نهمین کنفرانس بین‌المللی
مطالعات میان‌رشته‌ای در نانو فناوری
9th International Conference on
Interdisciplinary Studies in Nanotechnology
www.nano.bcnf.ir

Bioinformatic Analysis of *Salmonella* and *Shigella* Genomic sequences for the designing of Specific probe and primer Sequences for nanosensing and PCR Diagnostic methods

Haniyeh Pourtaghi Khorsand , Farrokh Karimi*, Saba Jalali
Department of biotechnology, faculty of science, university of Maragheh, Maragheh, Iran
karimifm@maragheh.ac.ir

ABSTRACT

This study provides a comprehensive bioinformatic analysis of the genomic sequences of Salmonella and Shigella with the aim of identifying species-specific sequences suitable for diagnostic applications. Given the clinical significance and pathogenic potential of these bacteria, rapid and accurate detection is critical. Although conventional molecular methods remain effective, emerging approaches, including nanotechnology-based diagnostics, offer potential advantages owing to their enhanced sensitivity and precision. Genomic sequences of Salmonella and Shigella were retrieved from the NCBI database and subjected to BLAST analysis to identify unique regions specific to each species. Distinct species-specific sequences were successfully identified for both organisms, providing a foundation for the development of targeted molecular assays. These sequences can be employed in the design of primers and probes for diagnostic platforms, including PCR and nanotechnology-based methods, thereby improving both the accuracy and speed of pathogen detection. The results highlight the value of integrating bioinformatic analyses with advanced molecular diagnostic strategies to achieve precise identification of clinically relevant bacterial pathogens.

Keywords: In silico, Salmonella, Shigella, nanotechnology, polymerase chain reaction (PCR)

INTRODUCTION

Salmonella and Shigella are among the most clinically significant bacterial pathogens and constitute major etiological agents of human intestinal diseases. These organisms can cause severe gastrointestinal infections, typically manifesting as bloody diarrhea and fever, and may result in serious complications, particularly in vulnerable populations such as children, the elderly, and immunocompromised individuals. Beyond their direct clinical impact, infections caused by these bacteria impose substantial burdens on public health systems and incur significant economic costs in healthcare and food industry sectors. These challenges underscore the need for comprehensive and sustained strategies for diagnosis, treatment, prevention, and control across the entire food supply chain [1]. Salmonella and Shigella are well-established zoonotic pathogens, with outbreaks frequently linked to contaminated animal reservoirs and products of animal origin. The transmission of these bacteria from animals to humans represents a pivotal mechanism in the propagation of foodborne epidemics [2]. The rapid and precise detection of these pathogens is critical for public health and the effective prevention of disease outbreaks. Prompt identification not only limits pathogen transmission but also optimizes clinical management and decreases the associated healthcare burden [3]. Conventional approaches for the identification of pathogenic bacteria primarily include microbial culture on selective media and immunoassays. Microbial culture is widely regarded as the diagnostic gold standard; however, it necessitates extended incubation periods and trained personnel, and its

sensitivity is markedly reduced in samples containing bacteria in a viable but non-culturable (VBNC) state. While immunoassays are capable of detecting specific antigens, they frequently demonstrate limited sensitivity and specificity in complex sample matrices, and cross-reactivity with non-target organisms can compromise the reliability of the results [4]. Advancements in molecular technologies have facilitated the development and widespread implementation of contemporary diagnostic methodologies, including conventional polymerase chain reaction (PCR), real-time PCR, next-generation sequencing (NGS), and innovative nanosensors and biosensors for pathogen detection. PCR-based assays, which depend on precise primer design, enable the targeting of pathogen-specific genomic regions, thereby enhancing both sensitivity and specificity relative to traditional approaches [۵]. In recent years, nanosensor platforms have garnered significant attention owing to their capacity to support rapid, highly sensitive diagnostic systems that are deployable in field settings. Empirical evidence demonstrates that these sensors can detect bacterial species with high specificity within brief timeframes [۶]. In numerous molecular diagnostic frameworks, pathogen-specific genomic sequences-particularly loci exhibiting interspecies variability-serve as precise targets for probe or detector design, thereby playing a pivotal role in optimizing diagnostic accuracy [۷]. Recent bioinformatic studies have applied comprehensive genomic analyses, encompassing comparative genomics and pan-genome approaches, to identify species- or serovar-specific genomic markers for Salmonella and Shigella. These analyses leverage complete genome sequence datasets to identify sequences conserved within target populations yet absent in non-target species, thereby yielding molecular targets with high diagnostic specificity [۸]. Furthermore, investigations into pathogen genome architecture and population diversity indicate that comprehensive genomic analysis not only facilitates accurate species identification but also provides critical insights into genetic variation and evolutionary dynamics within bacterial populations. The 16S rRNA gene is widely employed as a molecular target in bacterial diagnostics due to its evolutionary conservation across bacterial taxa. Its incorporation into diagnostic workflows has been shown in multiple studies to enable species-level discrimination and to enhance overall diagnostic sensitivity.[۹]

