CONTIG · VOL. 1 · ISSUE 202603 · 2026 · REVIEW · OPEN ACCESS · CC BY 4.0 · UDC: 579.25, 579.26

Genetic and Genomic Perspectives on Emerging and Re-Emerging Plant Diseases

A review of genomic surveillance, HTS diagnostics, pathogen pangenomes, and genomic risk forecasting.

Authors

- Dilyara Gritsenko, PhD, Assoc. Prof., head of laboratory, laboratory of molecular biology, Institute of plant biology and biotechnology, ORCID, Scopus
- Alexandr Pozharskiy, PhD, senior researcher, laboratory of molecular biology, Institute of plant biology and biotechnology, ORCID, Scopus

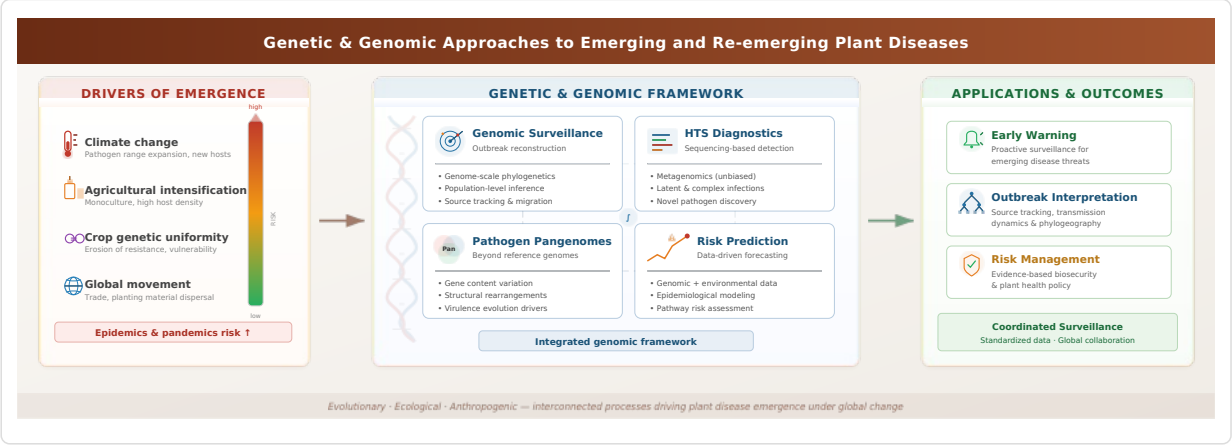
Institute of Plant Biology and Biotechnology, Kazakhstan

DOI	10.66273/3134-6359.2026.1.1.003
Dates	Received: 10.01.2026 · Accepted: 31.01.2026 · Published online: 31.01.2026
Edited & reviewed by	Ruslan Kalendar
Citation	Dilyara Gritsenko, Alexandr Pozharskiy (2026) Genetic and Genomic Perspectives on Emerging and Re-Emerging Plant Diseases. Contig. 1:202603. DOI: 10.66273/3134-6359.2026.1.1.003
Keywords	emerging plant diseases; genomic surveillance; high-throughput sequencing; metagenomics; pathogen pangenome; effector-mediated virulence; risk assessment

Abstract

Emerging and re-emerging plant diseases are becoming more frequent and severe due to climate change, agricultural intensification, crop genetic uniformity, and the expanding global movement of planting materials. These drivers have altered host-pathogen-environment interactions, facilitating pathogen range expansion, host shifts, and erosion of resistance, thereby increasing the risk of epidemics and pandemics in major crops. In this context, genetic and genomic approaches have become central to modern plant pathology, providing high-resolution tools to investigate disease emergence, spread, and adaptation. This review summarizes recent advances in genomic surveillance, high-throughput sequencing-based diagnostics, pathogen pangenomes, and data-driven prediction as interconnected components of an integrated framework for studying emerging and re-emerging plant diseases. Genome-scale analyses have improved outbreak reconstruction and population-level inference, while unbiased sequencing and metagenomic approaches have expanded diagnostic capacity to include complex, latent, and previously undetected infections. Pangenomic studies further demonstrate that virulence evolution is frequently driven by gene content variation, structural rearrangements, and demographic processes, challenging reference-centric models of pathogen adaptation. Predictive and risk assessment approaches increasingly integrate genomic information with environmental, epidemiological, and pathway data to support proactive surveillance and management. Taken together, the evidence indicates that plant disease emergence results from interacting evolutionary, ecological, and anthropogenic processes. Genetic and genomic tools, when embedded within coordinated surveillance systems and supported by standardized data practices, provide a robust foundation for early warning, improved outbreak interpretation, and risk-informed plant health management under accelerating global change.

