VETERINARIA
ITALIANA

RIVISTA DI
SANITÀ PUBBLICA
VETERINARIA

SPREAD: Spatiotemporal Pathogen Relationships and Epidemiological Analysis Dashboard

Andrea de Ruvo[1*], Alessandro De Luca[2], Andrea Bucciacchio[2], Pierluigi Castelli[2], Alessio Di Lorenzo[2], Nicolas Radomski[2], Adriano Di Pasquale[2]

[1]Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "G. Caporale", Teramo, Italy; Computer Science, Gran Sasso Science Institute, L'Aquila, Italy - IT
[2]Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "G. Caporale" - IT

[*]Corresponding author at: Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "G. Caporale", Teramo, Italy; Computer Science, Gran Sasso Science Institute, L'Aquila, Italy - IT
E-mail: a.deruvo@izs.it

Abstract

In the scope of public health, the rapid identification and control of infectious disease outbreaks are a paramount concern. Traditional surveillance methods often face challenges in effectively combining genetic, geographical, and temporal data, which is crucial for a comprehensive understanding of disease transmission dynamics. Addressing this critical need, the Spatiotemporal Phylogenomic Research and Epidemiological Analysis Dashboard (SPREAD) emerges as an innovative standalone web-based application. SPREAD integrates several modules for detailed genomic relationships, pinpointing genetically close pathogens, and spatial mapping, providing in-depth views of how diseases spread across populations and territories, with significant advantage to manage both bacteria and viruses based on allele and variant calling, respectively. Designed for broad accessibility, SPREAD operates seamlessly within web browsers, eliminating the need for sophisticated IT infrastructure and facilitating its use across various public health contexts. Its intuitive interface ensures that users can effortlessly navigate complex datasets, facilitating widespread access to advanced surveillance capabilities. Through its initial deployments, SPREAD has proven instrumental in quickly identifying transmission clusters, significantly aiding in the formulation of prompt and targeted public health responses. Through the integration of state-of-the-art technology with a focus on user-centered design, SPREAD offers a promising solution that highlights the potential of digital health innovations.

Keywords

Infectious Disease Surveillance, Spatiotemporal Visualization, Standalone Dashboard, Phylogenomic Analysis, Public Health Informatics

Introduction

In the domain of global health, a profound comprehension of the transmission pathways of infectious diseases is imperative for enabling effective surveillance and swift outbreak response (Nsoesie et al., 2015). The intricate nature of tracking and interpreting the spread of these diseases underscores the necessity for more sophisticated and adaptable national surveillance systems based on genomic and metagenomic data (Akther et al., 2020). The ability to effectively track and understand the transmission pathways of pathogens is crucial for public health decision-making (Grubaugh et al., 2018) and retrospective research.The current constraints of these genomic data-based national surveillance systems are directly associated with the main goals of users in charge of surveillance activities (Perry et al., 2007). Firstly, these users need to integrate spatiotemporal data with genomic data to detect genetically related clusters of pathogens in real-time (e.g., bacteria and viruses) (Payne et al., 2024). Since the typical static genomic relationships from cumulative long-term genomic surveillance are usually too large to be presented in a human-readable format, it is also necessary to be able to easily navigate comprehensively through genetically related clusters of pathogens (Letunic and Bork, 2021). Finally, the specialists of national surveillance activities deal with a vast amount of sensitive spatiotemporal genomic data. Their tendency to manage such data internally, within their own self-

supported computing facilities (Pronyk et al., 2021), is not merely a matter of choice but is also imposed by legislation, ensuring stringent adherence to data privacy principles (Arshad et al., 2021). This legal mandate underscores the critical importance of safeguarding sensitive information, aligning operational practices with legal and ethical standards (Kędzior, 2020).Addressing some of these concerns raised by experts in the field of genomic-based national surveillance activities, the objectives of the present study were to (i) link spatiotemporal data to genomic relationships of pathogens' spreading, (ii) provide a solution to enable in-depth navigation among genetically close pathogens, and (iii) ensure data privacy through a versatile, open-source, standalone dashboard capable of handling data from various pathogens. The present study introduces an innovative standalone dashboard named Spatiotemporal Pathogen Relationships and Epidemiological Analysis Dashboard (SPREAD), integrating the functionalities of GrapeTree (Zhou et al., 2018) and ReportTree (Mixão et al., 2023), alongside a portable data visualization tool, to significantly improve the understanding of disease transmission pathways.

Materials and methods

Central to SPREAD is the integration of Leaflet (version 1.7.1), a JavaScript library designed for interactive mapping, with the OpenStreetMap free geographic database (implemented as a service). Together, they form a comprehensive Geographic Information System (GIS) (Herfort et al., 2021; Edler and Vetter, 2019), enhancing the platform's ability to visualize and analyze spatial data effectively.As a proof of concept, SPREAD was tested using a large genomic data collection of four microorganisms: *Listeria monocytogenes*, Severe Acute Respiratory Syndrome-related virus 2 (SARS-CoV-2), African Swine Fever Virus (ASFV) and West Nile Virus (WNV).

Sample collection

The genomic samples used in this proof of concept were collected by IZSAM via the GENPAT information system, a national platform based on a fork of the COHESIVE information system (Mangone et al., 2021), within the framework of both national and European surveillance initiatives (Table 1). Additionally, the NGSManager bioinformatic pipeline (https://github.com/genpat-it/ngsmanager) of the GENPAT platform (https://genpat.izs.it/) was used to process paired-end reads and/or long reads. More precisely, bacterial genomic samples of *L. monocytogenes* underwent processing through a *de novo* assembly-based cgMLST workflow, developed for the *Listeria* National Reference Laboratory. This workflow involved the use of fastp (version 0.23.1) (Chen, 2023), shovill (version 1.1.0) (https://github.com/tseemann/shovill) and chewBBACA (version 2.8.5) (Silva et al., 2018), culminating in the generation of allelic profiles combined into a single allele matrix file. Concerning viruses, genomic samples of SARS-CoV-2, ASFV, and WNV underwent variant calling workflows, implementing Trimmomatic (version 0.36) (https://github.com/usadellab/Trimmomatic), snippy (version 4.5.1) (https://github.com/tseemann/snippy) and vcf2mlst (Di Pasquale et al., 2021). The used reference genomes were Wuhan-Hu-1/2019, Georgia 2007/1, and NY99/Héja for SARS-CoV-2, ASFV, and WNV, respectively. The cgMLST-like outputs from both bacterial and viral workflows were also compatible with GrapeTree (version 1.5.0) (Zhou et al., 2018) and ReportTree (version 2.0.3) (Mixão et al., 2023). In summary, the developed standalone dashboard underwent testing using a substantial dataset, comprising 8832*L. monocytogenes*, 9216 SARS-CoV-2, 158 ASFV, and 335 WNV samples (Table 1).

Pathogen	Number of samples	Study Period (start year-end year)	Bioinformatics workflows (NGSManager)
SARS-CoV-2	9202	2020-2024	Variant calling
Listeria monocytogenes	8832	1999-2024	*De novo* assembly based cgMLST
ASFV	158	2015-2023	Variant calling
WNV	335	2017-2023	Variant calling

Table 1. Bacterial and viral genomic samples, as well as related preprocessings, used to prepare input data for SPREAD.