MATERIALS AND METHODS

Candidate Gene Selection

A systematic review of the literature was performed to identify species-specific sequences in Salmonella and Shigella, with the objective of determining genes exhibiting the highest specificity for molecular detection of these pathogens. The invA gene has been established as a reliable and highly specific genetic marker for the genus Salmonella, as it is conserved across all recognized serotypes and absent in closely related bacterial taxa. Accordingly, this gene has been extensively utilized in molecular diagnostic assays as a marker for Salmonella detection. In contrast, the ipaH gene, present on both plasmid and chromosomal elements in Shigella, is conserved across all species within this genus. The presence of multiple copies and the relatively high sequence conservation render ipaH a robust marker for Shigella identification using PCR and other molecular approaches (employed as a species-specific marker for Shigella). Based on these characteristics, these two genes were selected as primary targets in the present study to identify reliable, species-specific sequences for each pathogen.

Data Collection and Genome Retrieval

Following the selection of target genes, reference sequences were retrieved from the National Center for Biotechnology Information (NCBI) genomic database. For Shigella, the sequence corresponding to the ipaH gene (NCBI accession number: M32063.1) was extracted, representing the complete gene sequence from a reference isolate. Similarly, the invA gene sequence in Salmonella (NCBI accession number: M90846.1) was retrieved as a representative reference. These sequences served as primary input for subsequent bioinformatic analyses aimed at identifying and validating species-specific sequences within broader genomic datasets.

Sequence Analysis and BLAST

Following retrieval of the reference sequences for ipaH and invA, these sequences were analyzed using BLASTn to identify regions absent in non-target bacteria and uniquely present within the target genus. BLASTn facilitates rapid comparison of query sequences against comprehensive bacterial genomic

databases and is widely employed for the accurate identification of species- or genus-specific loci.

In this study, the results were systematically evaluated to select gene segments exhibiting minimal similarity to homologous sequences in related taxa, particularly other members of the Enterobacteriaceae family. The regions thus identified provide a basis for the development of molecular diagnostic assays and the detection of genus- or species-specific genomic markers in subsequent investigations.

Analysis of the 16S rRNA Gene and Its Variable Regions

To support the assessment of specificity and the validation of molecular markers, the 16S rRNA gene was examined as a conserved yet adaptable genetic marker in bacteria. This gene comprises nine hypervariable regions (V1–V9) that display sequence heterogeneity across species, facilitating discrimination at both the genus and species levels (Figure1). The conserved regions flanking these hypervariable segments enable the design of universal primers. The interplay between conserved and variable regions thus renders the 16S rRNA gene a robust tool for the precise identification of bacterial taxa [1 •]. In the present study, the V3 and V6 hypervariable regions were selected as candidate targets for the development of genus-specific markers for Salmonella and Shigella. This selection was guided by the pronounced sequence divergence of these regions relative to closely related taxa, alongside the high degree of sequence conservation observed with the target genera.

