With Patient Data in Influenza Surveillance. Front Public Health. 2022 May 2;10:889987. Available from: https://doi.org/10.3389/fpubh.2022.809887

[2] Mtambo SE, Mulaudzi T, Shonhai A. Influenza Viruses: Harnessing the Crucial Role of the M2 Ion-Channel and Neuraminidase toward Inhibitor Design. Molecules. 2021 Feb 4;26(4):880. Available from: https://doi.org/10.3390/molecules26040880

[3] Iuliano AD, Roguski SK, Chang HH, Muscatello DJ, Palekar R, et al. Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. Lancet. 2018 Mar 31;391(10127):1285-1302. Available from: https://doi.org/10.1016/S0140-6736(17)33293-2

[4] Smith GJD, Vijaykrishna J, Bahl J, Lycett SJ, Lewis NS, et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature. 2009 Jun 25;459(7250):1122-5. Available from: https://doi.org/10.1038/nature08182

[5] Skowronski DM, Chambers C, Sabaiduc S, De Serres G, et al. Integrated sentinel surveillance linking genetic, antigenic, and epidemiologic monitoring of influenza vaccine-virus relatedness and effectiveness during the 2013-2014 season. J Infect Dis. 2015 Sep 1;212(5):726-36. Available from: https://doi.org/10.1093/infdis/jiu644

[6] Van den Hoecke S, Verhelst J, Vuylsteke M, Saelens X. Analysis of the genetic diversity of influenza A viruses using next-generation DNA sequencing. BMC Genomics. 2015 Feb 26;16:150. Available from: https://doi.org/10.1186/s12864-015-1284-z

[7] McGinnis J, Laplante J, Shudt M, George KS. Next generation sequencing for whole genome analysis and surveillance of influenza A viruses. J Clin Virol. 2016 Mar;76 Suppl 1:S13-S17. Available from: https://doi.org/10.1016/j.jcv.2016.01.011

[8] Gill MS, Pepin J, Labbé AC, Ndiaye Y, et al. Online Bayesian phylodynamic inference in BEAST with application to epidemic reconstruction. Mol Biol Evol. 2020 Jun 1;37(6):1832-1836. Available from: https://doi.org/10.1093/molbev/msaa040

[9] Jayadas TTP, Jeewandara C, Senadheera B, et al. Genomic surveillance and evolutionary dynamics of influenza A virus in Sri Lanka. Virol J. 2024 Apr 24;21(1):97. Available from: https://doi.org/10.1186/s12985-024-02555-z

[10] Williams TGS, Snell LB, Alder C, et al. Feasibility and clinical utility of local rapid Nanopore influenza A virus whole genome sequencing for integrated outbreak management, genotypic resistance detection, and timely surveillance. Microb Genom. 2023 Sep;9(9):001083. Available from: https://doi.org/10.1099/mgen.0.001083

[11] Borges V, Pinto M, Osório H, et al. INSaFLU: an automated open web-based bioinformatics suite “from-reads” for influenza whole-genome-sequencing-based surveillance. Genome Med. 2018 Jun 28;10(1):50. Available from: https://doi.org/10.1186/s13073-018-0555-0

[12] Moshiri N, Moshiri B, Moshiri N. The ViReflow pipeline enables user friendly large scale viral sequence analysis. Sci Rep. 2022 Mar 24;12(1):4999. Available from: https://doi.org/10.1038/s41598-022-09035-w

[13] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep;20(9):1297-303. Available from: https://doi.org/10.1101/gr.107524.110

[14] Wilm A, Aw PPK, Bertrand D, Yeo GHT, Ong SH, Wong CH, Khor CC, Petric R, Hibberd ML, Nagarajan N. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. 2012 Dec;40(22):11189-201. Available from: https://doi.org/10.1093/nar/gks918

[15] Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Kokocinski T, Li H, Birol G, Cox AJ, Grocock AJ, Heger A, O'Keeffe S, Stalker J, Woodcock C, Smith GR, Tassé AM, Keane TM, Flicek P, Li H, Durbin R. Twelve years of SAMtools and BCFtools. Gigascience. 2021 Mar 4;10(2):giab008. Available from: https://doi.org/10.1093/gigascience/giab008

[16] Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol. 2015 Mar 1;1(1):vev003. Available from: https://doi.org/10.1093/ve/vev003

[17] Nagarajan N, Kingsford C. GiRaF: robust, computational identification of influenza reassortments via graph mining. Nucleic Acids Res. 2011 Mar;39(6):e34. Available from: https://doi.org/10.1093/nar/gkq1198

[18] Hurt AC, Holien JK, Parker MW, Barr IG. Oseltamivir resistance and the H274Y neuraminidase mutation in seasonal, pandemic and highly pathogenic influenza viruses. Drugs. 2009;69(18):2523-31. Available from: https://doi.org/10.2165/11531450-000000000-00000

[19] Cheng PKC, To APC, Leung TWC, Li KSM, Lee WM, Lim W. Oseltamivir- and amantadine-resistant influenza virus A (H1N1). Emerg Infect Dis. 2010 May;16(5):856-8. Available from: https://doi.org/10.3201/eid1605.091763

[20] Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Zaslavsky L, Tatusova T, Ostell J, Lipman D. The Influenza Virus Resource at the National Center for Biotechnology Information. J Virol. 2008 Feb;82(2):596-601. Available from: https://doi.org/10.1128/JVI.02005-07

[21] Shafer RW. Rationale and uses of a public viral drug resistance database. J Infect Dis. 2006 May 15;193(10):1346-8. Available from: https://doi.org/10.1086/503419

[22] Asante IA, Hsu SN, Boatemaa L, Kwasah L, et al. Repurposing an integrated national influenza platform for genomic surveillance of SARS-CoV-2 in Ghana: a molecular epidemiological analysis. Lancet Glob Health. 2023 Oct;11(10):e1553-e1562. Available from: https://doi.org/10.1016/S2214-109X(23)00189-4