Main dependencies

The GrapeTree (Zhou et al., 2018), Leaflet-powered OpenStreetMap (Herfort et al., 2021; Edler and Vetter, 2019) and ReportTree (Mixão et al., 2023) resources were selected to manage the genomic relationships, GIS, and detection of genetically close samples, respectively. The resources were integrated into SPREAD for their unique contributions in managing intricate and heterogeneous datasets, typical of infectious disease surveillance. GrapeTree (Zhou et al., 2018) is a fully interactive tree visualization program originally developed for EnteroBase (Zhou et al., 2019), a database for genomic data of enteric pathogens. Notably, it stands out for its ability to handle large numbers of allelic profiles, supporting interactive visualizations of extensive Minimum Spanning Trees (MSTs). GrapeTree (Zhou et al., 2018) allows users to manipulate both tree layout and associated metadata, making it an invaluable tool for exploring core genomic relationships among bacterial pathogens. Its functionality extends to investigating genomic relationships, enabling a comprehensive understanding of genomic diversity and relationships. The OpenStreetMap free world map (Herfort et al., 2021) and the Leaflet open-source JavaScript library for interactive maps (Edler and Vetter, 2019) were selected because they are known for their flexibility in creating web mapping applications. This combination offers a solid platform for visualizing and interacting with geographical data. ReportTree (Mixão et al., 2023) is a recent surveillance-oriented tool, designed to strengthen the linkage between pathogen genetic clusters and epidemiological data. This software facilitates the analysis of genetic sequencing data, allowing researchers and public health professionals to trace potential epidemiologically related clusters of sequenced genomes, based on their genomic proximity. ReportTree (Mixão et al., 2023) automates the identification and characterization of pathogen clusters, providing rapid, flexible, and reproducible insights that are essential in today's fast-paced epidemiological landscape.

Minimal input data requirements

The effectiveness of SPREAD in infectious disease research relies on its ability to handle various data formats, a crucial feature, considering the diverse data structures involved in genomic and epidemiological studies. SPREAD processes Newick (NWK) format for phylogenetic trees, essential for understanding genomic diversity and relationships among pathogens. While the NWK file alone forms the minimal input for initiating an analysis, it does not provide sufficient any spatiotemporal insights, highlighting the importance of integrating additional data types. To this end, SPREAD utilizes Tab-Separated Values (TSV) to incorporate indispensable metadata, including sample identifiers, collection locations, and dates, thereby embedding the analysis within a meaningful spatiotemporal framework. Alternatively, SPREAD also utilizes JavaScript Object Notation (JSON) to encapsulate both the NWK information and metadata into a single file, facilitating streamlined data analysis. For a thorough comprehension of the structural organization of these inputs, please consult the project's GitHub repository (https://github.com/genpat-it/spread/). When utilizing ReportTree (Mixão et al., 2023) to generate these inputs, it is recommended to refer to the technical documentation of the tool for comprehensive guidance (https://github.com/insapathogenomics/ReporTree).

Integration of Genomic, Spatial, and Temporal Data

SPREAD was mainly developed with JavaScript to implement GrapeTree (Zhou et al., 2018) and OpenStreetMap/Leaflet (Herfort et al., 2021; Edler and Vetter, 2019) and manages outputs from ReportTree (Mixão et al., 2023) (Figure 1). Specifically, data input handling is designed to be versatile, enabling the application to proficiently accept inputs in NWK, TSV and JSON formats, not only via query strings but also through user-friendly drag-and-drop functionality or conventional file uploads. Such a design decision provides a high degree of flexibility in data management, accommodating a variety of data types and sources prevalent in genomic and epidemiological research. The user interface, developed using CSS3 and HTML5, focuses on user engagement and implements techniques for internationalization readiness. This ensures that the website or application can be easily adapted to different languages and regions, making it accessible and usable for a global audience. Although it currently lacks a responsive design adaptable to various devices and screen sizes, it maintains a clear and functional layout, ensuring that the application remains accessible and user-friendly up to tablet-sized screen resolutions, providing a fluid experience for users across a wide range of devices. The dashboard embraces modern web standards in its design and execution, fostering intuitive user interaction and a positive user experience. In the context of management and display of large and complex datasets, the high-performance JavaScript grid/spreadsheet component Ag-Grid is utilized (https://www.ag-grid.com/), as it's known for its efficiency and adaptability. This component is crucial for managing extensive data, ensuring an efficient user experience, even when navigating through large-scale datasets. The choice to use open-source libraries not only fosters transparency and community-driven development but also enables cost-effective and sustainable software solutions.
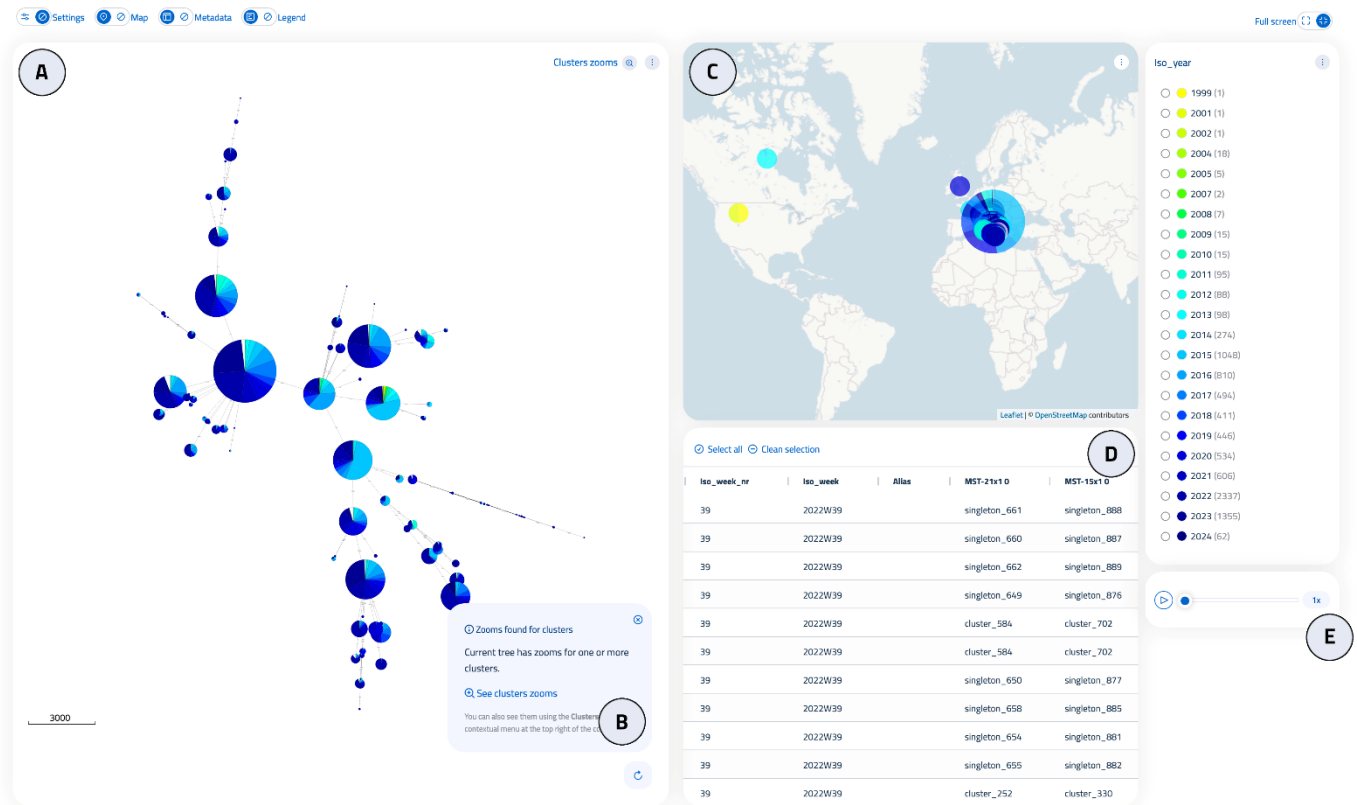
Figure 1. Overview of SPREAD: GrapeTree for the visualization of the tree (A), ReportTree for zoom functionality (B), OpenStreetMap and Leaflet to map nodes to represent geographic distribution (C), Ag-Grid for listing and filtering metadata (D), and JavaScript for a timeline component to visualize the context over time (E).