Graphical Abstract



Introduction

Emerging and re-emerging plant diseases are increasing in frequency and severity. This increase is driven by climate change, intensified agriculture, larger monocultures, and the unprecedented connectivity in global trade and movement of planting materials (Fisher et al. 2012; D.P. Bebber et al. 2014; Ristaino et al. 2021; Raza and Bebber 2022). These factors alter host–pathogen interactions and environmental dynamics, allowing pathogens to extend their geographic ranges, explore new host species, and overcome existing resistances (Ristaino et al. 2021; Raza and Bebber 2022). As a result, the likelihood of epidemics and disease pandemics in key crops rises. Analyzing historical and recent outbreak strains through genomic mining, combined with enhanced surveillance, is crucial for developing early warning systems and understanding how pathogens are introduced and spread (Ristaino et al. 2021).

Over the last ten years, advances in genetic and genomic technologies have revolutionized the detection, characterization, and management of threats to plant health. Whole-genome sequencing (WGS) and population genomics now allow for strain-level identification of bacteria, fungi, oomycetes, and viruses (Grünwald et al. 2016; Everhart et al. 2021). These methods enable detailed insights into outbreak origins, transmission routes, and evolutionary patterns when combined with spatial–temporal data. This “genomic epidemiology” method provides practical guidance for quarantine, eradication, and management by distinguishing between single and multiple introductions, identifying reservoirs, and tracking the emergence of adaptive trait variants (Grünwald et al. 2016).

In parallel, high-throughput sequencing (HTS) has become a cornerstone for unbiased pathogen detection and discovery (Massart et al. 2014; Roossinck et al. 2015; Villamor et al. 2019). Metagenomic and viromic methods are especially useful for diagnosing complex infections, uncovering hidden co-infections, and detecting new or unexpected viruses that targeted tests might miss. These techniques are growing in importance for clonally propagated crops and vegetatively spread planting materials, where latent infections can remain undetected and cause re-infection across regions and seasons (Boykin et al. 2019).

A significant conceptual change has happened in pathogenomics: relying on a single reference genome is often insufficient to capture the full spectrum of variation important for virulence and host interactions adaptation (Möller and Stukenbrock 2017; Badet and Croll 2020). Pangenome and graph-based frameworks identify gene presence–absence variations, structural variants, and copy-number changes, which are crucial for effector repertoires and the swift loss of resistance in the field (Plissonneau et al. 2018). Recent research demonstrates that pangenome-level analyses can uncover significant effector copy-number variation and other forms of genomic plasticity that are often undetected by reference-centric methods. This enhances our capacity to predict and track high-risk pathogens' lineages (Plissonneau et al. 2018; Badet and Croll 2020).

On the host side, advances in genomics, including genome-wide association studies (GWAS), genomic

selection, and pangenome-informed analysis of resistance loci, are enhancing the identification and application of durable resistance (Montesinos-López et al. 2021; Upadhyaya et al. 2024). The integration of machine learning with genomic datasets further supports resistance prediction and the prioritization of loci for breeding, offering a path to faster and more precise responses to evolving pathogen populations (Sperschneider 2020; Upadhyaya et al. 2024).

Thus, genetic and genomic approaches enable comprehensive analysis of emerging and re-emerging plant diseases. In this review, the following topics will be discussed: (i) genomic surveillance for plant disease outbreak reconstruction, (ii) HTS-based diagnostics and metagenomics, (iii) pathogen pangenomics and effector-mediated virulence, and (iv) prediction and risk assessment of emerging and re-emerging plant diseases.