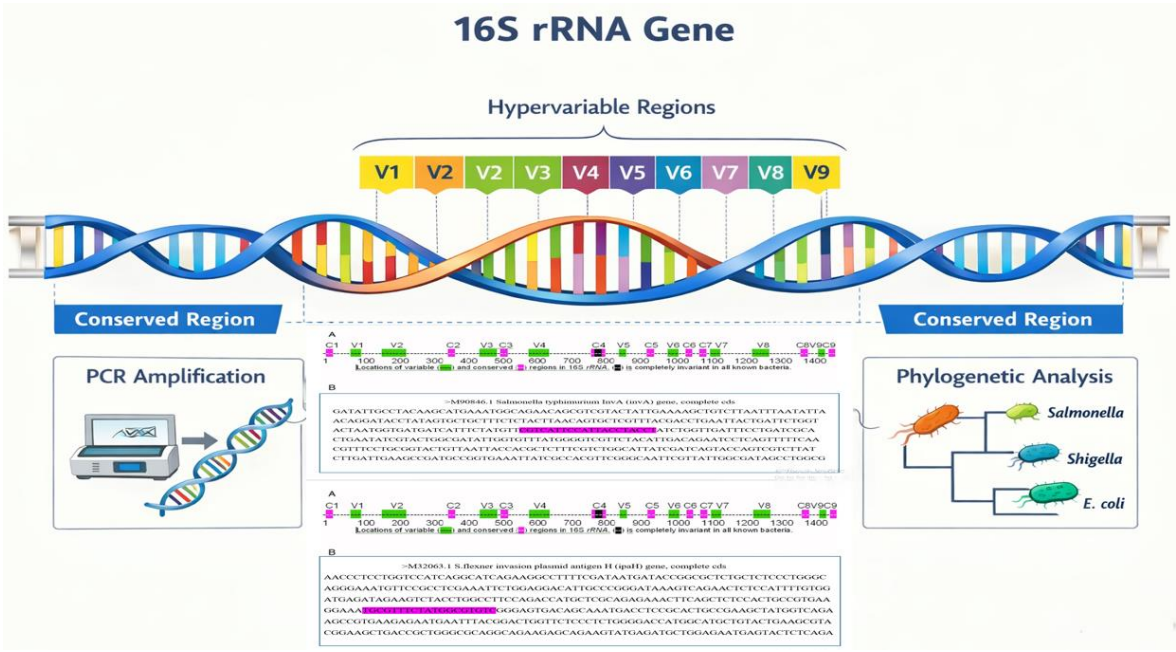

Fig. 1. *A schematic illustration of the 16S rRNA gene depicting its conserved and variable regions, accompanied by the species-specific sequences selected for Shigella and Salmonella.*

Primer and Probe Design Based on Nanotechnology for Molecular Detection

Beyond conventional PCR-based methodologies, the species-specific sequences identified in this study were systematically evaluated for their suitability in nanotechnology-enabled biosensing platforms. Genomic regions derived from Salmonella enterica (invA gene) and Shigella spp. (ipaH gene) were rigorously validated using BLAST analyses and subsequently selected as molecular targets for the design of highly specific oligonucleotide primers and probes. These targets were chosen to ensure dual compatibility with both nucleic acid amplification assays and nanomaterial-based sensing systems.

The incorporation of nanotechnology into molecular diagnostics has been shown to improve analytical sensitivity, enhance signal transduction efficiency, and reduce the limit of detection relative to conventional techniques. In this context, the selected gene fragments were comprehensively evaluated with respect to thermodynamic stability, GC content, propensity for secondary structure formation, and probe–target hybridization efficiency. These parameters were optimized to ensure robust performance within nanoparticle-assisted detection platforms, including carbon-based nanomaterials and graphene-derived biosensors. Accordingly, the identified species-specific genomic regions fulfill a dual functional role: they serve as amplification targets in PCR-based diagnostic assays and as hybridization targets in nanoscale biosensing architectures. This integrated design strategy strengthens the translational applicability of the developed molecular marker and supports the advancement of rapid, sensitive, and field-deployable diagnostic tools for clinical diagnostics and food safety monitoring. The integration of nanotechnology-oriented approaches within the present bioinformatic framework positions this study in alignment with current advances in nanodiagnostics and provides a foundation for the development of next-generation molecular biosensing platforms.

RESULT AND DISCUSSION

Species-Specific Sequences of invA and ipaH Genes for Nanosensing and PCR Applications