The data processing pipeline of SPREAD (Figure 2) begins with entering genomic data (i.e., NWK and TSV files to provide tree and metadata, or a JSON format that combines both) and associated metadata from a standardized bioinformatics services platform, which includes essential spatiotemporal information. The enriched cluster data, associated with temporal information and geographical coordinates, is then conveyed into SPREAD (Figure 2). The result is a dynamic and multifaceted visualization tool that enables a deeper understanding of the routes and patterns of infectious disease transmission. SPREAD users can navigate through the MSTs over time and across maps. Additionally, SPREAD offers a functionality that enables users to visualize the progression of the trees in a video-like format, enhancing user-friendly exploration of the data.

In-depth navigation through genetically close pathogens

ReportTree (Mixão et al., 2023) is also used and configured to produce zoomed-in views (Figure 3) and in this case, SPREAD empowers users with the ability to navigate directly to a more detailed view of a selected cluster with just a single click. This action maintains all the settings from the initial broader view, ensuring a continuous transition between different levels of data granularity and providing an efficient and user-friendly means to explore specific clusters within the larger dataset, without losing sight of the overall genomic landscape (Figure 3). The user interface has been meticulously crafted to ensure easy navigation through this complex amalgamation of information. SPREAD gives users the ability to conduct a detailed inspection of pathogen spread by facilitating interactive exploration of MSTs, tracking temporal changes, and visualizing geographic distribution, thereby simplifying the complex task of spatiotemporal data analysis for public health professionals.
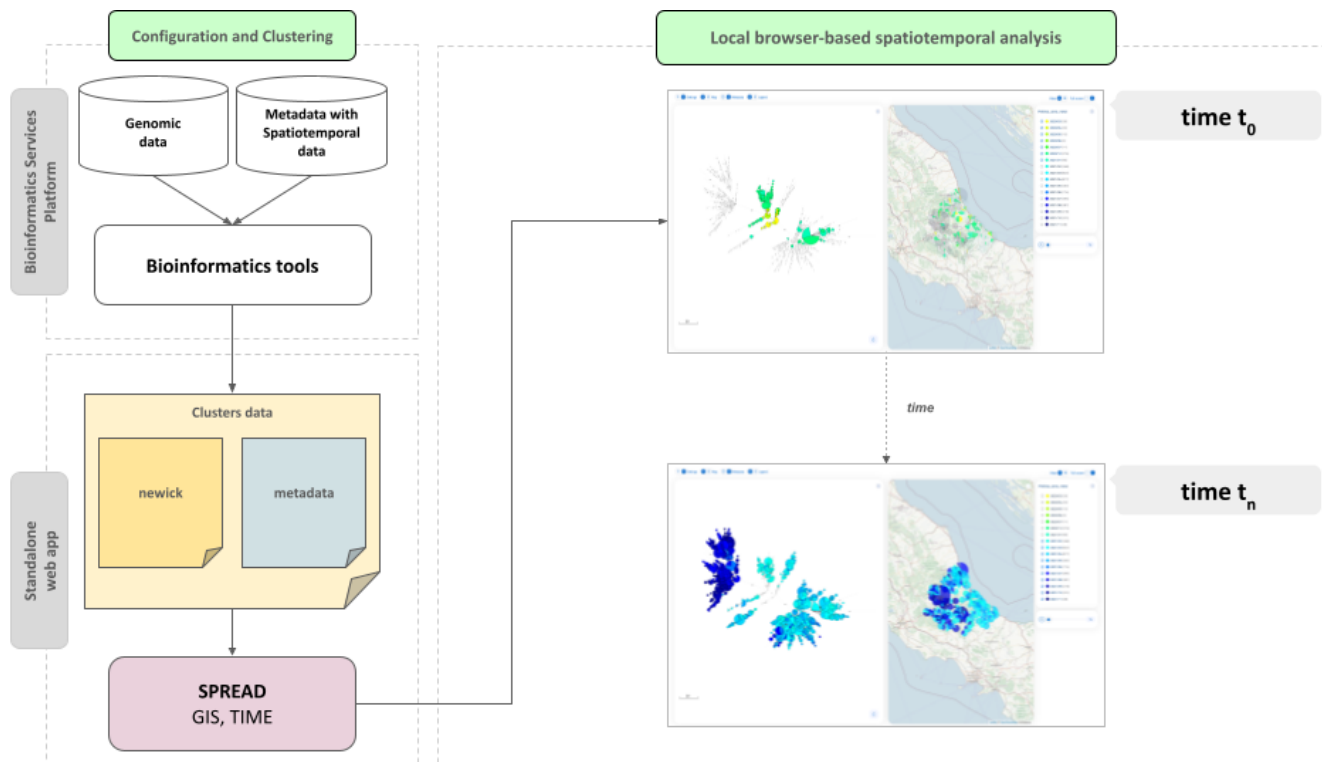
Figure 2. Schematic representation of the data analysis workflow within SPREAD, emphasizing the seamless integration of genomic, temporal, and geographic data produced by bioinformatics service platform.
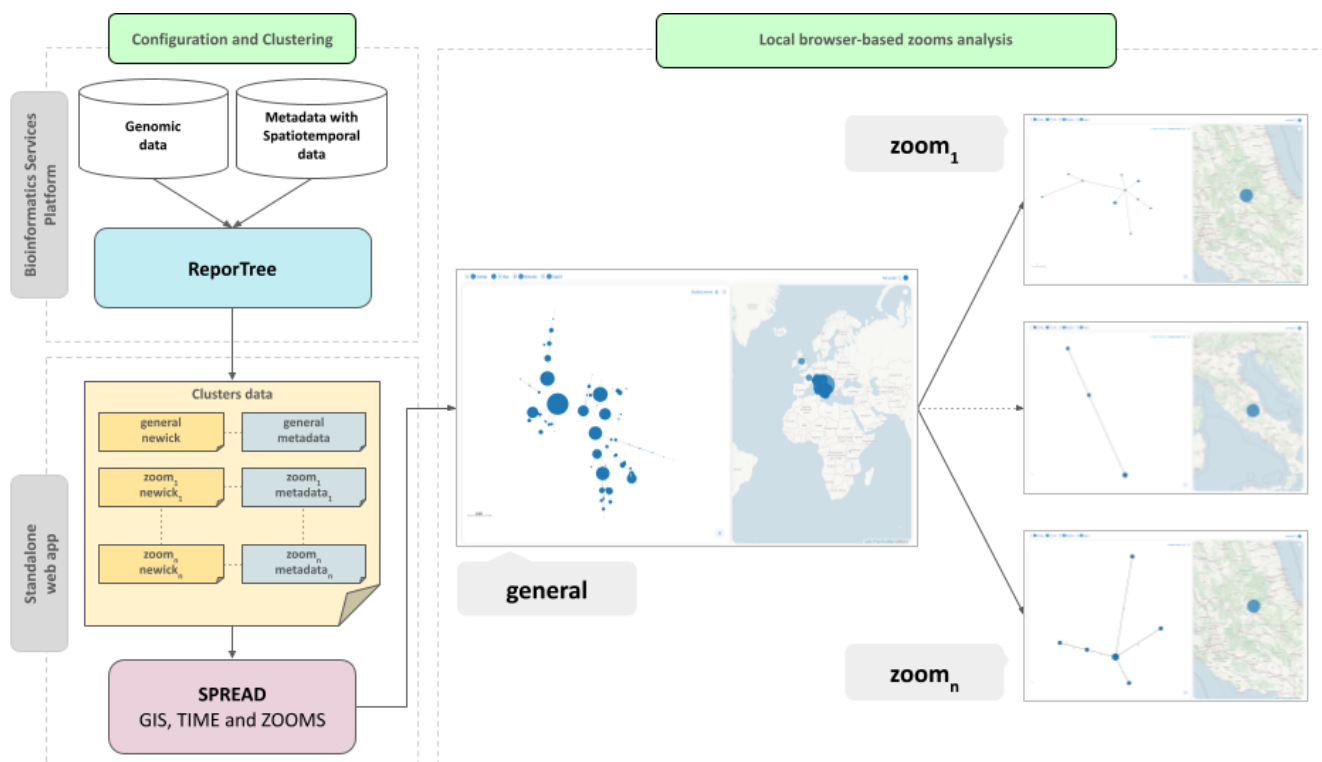
Figure 3. Schematic representation of the data analysis workflow within SPREAD, highlighting the integration of genomic, temporal, and geographic data based on ReportTree output. SPREAD enables one-click zoom functionality for detailed cluster examination, without losing the context of the broader genomic landscape.