Genomic surveillance for outbreak reconstruction

Genomic surveillance now plays a key role in modern plant pathology, offering high resolution for tracing disease outbreaks and studying how plant pathogens emerge and re-emerge. Unlike traditional methods that depend on phenotypic traits or limited genetic markers, whole-genome sequencing (WGS) allows researchers to determine pathogen origins, migration paths, population structures, and evolutionary patterns at the strain level. This progress has been fueled by advancements in sequencing technology, population genomics analysis, and increased access to existing collections of historical and current pathogen samples .

McDonald and Stukenbrock (2016) provided a comprehensive evolutionary framework for understanding the emergence and re-emergence of plant diseases. They explained that pathogen emergence arises from interconnected processes occurring at various spatial and temporal scales, including long-distance migration, recombination, natural selection, and host adaptation (McDonald and Stukenbrock 2016). By combining population genetics with epidemiological principles, this study demonstrated how genome-scale data can trace outbreak histories and distinguish between local evolutionary changes and multiple separate introductions. This differentiation is essential for accurate outbreak reconstruction and the development of effective quarantine, surveillance, and disease control measures.

The significance of global genomic surveillance in tracking plant disease outbreaks is evident, as research indicates that current epidemic patterns are increasingly influenced by global connectivity, climate change, and intensified farming practices (Ristaino et al. 2021). These factors collectively increase the frequency, scale, and complexity of outbreaks, emphasizing the importance of coordinated genomic monitoring for early detection and rapid response. While genome-scale analyses provide detailed insights into pathogen dissemination and entry paths, they also show that genomic data alone are insufficient for precise outbreak reconstruction. Accurate conclusions depend on integrating genomic information with spatial, temporal, host, and trade-related data. The absence of comprehensive datasets acts as a major obstacle to effective surveillance and phytosanitary decision-making.

Numerous landmark studies demonstrate the effectiveness of genomic surveillance in reconstructing both historical and recent outbreaks. A notable example is Yoshida et al.'s work, which used historical herbarium specimens to reconstruct the pathogen behind the Irish potato famine, Phytophthora infestans (Yoshida et al. 2013). This study demonstrated that the extinct lineage HERB-1 was responsible for 19th-century pandemics and showed how temporal genomic data can uncover lineage replacements that might otherwise go unnoticed. It set a benchmark for “historical pathogenomics" and emphasized technical challenges such as DNA degradation, limited sample access, and the importance of strict ancient DNA authentication.

Genomic surveillance has demonstrated its effectiveness in reconstructing ongoing epidemics. Hubbard et al. coined the term “field pathogenomics” by employing large-scale sequencing to monitor the emergence and dissemination of wheat yellow rust (Puccinia striiformis) (Hubbard et al. 2015). This

approach enabled near real-time insights into population turnover and the appearance of aggressive genotypes. However, subsequent analyses have shown that such reconstructions are highly sensitive to sampling design; uneven geographic coverage can bias phylogeographic inference and lead to incorrect conclusions about the direction and timing of pathogen spread.

Although genomic surveillance has great potential for transformation, it encounters several key limitations that should be clearly addressed when considering emerging and re-emerging plant diseases. First, sampling bias is widespread; genomic data often favors regions with robust research infrastructure while neglecting outbreak areas. Second, incomplete or inconsistent metadata can hinder epidemiological analysis, turning genomic data into mere descriptive phylogenies rather than useful reconstruction tools. Third, biological complexities such as recombination, hybridization, and varying clonality statuses make standard phylogenetic analysis challenging and necessitate pathogen-specific approaches.

Recent studies are increasingly highlighting the importance of shifting from single-reference methods to surveillance frameworks that consider populations and pangenomes. Although these approaches are still developing, they are especially important for pathogens with high variability in their accessory genomes, where critical virulence factors might not be present in standard reference genomes and could therefore be missed during outbreak investigations.