The invA and ipaH genes were selected as candidate markers for the design of species-specific sequences for Salmonella and Shigella, based on a comprehensive review of prior studies demonstrating that these genes are both conserved within the genomes of the target species and uniquely characteristic. BLASTn analysis identified the short sequence CGTCATTCCATTACCTACCT as a species-specific fragment for Salmonella and TGCGTTTCTATGGCGTGTC as a species-specific fragment for Shigella (Figure 2). Validation of these sequences against thousands of entries in public genomic databases indicated minimal occurrence in non-target species, confirming their suitability as diagnostic markers in nanomaterial-based assays, nanosensors, and other molecular methodologies. These sequences are appropriate for the design of primers and probes in polymerase chain reaction (PCR) and nanosensor-based platforms, enabling rapid and precise pathogen detection. Furthermore, BLAST analyses corroborate the utility of these species-specific sequences in the development of phylogenetic markers and taxonomic classification schemes. Integrating bioinformatic analyses with molecular diagnostic techniques is crucial for enhancing assay reliability. Multiple lines of evidence suggest that comprehensive genomic analysis and identification of unique regions can significantly improve the sensitivity and specificity of nanomaterial-based and molecular assays, including PCR. Primer and probe design guided by species-specific sequences reduces the risk of cross-reactivity, thereby improving diagnostic accuracy [11]. Emerging technologies, particularly the combination of PCR with nanotechnology, facilitate accelerated pathogen detection. Incorporating nanosensors and nanomaterials into assay design allows development of platforms with reduced detection times, increased sensitivity, and lower limits of detection compared to conventional methods. For instance, nanomaterial-based sensors can detect subtle alterations in the presence of target sequences through electrochemical or optical signals, enhancing diagnostic performance in both clinical and environmental samples [12]. The utilization of nanomaterials increases the effective surface area available for probe-target interactions, enabling detection of extremely low concentrations of pathogen DNA. This capability is particularly critical in clinical specimens with low bacterial loads or environmental samples where

conventional detection methods may be inadequate. Integrating bioinformatically designed species-specific sequences with nanotechnology supports the creation of rapid, multiplexed, and field-deployable diagnostic systems. This approach enables simultaneous identification of Salmonella and Shigella and facilitates rapid phylogenetic analysis. The findings underscore the potential of nanotechnology to enhance molecular diagnostic sensitivity and accuracy, reduce detection time, and improve assay reliability, thereby laying the groundwork for next-generation molecular diagnostic tools. Despite these advances, bioinformatic analyses possess inherent limitations. A primary constraint is dependence on the quality and completeness of genomic data in public repositories; incomplete or low-quality datasets can compromise analytical accuracy. Additionally, comparative algorithms such as BLAST and other sequence search tools may yield variable results depending on parameter selection, necessitating careful optimization and expert interpretation [13]. Nonetheless, conducting bioinformatic analyses prior to molecular assays offers substantial advantages: genomic analyses guide precise target selection, optimize primer and probe design, and minimize potential experimental errors, ultimately conserving time and resources. These approaches also provide a comprehensive understanding of genetic diversity and pathogen evolution, supporting the development of advanced diagnostic methodologies [14]. Collectively, findings from this study demonstrate that integrating bioinformatic analyses with molecular diagnostics and nanosensors establishes a robust framework for rapid, accurate, and species-specific pathogen detection, addressing contemporary public health demands.

16S rRNA Sequences for Species-Specific Detection of Salmonella and Shigella
The 16S rRNA gene is a highly conserved component of bacterial genomes, interspersed with hypervariable regions. This combination of conserved and variable sequences renders the 16S rRNA gene a widely employed molecular marker for bacterial identification and phylogenetic analysis. In molecular diagnostics, 16S rRNA analysis is routinely used to identify bacteria at the genus or species level through PCR amplification and sequencing of specific gene regions [9]. For species-specific identification of Salmonella, sequencing selected regions of the 16S rRNA gene has demonstrated high accuracy in isolate classification and phylogenetic confirmation. In one study, 16S rRNA sequences from multiple Salmonella isolates closely matched reference strains, enabling their differentiation into distinct phylogenetic clusters. These results indicate that 16S rRNA-based analysis can serve as a reliable component of molecular identification protocols for Salmonella in both epidemiological investigations and diagnostic laboratories. Accurate identification of Shigella presents additional challenges, as the 16S rRNA gene in this genus shares high sequence similarity with certain Escherichia coli strains. This high degree of homology complicates species-level discrimination based solely on conserved 16S regions. However, targeting hypervariable regions of the 16S gene, or combining 16S analysis with species-specific genetic markers, can markedly improve diagnostic accuracy. Intrasystem studies indicate that primers designed to amplify selected variable regions of the 16S rRNA gene permit more precise identification of Salmonella and Shigella in multiplex PCR assays and dual- or multiplex nanosensor platforms. Overall, 16S rRNA sequencing remains a foundational technique for bacterial identification, with broad applicability across taxa and compatibility with standard PCR and sequencing workflows. For closely related evolutionary groups, such as members of Enterobacterales with high 16S sequence similarity, integration of 16S analysis with targeted gene markers or whole-genome sequencing substantially enhances the accuracy and reliability of bacterial classification.