Portability

The devised software solution is self-contained, offering advantages in terms of portability and an easy integration within a Docker framework (https://github.com/genpat-it/spread/pkgs/container/spread). This autonomous feature ensures that the application can be effortlessly deployed and operated across various systems, simplifying the installation process and enabling smooth integration into existing software ecosystems.

Results

The outcomes of the research demonstrate that the three main objectives (i) link spatiotemporal data to genomic relationships of pathogens' spreading, (ii) provide a solution to navigate through genetically close pathogens, and (iii) ensure data privacy based on open source web application for multiple pathogens have been effectively achieved.

Linkage of spatiotemporal data to genomic reconstruction of pathogens

The developed standalone dashboard links spatiotemporal data managed by Leaflet (Herfort et al., 2021; Edler and Vetter, 2019) to genomic relationships handled by GrapeTree (Zhou et al., 2018). This integration of a large amount of spatiotemporal data with genomic information (Table 1) enables users of SPREAD to effortlessly merge both datasets, facilitating the rapid identification of the geographic origins of pathogen sub-lineages, especially when dealing with compact and easily interpretable MSTs (Figure 4).
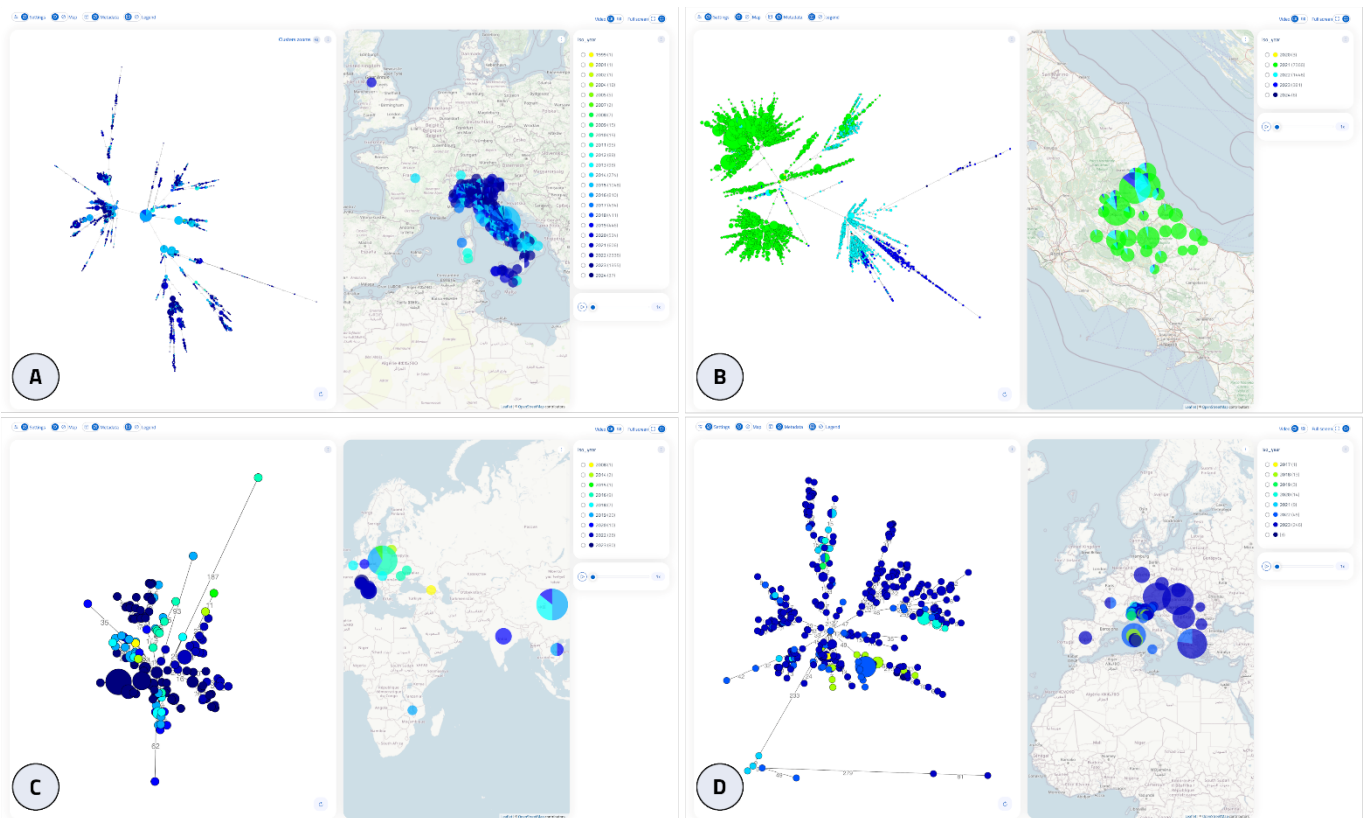

Figure 4. The integration of spatiotemporal data, shown on the OpenStreetMap map using Leaflet, with genomic relationships using GrapeTree, enables seamless merging of datasets. This facilitates rapid identification of geographic origins of pathogen sub-lineages, particularly effective with compact and interpretable MSTs. Examples include Listeria monocytogenes (A), SARS-CoV-2 (B), ASFV (C) and WNV (D).

In-depth navigation through genetically close pathogens

In instances of large genomic relationships, where identifying samples of interest becomes challenging, SPREAD is able to highlight, into the tree (i.e., GrapeTree (Zhou et al., 2018)) and related map (i.e., OpenStreetMap/Leaflet (Herfort et al., 2021; Edler and Vetter, 2019)), those pathogens from the collection which are genetically close to pathogens defined by users, based on a threshold setting used to delineate clusters of interest identified by ReportTree (Mixão et al., 2023) (Figure 5). This mechanism is typically employed in surveillance activities when new samples must be used to query a historical database of samples, in order to swiftly retrieve genetically close pathogens which may share the same epidemiological origin.In addition to the navigation through genetically close