HTS-based diagnostics and metagenomics

High-throughput sequencing (HTS), also known as next-generation sequencing (NGS), has dramatically changed the landscape of plant pathology diagnostics. It enables broad, unbiased detection of multiple pathogens from a single sample, without prior knowledge of the organisms. Traditional methods such as PCR, ELISA, and culture-based tests typically detect only known pathogens for which specific probes, antibodies, or primers are available. In contrast, HTS offers high sensitivity and coverage by producing millions of sequence reads, which can be aligned to reference databases or assembled de novo to identify both known and novel infectious agents (Villamor et al. 2019).

A primary application of HTS in plant diagnostics is to identify and characterize plant viruses and viroids. These pathogens are highly diverse, often asymptomatic, and challenging to detect with traditional methods. HTS enables the detection of entire viral genomes, low-abundance pathogens, and mixed infections, providing a far more detailed overview of viral diversity within plant tissues and associated environments. This method significantly outperforms targeted approaches in both sensitivity and comprehensiveness (Villamor et al. 2019). Performing HTS on total RNA (shotgun RNA-seq) or small interfering RNA fractions enables virus detection without prior knowledge of their identity. This method facilitates diagnosing known viruses and discovering new viruses and strains that may contribute to disease causes (Jaksa-Czotter et al. 2024).

HTS-based metagenomic methods have been effectively used beyond viruses, extending to bacterial, fungal, oomycete, and mixed-pathogen communities. By sequencing all nucleic acids from both symptomatic and asymptomatic plant samples, metagenomics can identify the presence, relative levels, and genomic traits of various taxa, including those hard to culture or target with specific tests. This “shotgun” approach is particularly effective for complex diseases (Nizamani et al. 2023). A key advantage of HTS diagnostics is its ability to analyze the plant virome and microbiome at scale, offering valuable insights into the ecological context of disease. Studies using HTS have shown that plant samples often harbor a rich assemblage of viral and microbial sequences, including latent or cryptic pathogens that targeted testing would miss. By integrating HTS data with bioinformatic classification pipelines and expanding virus reference databases, researchers have uncovered previously undocumented viruses and sequence variants in crops and their wild relatives, thereby enhancing understanding of pathogen diversity and evolutionary dynamics (Nizamani et al. 2023).

While HTS provides powerful diagnostic capabilities, several challenges limit its routine application in

plant health monitoring. First, handling, assembling, and analyzing large volumes of sequence data require specialized bioinformatics skills and validated pipelines, which many diagnostic labs do not possess. Furthermore, precise taxonomic classification depends on comprehensive, well-annotated reference databases. However, many pathogen groups, particularly viruses, are underrepresented compared to their estimated diversity (Maina et al. 2024). Second, identifying resident, commensal, and pathogenic sequences in metagenomic datasets remains challenging without a robust epidemiological context and standardized sampling methods. Lastly, although the costs and infrastructure needs for HTS are decreasing, they remain significant compared to traditional diagnostics, which can hinder adoption in resource-limited environments (Maina et al. 2024).