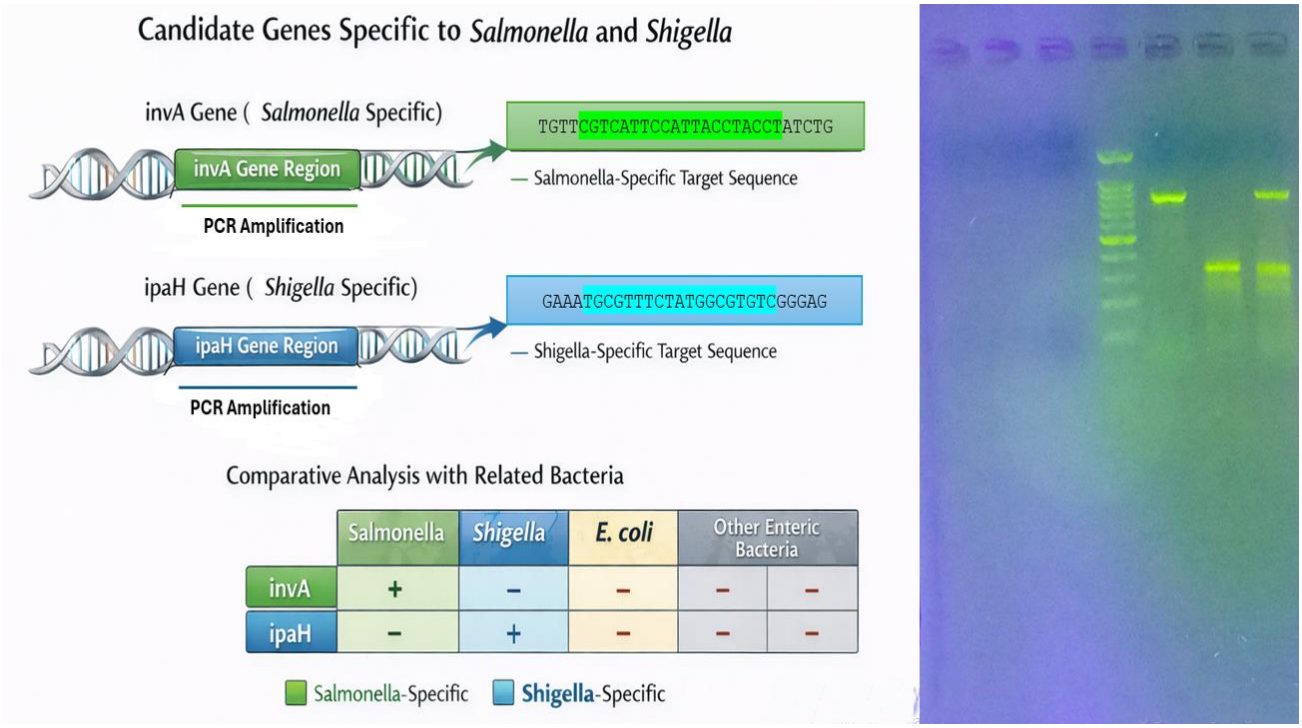

Fig. 2. *Experimental results indicated that the invA gene serves as a specific marker for Salmonella, whereas the ipaH gene is specific to Shigella. The species-specific sequences identified from these genes are well-suited for the design of primers and probes and can be employed in single, duplex, and multiplex assays, encompassing both nanotechnology-based and PCR-based detection platforms.*

CONCLUSION

This study demonstrates that comprehensive bioinformatic analysis of Salmonella and Shigella genomes enables the precise identification of species-specific sequences suitable for advanced diagnostic applications. Genomic datasets obtained from public repositories, including the National Center for Biotechnology Information (NCBI), were systematically analyzed, and comparative sequence analyses using BLAST identified conserved genes and unique genomic regions. These regions serve as optimal targets for the rational design of highly specific primers and probes compatible with both conventional nucleic acid amplification assays and nanotechnology- based biosensing platforms. The results underscore the practical utility of bioinformatic-guided strategies in the development of sensitive, rapid, and reliable diagnostic assays. By integrating genomic analyses with nanomaterial-based detection systems, this framework provides a scalable and translational approach to enhance pathogen surveillance, support disease management, and improve public health outcomes related to Salmonella and Shigella infections. Collectively, these findings establish a robust foundation for the development of next-generation molecular biosensors and highlight their significance for high-impact applications in nanoscience and nanotechnology.