pathogens, the users can also restrict the view of tree and map to samples belonging to specific genomic clusters, in order to refine the hypotheses about transmission pathways and sources (Figure 6).
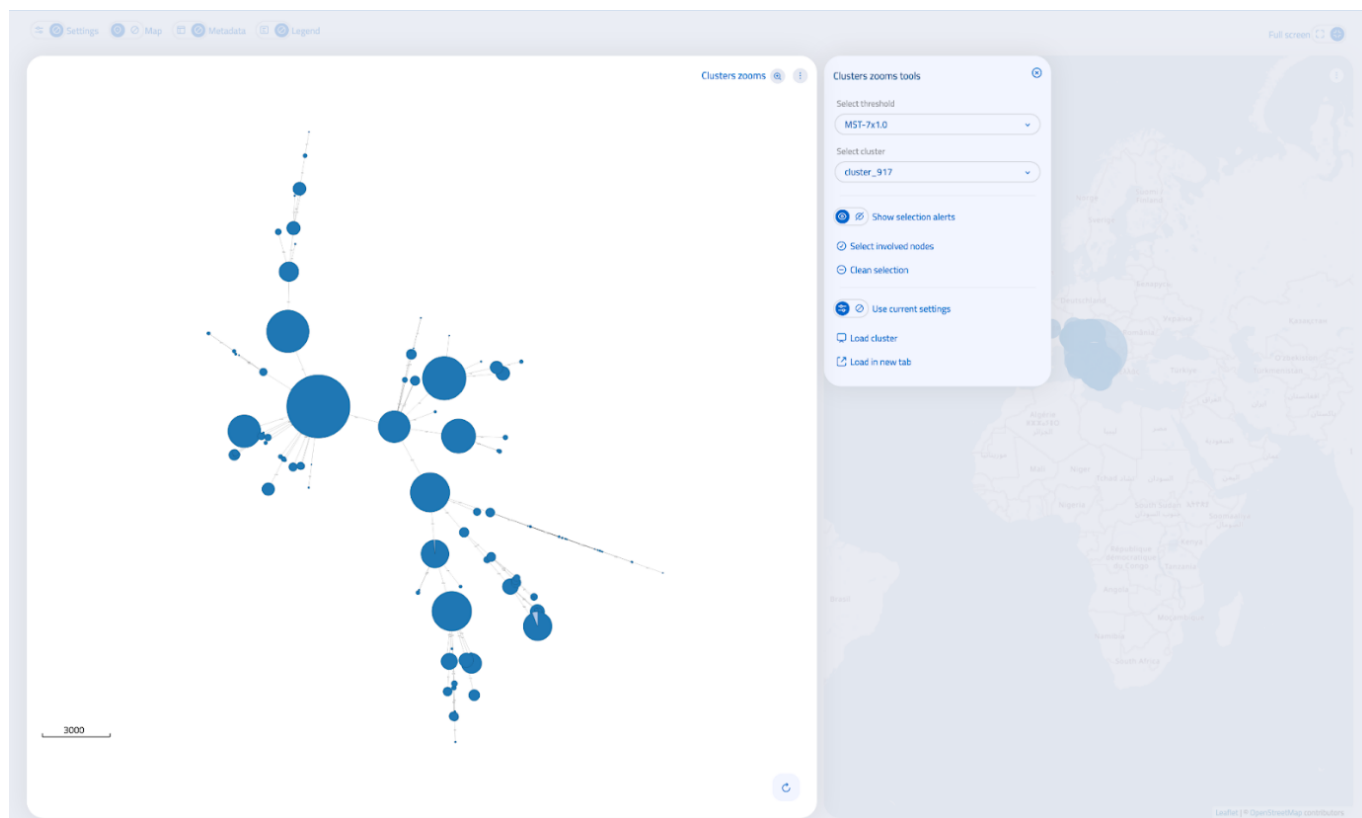
Figure 5. Illustration of SPREAD's functionality allowing highlight of pathogens close to user-defined samples within large genomic relationships. This aids in querying historical sample databases for swift retrieval of genetically linked samples, crucial in surveillance activities.
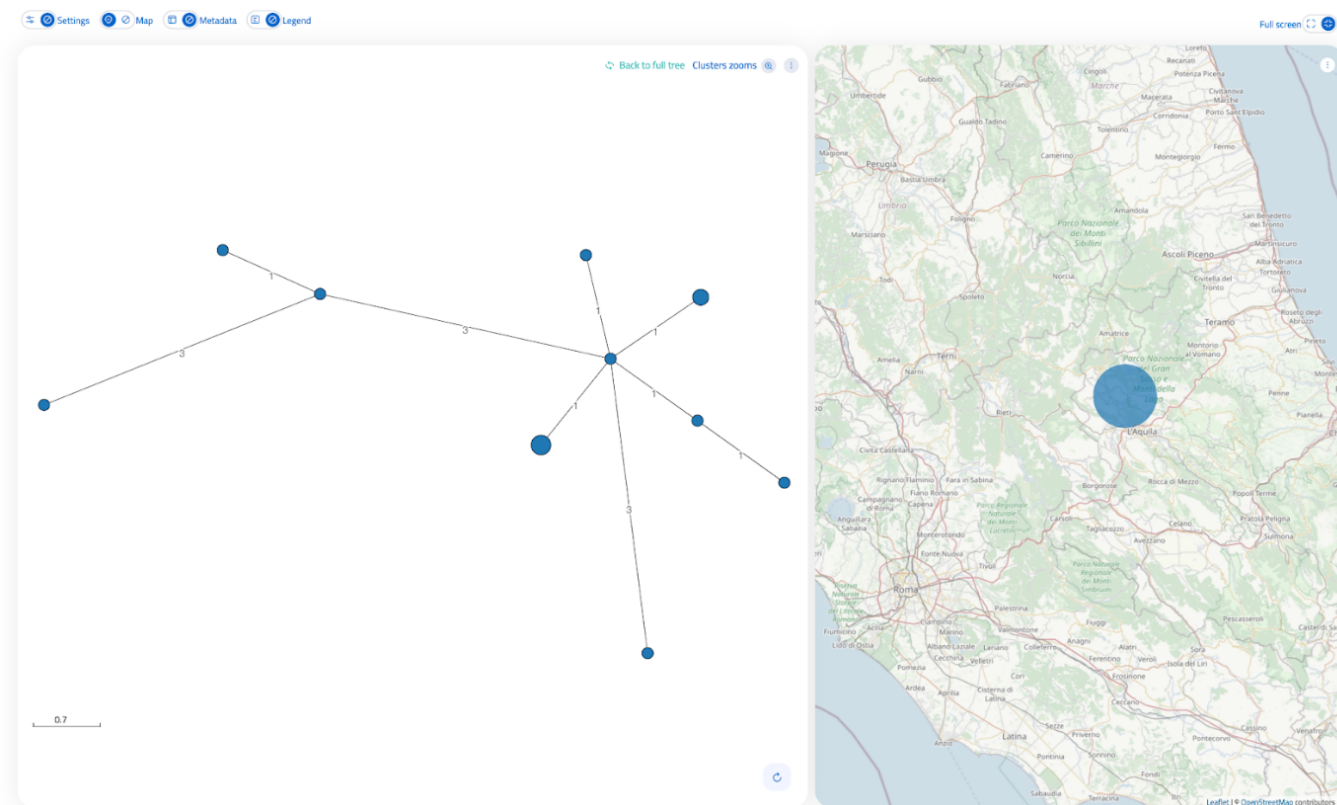
Figure 6. Illustration of SPREAD's functionality depicting the restriction of the MST and map to samples within specific genomic clusters, facilitating the refinement of hypotheses concerning transmission pathways and sources.

Data privacy based on open source standalone dashboard for multiple pathogens

SPREAD requires just a standard device equipped with a web browser, allowing users to easily drag and drop NWK, TSV or JSON files sourced from their own computing facilities or other organizations. This self-contained behavior of SPREAD contributes to enforce data privacy, since only users involved in data exchange can access it via the standalone dashboard. In addition, the presented proof of concept showed that this web application can manage large amount (Table 1) of bacterial (i.e., *L. monocytogenes*) and viral (i.e., SARS-CoV-2, ASFV and WNV) (Figure 4) pathogens. For instance, the developed tool enabled the identification of SARS-CoV-2 transmission clusters, linking them to specific events and locations, thereby supporting the choice of containment strategies and highlighting areas requiring intensified public health interventions. Concerning *L. monocytogenes* outbreaks, the application pinpointed the contamination source to specific food products distributed across various regions. By collapsing spatiotemporal and genomic data, it facilitated swift action, including product recalls and the implementation of preventative measures, underscoring the tool's usefulness in managing foodborne disease outbreaks. Furthermore, the application's analysis of ASFV and WNV presented a groundbreaking approach to understanding these devastating viruses, affecting swine populations globally and humans in southern, eastern and western Europe, respectively. The rapid spread of ASFV and WNV poses significant challenges, necessitating precise and rapid identification of transmission vectors to mitigate their impact. Through the application, genetic sequences of ASFV and WNV viruses from different outbreaks were analyzed, revealing not only the genomic relationships among strains, but also their movement across landscapes. The ability of SPREAD to manage multiple pathogens, such as bacteria and viruses, is made possible by the vcf2mlst program (Di Pasquale et al., 2021), which we developed to ensure compatibility between the output format from variant calling (i.e., snippy for viruses) and the input format of GrapeTree (Zhou et al., 2018), originally designed to process the cgMLST output format (i.e., chewBBACA for bacteria) (Silva et al., 2018).