HTS allows for the simultaneous detection, discovery, and genomic analysis of a broad range of plant pathogens from single tests, placing it in a distinctive position within modern diagnostics. The sequence data obtained through HTS can be used for further analyses such as phylogenetic studies, comparative genomics, and resistance marker identification, which aid in reconstructing outbreaks, assessing risks, and shaping response plans. Combining metagenomics with powerful bioinformatics tools has thus become crucial for next-generation plant disease diagnostics, especially for emerging and re-emerging pathogens that are difficult to detect with targeted methods (Kutnjak et al. 2021). Pathogen pangenomics and effector-mediated virulence Pathogen pangenomics has fundamentally transformed the interpretation of virulence in emerging and re-emerging plant diseases by illustrating that a single reference genome often does not encompass the full extent of genetic variation pertinent to host adaptation. Comparative analyses across multiple pathogen genomes have revealed that crucial virulence determinants, notably effector genes, are frequently situated within the accessory genome, where gene presence–absence variability, structural rearrangements, and transposable element activity are prevalent. These genomic compartments demonstrate a higher rate of evolutionary turnover compared to the conserved core genome, facilitating the rapid acquisition and loss of virulence-associated genes. The methodological frameworks established for the construction and analysis of pathogen pangenomes have further demonstrated that the distribution patterns of accessory genes are markedly affected by ecological and evolutionary factors such as host range, geographic origin, and population structure. Consequently, this provides insight into rapid adaptation and resistance breakdown (Badet and Croll 2020). A key strength of pangenomic approaches is their ability to reveal presence-absence variation (PAV) and structural variation (SV) that underpin virulence and resistance breakdown (Table 1). A recent pangenome study of the wheat pathogen Zymoseptoria tritici assembled multiple high-quality genomes and demonstrated that genome plasticity, including rearrangements, gene deletions, and variable gene content, is central to host specialization and adaptation, moving the field beyond SNP-only narratives (Lovelace et al. 2023). Additionally, a large, reference-quality pangenome resource was created that enabled more systematic mapping of gene content diversity and provided a platform to ask how pangenome structure affects host–pathogen interactions at the population scale (Badet et al. 2020). Together, these studies established an important practical lesson for outbreak-relevant research: virulence-related diversity is often driven by gene content and SVs, so surveillance that relies exclusively on read mapping to a single reference can miss lineage-defining virulence determinants. Effector-mediated virulence is particularly well served by pangenomics because effector repertoires are often shaped by duplications, deletions, and copy-number variation (CNV) in dynamic genomic regions. Research on transposon-driven genome evolution in Verticillium dahliae revealed that transposable elements can promote adaptive evolution through genome rearrangements and duplications, mainly in lineage-specific regions (Faino et al. 2016). These regions are exactly the type of genomic architecture that can hasten alterations in effector complements and other virulence traits. Recent research has critically analyzed the conceptual frameworks of 'two-speed' genome evolution. These studies support the observed connection between dynamic, repeat-rich regions and effector-like genes. However, they also caution against universally applying this model across all pathogen groups (Torres et al. 2020). The consensus is nuanced: while “fast” compartments can enable quick adaptation, the mechanisms and predictability of effector evolution vary among taxa, which is important when applying genomics to risk forecasting. Methodologically, the field has advanced from