REFERENCES

[1] Amare A, Eshetie S, Kasew D, Amare A, Abebe W, Moges F. (2024). Prevalence of Salmonella spp., Shigella spp., and intestinal parasites among food handlers working in University of Gondar student's cafeteria, Northwest Ethiopia. Frontiers in Public Health, vol. 12, p. 1370338.
[2] Abdelghaffar MA, Arab WS, Heikal GI, Shaltout FAE. (2025). Prevalence of multidrug resistant Shigella and Salmonella in raw chicken cuts and giblets. Benha Veterinary Medical Journal, vol. 48, no. 2, p. 62-67.
[3] Omar RF, Boissinot M, Huletsky A, Bergeron MG. (2024). Tackling infectious diseases with rapid molecular diagnosis and innovative prevention. Infectious Disease Reports, vol. 16, no. 2, p. 216-227.
[4] Kabiraz MP, Majumdar PR, Chayan Mahmud MM, Bhowmik S, Ali A. (2023). Conventional and advanced detection techniques of foodborne pathogens: A comprehensive review. Heliyon, vol. 9, p. e15482.
[5] Oon Y-L, Oon Y-S, Ayaz M, Deng M, Lu L, Kang S. (2023). Waterborne pathogens detection technologies: advances, challenges, and future perspectives. Frontiers in Microbiology, vol. 14, p. 1286923.
[6] Banerjee T, Panchal N, Sutton C, Elliott R, Patel T, Kajal K, Arogunyo E, Koti N, Santra S. (2023). Tunable magneto-plasmonic nanosensor for sensitive detection of foodborne pathogens. Biosensors, vol. 13, no. 1, p. 109.
[7] Lim YJ, Asadi Tokmedash M, Allen M, Gruich C, Choi JH, Kim C, Falkowski N, Goldsmith BR, Stringer KA, Dickson RP, Bong KW, Min J. (2026). MIDAS: rapid, multiplexed molecular profiling for integrated host–pathogen analysis. Nature Communications, vol. 17, p. 349.
[8] Ye Q, Shang Y, Chen M, Pang R, Li F, Xiang X, Wang C, Zhou B, Zhang S, Zhang J, Yang X, Xue L, Ding Y, Wu Q. (2021). Identification of novel sensitive and reliable serovar-specific targets for PCR detection of Salmonella serovars Hadar and Albany by pan-genome analysis. Frontiers in Microbiology, vol. 12, p. 605984.
[9] Moorlag SJFCM, Coolen JPM, van den Bosch B, Hui-Mei Jin E, Buil JB, Wertheim HFL, Melchers WJG. (2023). Targeting the 16S rRNA gene by reverse complement PCR next-generation sequencing: specific and sensitive detection and identification of microbes directly in clinical samples. Microbiology Spectrum, vol. 11, no. 3, p. e04483-22.
[10] Sirinthorn S, Rujipat W, Pimonwan P, Nirinya S, Alisa W, Chumpol N, Wanwisa C, Sissades T. (2025). Unveiling the impact of 16S rRNA gene intergenomic variation on primer design and gut microbiome profiling. Frontiers in Microbiology, vol. 16, p. 1573920.
[11] Haubold B, Klötzl F, Hellberg L, Thompson D, Cavalar M. (2021). Fur: Find unique genomic regions for diagnostic PCR. Bioinformatics, vol. 37, no. 15, p. 2081-2087.
[12] Lin X, Mei Y, He C, Luo Y, Yang M, Kuang Y, Ma X, Zhang H, Huang Q. (2021). Electrochemical biosensing interface based on carbon dots Fe3O4 nanomaterial for the determination of Escherichia coli O157:H7. Frontiers in Chemistry, vol. 9, p. 769648.
[13] Zhang C, Freddolino L. (2024). A large scale assessment of sequence database search tools for homology based protein function prediction. Briefings in Bioinformatics, vol. 25, no. 4, p. bbae349.
[14] Low SJ, O'Neill M, Kerry WJ, Wild N, Krysiak M, Nong Y, Azzato F, Hor E, Williams L, Taiaroa G, Steinig E, Pasricha S, Williamson DA. (2025). PathoGD: an integrative genomics approach to primer and guide RNA design for CRISPR based diagnostics. Communications Biology, vol. 8, no. 1, p. 147.