Discussion

These technical implementations showcase an innovative approach for data visualization in infectious disease surveillance, enhancing user experience and providing researchers and public health professionals with actionable insights.

Fit for purpose standalone dashboard

The integration of ReportTree output (Mixão et al., 2023), OpenStreetMap/Leaflet (Herfort et al., 2021; Edler and Vetter, 2019), and GrapeTree (Zhou et al., 2018) within a unified dashboard provides an intuitive means for navigating complex data sets (Figure 1). The user-centric interface of the application caters to the requirements of both experts and non-experts in the field, enabling seamless exploration and interaction with MSTs, geographical data, and temporal sequences (Figure 2 and Figure 3). As recently proposed (Bhatia et al., 2021; Di Lorenzo et al., 2023), this functionality empowers users to interact with data dynamically, facilitating real-time visualization of the spread of infectious diseases across diverse geographic landscapes and temporal phases with unprecedented clarity and detail (Figure 4). The development of a standalone dashboard integrating GrapeTree together with spatial and temporal data visualization represents a significant advancement in the field of infectious disease surveillance (Nsoesie et al., 2015). SPREAD features cohesive, user-friendly visualization through GrapeTree (Zhou et al., 2018) and OpenStreetMap/Leaflet (Herfort et al., 2021; Edler and Vetter, 2019) (Figure 1), as well as navigation of complex zooms generated by ReportTree output (Mixão et al., 2023) (Figure 5 and Figure 6), enhancing user experience, accessibility and interpretability of complex genomic relationships and transmission pathways (Stockdale et al., 2022). This enables effortless navigation and interaction with detailed MSTs and spatial and temporal data, bringing about advanced surveillance capabilities to a broader audience, including those without specialized training in bioinformatics (Carriço et al., 2018). Efforts are ongoing to refine SPREAD presentation of complex genomic and epidemiological data, aiming to empower researchers, public health professionals, and policymakers to interpret data in a clean, clear and straightforward way. The ultimate goal is fostering a more informed and rapid response to infectious disease outbreaks, contributing to global health security (Tweed et al., 2022). A notable technical feature is the sophisticated handling of in-depth navigation through genetically close pathogens (Figure 5 and Figure 6) which is based on compatibility with Reportree output (Mixão et al., 2023) and crucial for navigating complex MSTs derived from extensive datasets of infectious disease pathogens (Table 1 and Figure 4).

Real-time management of large amounts of multi-species data

The SPREAD design focuses on optimizing performance based on browser capabilities, including an intelligent system for adjusting tree visualization depth according to browser performance metrics (Schwind et al., 2019). This adaptive feature ensures clear and informative visualizations without system overload, crucial for maintaining a consistent and responsive user experience across a variety of devices and technological environments (Salem et al., 2017), especially where access to high-performance computing resources is limited. Against a backdrop of the need for rapid responses to future global outbreaks (Volkov, 2022), the case study on SARS-CoV-2 demonstrates the standalone dashboard utility in real-time analysis and understanding of infectious disease spread, facilitating intuitive data exploration and allowing for a comprehensive observation of disease dynamics across various geographic and temporal scales. Such capabilities aid in immediate data interpretation and timely public health response formulation. As proven by the needs to develop One Health surveillance systems (Hayman et al., 2023), the case studies on *L. monocytogenes*, SARS-CoV-2, ASFV and WNV emphasize that SPREAD can be used to perform surveillance with large amounts (Table 1) of diverse species of bacteria and viruses (Figure 4). While recent dashboards dedicated to surveillance activities were developed independently, to manage bacteria at the alleles level (Liu et al., 2022) or viruses at the variants level (Nasri et al., 2023), the presented SPREAD offers the significant advantage of managing both kinds of pathogens based on the vcf2mlst program (Di Pasquale et al., 2021). This software allows compatibility between the variant calling output and GrapeTree (Zhou et al., 2018), whose original version manages the cgMLST output (i.e., chewBBACA) (Silva et al., 2018). Even if SPREAD efficiency to manage large volumes of data is already very decent, as required by pathogen surveillance in the age of big data (Sattari et al., 2020), an improvement through the implementation of an enriched timeline visualization (Neher and Bedford, 2018) will be performed in the near future to minimize a lack of responsive design for various devices (Salem et al., 2017).

Open source principles and data privacy through standalone architecture

A notable feature of SPREAD is its standalone architecture, which enables autonomous operation within a web browser, eliminating the need for installations and dependencies on server-dependent databases, thus complying to data privacy regulations (Hayes et al., 2020). This portable design facilitates easy encapsulation into Docker containers (Potdar et al., 2020), making it straightforward to integrate into existing tools and systems while ensuring data privacy (Xu et al., 2020; Brady et al., 2020). This architectural decision significantly broadens the access to advanced surveillance tools by removing common barriers associated with server-dependent applications, such as the need for specialized IT infrastructure and support. Aligned with the principles of open science, the application is also open source and freely accessible, fostering a community-driven approach for development and knowledge exchange, thereby facilitating collaborative enhancements and adaptations by the global scientific community (Heron et al., 2013). The open-source, freely accessible nature of the application ensures that state-of-the-art surveillance technology is accessible to a wide range of users, laying the foundation for extensive collaboration and innovation.

Future implementation of analyses of interest

Analytical areas have been identified for future enhancement to extend the tool's utility and reach. Proposed improvements include the integration of machine learning for source attribution (Castelli et al., 2023), and/or surveillance of antimicrobial resistance (Baker et al., 2023) and biocide resistance (Gmeiner et al., 2023), enriching the application with dynamic timeline visualization for cluster tracking (Neher and Bedford, 2018), and incorporating collaborative functionalities, which would significantly increase SPREAD capabilities (Ștefan et al., 2022). Additionally, enhancing the user interface for smoother navigation of complex data visualizations will maintain the application's main goal in infectious disease surveillance technology. Subsequent advancements will focus on refining SPREAD capabilities, incorporating user feedback (Young and McCauley, 2019) to ensure that the tool remains at the forefront of epidemiological research and public health surveillance. One significant shortcoming is the lack of a feature for sharing the working environment among collaborators, which restricts the ability to conduct collaborative analyses and impedes the continuous exchange of data and findings. This limitation underscores the need for a more integrated approach for data sharing, which could significantly enhance the utility and impact of the dashboard by facilitating real-time collaboration among researchers and public health professionals (Kostkova, 2018). To address these challenges while preserving SPREAD's principle of being a standalone application, it is proposed to integrate a peer-to-peer (P2P) network for information exchange (Overbeek et al., 2004). This network would facilitate the secure and efficient sharing of data among users, ensuring privacy and data integrity through the use of advanced cryptography. Such an approach would not only overcome the limitations related to collaborative analysis and data sharing but also adhere to the application's standalone nature. Furthermore, the non-responsive design across various devices poses a challenge to its accessibility and user-friendliness (Salem et al., 2017). In the digital age, where mobile devices play a crucial role in accessing and interacting with web-based tools, it seems essential to make SPREAD adapted to

different screen sizes and resolutions for users in field settings or when access to desktop computers is limited.