multi-reference assemblies to pangenome graphs, which better capture SV and CNV than linear references. A recent pangenome graph analysis of an oomycete plant pathogen (spinach downy mildew) showed that virulence-related genes can cluster in large CNV-rich regions and that graph-based discovery can reveal variation at exceptional resolution (Skiadas et al. 2024). A key limitation in effector research is the common disconnect between effector prediction and actual effector function. Systematic evaluations of effector discovery methods reveal that silico predictions based on signal peptides, secretion motifs, subcellular localization, structural features, comparative genomics, and infection-stage expression identify only a portion of biologically active effectors. Additionally, delivery mechanisms vary greatly across different pathogen groups (Lovelace et al. 2023). Effector repertoires cannot be reliably determined using just a single generic pipeline; they need pathogen-specific optimization and validation. Meanwhile, comparative pangenome analyses show that many studies still depend heavily on effector-like annotations without functional proof, which results in inflated candidate sets and unclear connections to virulence. A more comprehensive analytical framework has developed, triangulating multiple evidence sources: gene presence-absence variation or copy-number changes detected through pangenome analysis, transcriptional support during infection, and genetic or molecular validation methods such as gene knockouts, allele replacement, or host recognition assays. This integrative approach is essential for linking genomic variation to phenotypic outcomes and for correctly interpreting resistance breakdown, because changes in effector copy number or accessory gene content are biologically meaningful only when they translate into altered infection dynamics or host responses. Pangenomics is increasingly used to link virulence evolution to recent outbreaks. For instance, population-genomic analyses of wheat blast reveal how different mating events and gene flow can generate new pathogenic groups, suggesting that effector repertoires and adaptability may be assembled through demographic history rather than solely through gradual local mutations. Table 1 Representative plant pathogens analyzed using pangenome approaches and key biological insights. CNV - Copy Number Variation; HGT - Horizontal Gene Transfer; RIP - Repeat-Induced Point mutation; SNP - Single Nucleotide Polymorphism Prediction and risk assessment Data-driven prediction and risk assessment increasingly depend on integrating diverse evidence streams, including pathogen genomic data, high-throughput diagnostics, climate and weather variables, host distributions, and trade and movement data. This integration helps to forecast where and when emerging or re-emerging disease events are most likely, guiding surveillance and intervention efforts. Global analyses have quantified crop losses from pathogens and pests, establishing a baseline for risk prioritization and highlighting the disproportionate burden on key food crops (Savary et al. 2019). Similarly, macroecological analyses reveal the continuous worldwide spread of crop pests and pathogens. They associate invasion patterns with environmental and socioeconomic factors, underscoring the importance of predictive, spatial risk mapping over reactive approaches (D. Bebber et al. 2014). Forecasting and risk assessment for plant diseases encounter unique challenges that impact model accuracy and transferability. These include issues like imperfect detection, significant spatial heterogeneity, reliance on host phenology and management, changing drivers due to climate change, and structural uncertainty when extrapolating beyond observed conditions. If these factors are not explicitly addressed, they can lead to overly confident risk predictions (Cunniffe et al. 2015). These limitations are particularly significant for emerging and re-emerging threats, where surveillance data are limited, emergence events might be infrequent but impactful, and pathway data are incomplete. Genomic surveillance frameworks indicate that outbreak analysis and risk assessment are more accurate when genomic signals such as evidence of multiple emergence events, rapid population growth, or new virulence profiles are combined with detailed spatial–temporal and pathway metadata. Without this integration, both reconstructing outbreaks and predicting their course become considerably more challenging (Ristaino et al. 2021). Overall, these studies support creating iterative, updateable risk models that assess uncertainty, address sampling bias and missing metadata, and integrate predictive analytics with surveillance planning. This strategy guarantees that data-driven outcomes remain effective for quarantine prioritization, targeted monitoring, and rapid deployment of control measures. Conclusions Emerging and re-emerging plant diseases pose an escalating threat to global agricultural production, driven by

climate change, intensified cropping systems, the genetic uniformity of cultivated plants, and increased connectivity through international trade and the movement of planting materials. As demonstrated throughout this review, genetic and genomic approaches have become central to understanding these processes, enabling a shift from predominantly reactive disease management toward proactive surveillance, analysis, and risk-informed intervention. Genomic surveillance has substantially improved the reconstruction of plant disease outbreaks by providing strain-level resolution of pathogen populations and insights into origins, transmission pathways, and evolutionary dynamics. However, the effectiveness of such approaches depends critically on representative sampling, standardized metadata, and analytical frameworks that account for population structure, recombination, and demographic history. Without integration of genomic data with spatial, temporal, host, and pathway information, outbreak reconstruction and epidemiological inference remain incomplete. In summary, genetic and genomic approaches now underpin modern strategies for understanding and managing emerging and re-emerging plant diseases. Future progress will require closer integration among genomic surveillance, diagnostics, pangenomics, and predictive modeling, supported by standardized data practices, international collaboration, and sustained investment in genomic infrastructure. These integrated frameworks are essential for translating genomic knowledge into effective disease prevention, containment, and long-term resilience of agricultural systems. Funding This research was funded by the Ministry of Agriculture of the Republic of Kazakhstan, BR22887230 “Development of an effective population management system for quarantine pests, with limited distribution in the Republic of Kazakhstan” Author contributions The manuscript was written through the contributions of all authors. All authors have given approval to the final version of the manuscript. Competing Interests The authors declare no competing interests. References Keywords emerging plant diseases; genomic surveillance; high-throughput sequencing; metagenomics; pathogen pangenome; effector-mediated virulence; risk assessment Citation Dilyara Gritsenko, Alexandr Pozharskiy (2026) Genetic and Genomic Perspectives on Emerging and Re-Emerging Plant Diseases. Contig. 1:202603. DOI: to be assigned Dates Received: 10.01.2026 · Accepted: 31.01.2026 · Published online: 31.01.2026 Edited and reviewed by Correspondence

---