Conclusions

In conclusion, the development and implementation of the standalone dashboard named SPREAD represent a significant advancement in the field of infectious disease surveillance, merging sophisticated data analysis capabilities with intuitive navigation interfaces. The proposed future enhancements aim to further elevate the platform, including integrating machine learning algorithms for source attribution, antimicrobial resistance, and biocide resistance, implementing a timeline feature that displays cluster numbers over time, and facilitating collaboration through shared workspaces among team members. Additionally, the generation of Artificial Intelligence (AI)-based reports is anticipated to streamline the analytical process, offering more nuanced insights into disease transmission patterns. Moving forward, the emphasis will be on improving usability and assimilating feedback from the global scientific community. This, in turn, will enhance our collective ability to combat infectious diseases effectively.

Acknowledgement

Data Availability Statement

The study's GitHub repository, accessible at https://github.com/genpat-it/spread/, serves as a central platform for sharing project resources, including source code and documentation. This public repository promotes collaboration, transparency, and community engagement, allowing users and contributors to participate in the project's development, issue reporting, and suggestion of improvements. Its open access fosters knowledge sharing among researchers, developers, and stakeholders, aligning with the project's commitment to openness and community involvement.

References

Akther, S., Bezrucenkovas, E., Sulkow, B., Panlasigui, C., Li, L., Qiu, W., & Di, L. (2020). CoV Genome Tracker: tracing genomic footprints of Covid-19 pandemic. Cold Spring Harbor Laboratory. https://doi.org/10.1101/2020.04.10.036343

Arshad, S., Arshad, J., Khan, M. M., & Parkinson, S. (2021). Analysis of security and privacy challenges for DNA-genomics applications and databases. In Journal of Biomedical Informatics (Vol. 119, p. 103815). Elsevier BV. https://doi.org/10.1016/j.jbi.2021.103815

Baker, K. S., Jauneikaite, E., Hopkins, K. L., Lo, S. W., Sánchez-Busó, L., Getino, M., Howden, B. P., Holt, K. E., Musila, L. A., Hendriksen, R. S., Amoako, D. G., Aanensen, D. M., Okeke, I. N., Egyir, B., Nunn, J. G., Midega, J. T., Feasey, N. A., & Peacock, S. J. (2023). Genomics for public health and international surveillance of antimicrobial resistance. In The Lancet Microbe (Vol. 4, Issue 12, pp. e1047–e1055). Elsevier BV. https://doi.org/10.1016/s2666-5247(23)00283-5

Bhatia, S., Lassmann, B., Cohn, E., Desai, A. N., Carrion, M., Kraemer, M. U. G., Herringer, M., Brownstein, J., Madoff, L., Cori, A., & Nouvellet, P. (2021). Using digital surveillance tools for near real-time mapping of the risk of infectious disease spread. In npj Digital Medicine (Vol. 4, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41746-021-00442-3

Brady, K., Moon, S., Nguyen, T., & Coffman, J. (2020). Docker Container Security in Cloud Computing. In 2020 10th Annual Computing and Communication Workshop and Conference (CCWC). 2020 10th Annual Computing and Communication Workshop and Conference (CCWC). IEEE. https://doi.org/10.1109/ccwc47524.2020.9031195

Carriço, J. A., Rossi, M., Moran-Gilad, J., Van Domselaar, G., & Ramirez, M. (2018). A primer on microbial bioinformatics for nonbioinformaticians. In Clinical Microbiology and Infection (Vol. 24, Issue 4, pp. 342–349). Elsevier BV. https://doi.org/10.1016/j.cmi.2017.12.015

Castelli, P., De Ruvo, A., Bucciacchio, A., D'Alterio, N., Cammà, C., Di Pasquale, A., & Radomski, N. (2023). Harmonization of supervised machine learning practices for efficient source attribution of Listeria monocytogenes based on genomic data. In BMC Genomics (Vol. 24, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1186/s12864-023-09667-w

Chen, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. In iMeta (Vol. 2, Issue 2). Wiley. https://doi.org/10.1002/imt2.107

Di Lorenzo, A., Mangone, I., Colangeli, P., Cioci, D., Curini, V., Vincifori, G., Mercante, M. T., Di Pasquale, A., Radomski, N., & Iannetti, S. (2023). One health system supporting surveillance during COVID-19 epidemic in Abruzzo region, southern Italy. In One Health (Vol. 16, p. 100471). Elsevier BV. https://doi.org/10.1016/j.onehlt.2022.100471

Di Pasquale, A., Radomski, N., Mangone, I., Calistri, P., Lorusso, A., & Cammà, C. (2021). SARS-CoV-2 surveillance in Italy through phylogenomic inferences based on Hamming distances derived from pan-SNPs, -MNPs and -InDels. In BMC Genomics (Vol. 22, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1186/s12864-021-08112-0

Edler, D., & Vetter, M. (2019). The Simplicity of Modern Audiovisual Web Cartography: An Example with the Open-Source JavaScript Library leaflet.js. In KN - Journal of Cartography and Geographic Information (Vol. 69, Issue 1, pp. 51–62). Springer Science and Business Media LLC. https://doi.org/10.1007/s42489-019-00006-2

Gmeiner, A., Ivanova, M., Kamau Njage, P. M., Hansen, L. T., Chindelevitch, L., & Leekitcharoenphon, P. (2023). Quantitative prediction of disinfectant tolerance inListeria monocytogenesusing whole genome sequencing and machine learning. Cold Spring Harbor Laboratory. https://doi.org/10.1101/2023.11.05.565740

Grubaugh, N. D., Ladner, J. T., Lemey, P., Pybus, O. G., Rambaut, A., Holmes, E. C., & Andersen, K. G. (2018). Tracking virus outbreaks in the twenty-first century. In Nature Microbiology (Vol. 4, Issue 1, pp. 10–19). Springer Science and Business Media LLC. https://doi.org/10.1038/s41564-018-0296-2

Hayes, D., Cappa, F., & Le-Khac, N. A. (2020). An effective approach to mobile device management: Security and privacy issues associated with mobile applications. In Digital Business (Vol. 1, Issue 1, p. 100001). Elsevier BV. https://doi.org/10.1016/j.digbus.2020.100001

Hayman, D. T. S., Adisasmito, W. B., Almuhairi, S., Behravesh, C. B., Bilivogui, P., Bukachi, S. A., Casas, N., Becerra, N. C., Charron, D. F., Chaudhary, A., Ciacci Zanella, J. R., Cunningham, A. A., Dar, O., Debnath, N., Dungu, B., Farag, E., Gao, G. F., Khaitsa, M., Machalaba, C., … Koopmans, M. (2023). Developing One Health surveillance systems. In One Health (Vol. 17, p. 100617). Elsevier BV. https://doi.org/10.1016/j.onehlt.2023.100617

Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J., & Zipf, A. (2021). The evolution of humanitarian mapping within the OpenStreetMap community. In Scientific Reports (Vol. 11, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41598-021-82404-z

Heron, M. J., Hanson, V. L., & Ricketts, I. (2013). Open Source and Accessibility: Advantages and Limitations. In Journal of Interaction Science (Vol. 1, Issue 1, p. 2). Springer Science and Business Media LLC. https://doi.org/10.1186/2194-0827-1-2

Kędzior, M. (2020). The right to data protection and the COVID-19 pandemic: the European approach. In ERA Forum (Vol. 21, Issue 4, pp. 533–543). Springer Science and Business Media LLC. https://doi.org/10.1007/s12027-020-00644-4

Kostkova, P. (2018). Disease surveillance data sharing for public health: the next ethical frontiers. In Life Sciences, Society and Policy (Vol. 14, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1186/s40504-018-0078-x

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. In Nucleic Acids Research (Vol. 49, Issue W1, pp. W293–W296). Oxford University Press (OUP). https://doi.org/10.1093/nar/gkab301

Liu, Y.-Y., Chen, C.-C., Yang, C.-H., Hsieh, H.-Y., He, J.-X., Lin, H.-H., & Lee, C.-C. (2022). LmTraceMap: A Listeria monocytogenes fast-tracing platform for global surveillance. In Z. Ruan (Ed.), PLOS ONE (Vol. 17, Issue 5, p. e0267972). Public Library of Science (PLoS). https://doi.org/10.1371/journal.pone.0267972

Mangone, I., Radomski, N., Di Pasquale, A., Santurbano, A., Calistri, P., Cammà, C., & Maassen, K. (2021). Refinement of the COHESIVE Information System towards a unified ontology of food terms for the public health organizations (COHESIVE). https://doi.org/10.5281/ZENODO.5482422

Mixão, V., Pinto, M., Sobral, D., Di Pasquale, A., Gomes, J. P., & Borges, V. (2023). ReporTree: a surveillance-oriented tool to strengthen the linkage between pathogen genetic clusters and epidemiological data. In Genome Medicine (Vol. 15, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1186/s13073-023-01196-1

Nasri, F., Kongkitimanon, K., Wittig, A., Cortés, J. S., Brinkmann, A., Nitsche, A., Schmachtenberg, A.-J., Renard, B. Y., & Fuchs, S. (2023). MpoxRadar: a worldwide MPXV genomic surveillance dashboard. In Nucleic Acids Research (Vol. 51, Issue W1, pp. W331–W337). Oxford University Press (OUP). https://doi.org/10.1093/nar/gkad325

Neher, R. A., & Bedford, T. (2018). Real-Time Analysis and Visualization of Pathogen Sequence Data. In C. S. Kraft (Ed.), Journal of Clinical Microbiology (Vol. 56, Issue 11). American Society for Microbiology. https://doi.org/10.1128/jcm.00480-18

Nsoesie, E. O., Kluberg, S. A., Mekaru, S. R., Majumder, M. S., Khan, K., Hay, S. I., & Brownstein, J. S. (2015). New digital technologies for the surveillance of infectious diseases at mass gathering events. In Clinical Microbiology and Infection (Vol. 21, Issue 2, pp. 134–140). Elsevier BV. https://doi.org/10.1016/j.cmi.2014.12.017

Overbeek, R., Disz, T., & Stevens, R. (2004). The SEED. In Communications of the ACM (Vol. 47, Issue 11, pp. 46–51). Association for Computing Machinery (ACM). https://doi.org/10.1145/1029496.1029525

Payne, M., Hu, D., Wang, Q., Sullivan, G., Graham, R. M., Rathnayake, I. U., Jennison, A. V., Sintchenko, V., & Lan, R. (2024). DODGE: Automated point source bacterial outbreak detection using cumulative long term genomic surveillance. Cold Spring Harbor Laboratory. https://doi.org/10.1101/2024.01.21.24301506

Perry, H. N., McDonnell, S. M., Alemu, W., Nsubuga, P., Chungong, S., Otten, M. W., Jr, Lusamba-dikassa, P. S., & Thacker, S. B. (2007). Planning an integrated disease surveillance and response system: a matrix of skills and activities. In BMC Medicine (Vol. 5, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1186/1741-7015-5-24

Potdar, A. M., D G, N., Kengond, S., & Mulla, M. M. (2020). Performance Evaluation of Docker Container and Virtual Machine. In Procedia Computer Science (Vol. 171, pp. 1419–1428). Elsevier BV. https://doi.org/10.1016/j.procs.2020.04.152

Pronyk, P. M., de Alwis, R., Rockett, R., Basile, K., Boucher, Y. F., Pang, V., Sessions, O., Getchell, M., Golubchik, T., Lam, C., Lin, R., Mak, T.-M., Marais, B., Twee-Hee Ong, R., Clapham, H. E., Wang, L., Cahyorini, Y., Polotan, F. G. M., Rukminiati, Y., … Sintchenko, V. (2023). Advancing pathogen genomics in resource-limited settings. In Cell Genomics (Vol. 3, Issue 12, p. 100443). Elsevier BV. https://doi.org/10.1016/j.xgen.2023.100443

Salem, B., Alves Lino, J., & Simons, J. (2017). A Framework for Responsive Environments. In Lecture Notes in Computer Science (pp. 263–277). Springer International Publishing. https://doi.org/10.1007/978-3-319-56997-0_21

Sattari, M., Jahanbakhsh, M., Ashrafi-Rizi, H., & Jangi, M. (2020). Big data in COVID-19 surveillance system: A commentary. In Journal of Education and Health Promotion (Vol. 9, Issue 1, p. 329). Medknow. https://doi.org/10.4103/jehp.jehp_303_20

Schwind, A., Midoglu, C., Alay, Ö., Griwodz, C., & Wamser, F. (2019). Dissecting the performance of YouTube video streaming in mobile networks. In International Journal of Network Management (Vol. 30, Issue 3). Wiley.

https://doi.org/10.1002/nem.205

Silva, M., Machado, M. P., Silva, D. N., Rossi, M., Moran-Gilad, J., Santos, S., Ramirez, M., & Carriço, J. A. (2018). chewBBACA: A complete suite for gene-by-gene schema creation and strain identification. In Microbial Genomics (Vol. 4, Issue 3). Microbiology Society. https://doi.org/10.1099/mgen.0.000166

Ștefan, I. A., Ștefan, A., Tsalapatas, H., Heidmann, O., & Gheorghe, A. F. (2022). Collaborative decision-making in software research projects: the innovation challenge. In Procedia Computer Science (Vol. 199, pp. 1318–1326). Elsevier BV. https://doi.org/10.1016/j.procs.2022.01.167

Stockdale, J. E., Liu, P., & Colijn, C. (2022). The potential of genomics for infectious disease forecasting. In Nature Microbiology (Vol. 7, Issue 11, pp. 1736–1743). Springer Science and Business Media LLC. https://doi.org/10.1038/s41564-022-01233-6

Tweed, S., Stewart, D., Hornsey, E., & Graham, W. (2022). Increasing role of Public Health Rapid Response Teams in infectious disease outbreaks. In European Journal of Public Health (Vol. 32, Issue Supplement_3). Oxford University Press (OUP). https://doi.org/10.1093/eurpub/ckac130.022

Volkov, V. (2022). System analysis of the fast global coronavirus disease 2019 (COVID-19) spread. Can we avoid future pandemics under global climate change? In Communicative & Integrative Biology (Vol. 15, Issue 1, pp. 150–157). Informa UK Limited. https://doi.org/10.1080/19420889.2022.2082735

Xu, X., Xu, A., Jiang, Y., Wang, Z., Wang, Q., Zhang, Y., & Wen, H. (2020). Research on Security Issues of Docker and Container Monitoring System in Edge Computing System. In Journal of Physics: Conference Series (Vol. 1673, Issue 1, p. 012067). IOP Publishing. https://doi.org/10.1088/1742-6596/1673/1/012067

Young, S. F., & McCauley, C. D. (2019). User-Driven Feedback Tools for Leader Development. In Feedback at Work (pp. 265–285). Springer International Publishing. https://doi.org/10.1007/978-3-030-30915-2_14

Zhou, Z., Alikhan, N.-F., Sergeant, M. J., Luhmann, N., Vaz, C., Francisco, A. P., Carriço, J. A., & Achtman, M. (2018). GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. In Genome Research (Vol. 28, Issue 9, pp. 1395–1404). Cold Spring Harbor Laboratory. https://doi.org/10.1101/gr.232397.117

Zhou, Z., Alikhan, N.-F., Mohamed, K., Fan, Y., & Achtman, M. (2019). The EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and Escherichia core genomic diversity. In Genome Research (Vol. 30, Issue 1, pp. 138–152). Cold Spring Harbor Laboratory. https://doi.org/10.1101/gr.251678.119