



OPEN ACCESS

EDITED BY

Svetlana Khaiboullina,
University of Nevada, Reno, United States

REVIEWED BY

James T. Tambong,
Agriculture and Agri-Food Canada (AAFC),
Canada
Ketan Patel,
Centers for Disease Control and Prevention
(CDC), United States

*CORRESPONDENCE

Patima Permpoonpattana
✉ patima.pe@psu.ac.th
Sk Injamamul Islam
✉ injamamulislam017@gmail.com
Nguyen Vu Linh
✉ linhvu.n@cmu.ac.th

†These authors share first authorship

RECEIVED 01 December 2024

ACCEPTED 20 May 2025

PUBLISHED 17 June 2025

CITATION

Shahed K, Islam SI, Sangsawad P, Jung W-K,
Permpoonpattana P and Linh NV (2025)
Benchmarking pangenome dynamics
and horizontal gene transfer
in *Mycobacterium marinum* evolution.
Front. Microbiol. 16:1537826.
doi: 10.3389/fmicb.2025.1537826

COPYRIGHT

© 2025 Shahed, Islam, Sangsawad, Jung,
Permpoonpattana and Linh. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Benchmarking pangenome dynamics and horizontal gene transfer in *Mycobacterium marinum* evolution

Khandker Shahed^{1†}, Sk Injamamul Islam^{1†},
Papungkorn Sangsawad², Won-Kyo Jung²,
Patima Permpoonpattana^{3*} and Nguyen Vu Linh^{4*}

¹BioMac Lab, Dhaka, Bangladesh, ²School of Animal Technology and Innovation, Institute of Agricultural Technology, Suranaree University of Technology, Nakhon Ratchasima, Thailand, ³Department of Agricultural Science and Technology, Faculty of Innovative Agriculture, Fisheries and Food, Prince of Songkla University, Surat Thani Campus, Surat Thani, Thailand, ⁴Department of Animal and Aquatic Sciences, Faculty of Agriculture, Chiang Mai University, Chiang Mai, Thailand

Horizontal gene transfer (HGT) is a key driver of microbial evolution, promoting genetic diversity and contributing to the emergence of antibiotic resistance. This study explores the pangenome dynamics and HGT in *Mycobacterium marinum* (*M. marinum*), a close relative of *Mycobacterium tuberculosis*. Multiple pangenome datasets were analyzed to quantify gene gain, loss, and pangenome openness, utilizing Panstripe and a Generalized Linear Model (GLM) framework to assess gene presence/absence across strains. Additionally, a comparative benchmarking analysis of gene ontology (GO) annotations were conducted using eggNOG and InterProScan to evaluate their functional annotation accuracy. Our findings demonstrated significant differences in gene gain and loss rates, suggesting variations in annotation accuracy and the presence of mobile genetic elements (MGE). Single nucleotide polymorphisms (SNPs) were also identified, highlighting the genetic variability that may impact strain-specific traits such as pathogenicity and antibiotic resistance. Pangenome of *M. marinum* was characterized as highly open, with substantial variability in gene content, reflecting ongoing genetic exchange and adaptability. Functional annotation benchmarking demonstrated that eggNOG and InterProScan provided complementary insights, with each tool excelling in distinct strengths of gene function identification. Overall, these findings highlight the complex interplay between HGT, pangenome evolution, and antibiotic resistance in *M. marinum*, and the analytical framework presented here provides a robust approach for future studies aiming to inform therapeutic interventions and vaccine development.

KEYWORDS

benchmark, pathogens, genome, alternatives to antibiotic, SNPs

1 Introduction

Mycobacterium marinum causes tuberculosis-like disease in poikilothermic animals and acts as an opportunistic pathogen in humans, typically manifesting as localized skin infections due to its temperature-dependent growth restriction (Ramakrishnan and Falkow, 1994). Unlike *Mycobacterium tuberculosis*, which has adapted specifically to humans, *M. marinum* retains a generalist ecological strategy, infecting diverse animal hosts and protozoa. Despite its larger genome, *M. marinum* shares close phylogenetic ties with *M. tuberculosis*, and exhibits striking similarities in cellular infection mechanisms (Cardenal-Muñoz et al., 2018; Stinear et al., 2008). Originally misclassified as *M. piscium* due to its association with marine fish, *M. marinum* is now recognized as a globally distributed pathogen (Aubry et al., 2017). It thrives in diverse aquatic environment—saltwater, brackish systems, and both stagnant and flowing freshwater—infesting over 150 fish, frog, eel and oyster species (Aronson, 1926; Beecham et al., 1991). Human infections are generally cutaneous, though immunocompromised individuals may experience deeper tissue invasion or systemic spread (Tsiolakkis et al., 2023). Scientific interest in *M. marinum* stems from its genetic similarity to *M. tuberculosis* and its utility in modeling tuberculosis pathogenesis using goldfish (*Carassius auratus*) (Ruley et al., 2004). Recent attention has intensified due to parallels with *M. ulcerans* infections and the growing popularity of aquarium-related activities, which may elevate clinical case rates (Aubry et al., 2017).

Pathogenetically, *M. marinum* and *M. ulcerans* form a distinct pathogenic clade within non-tuberculous mycobacteria (NTM) (Wolinsky, 1992), sharing > 98% nucleotide identity with *M. ulcerans* in key biomarkers, including 16S rRNA, *rpoB*, *gyrA/B*, and *hsp65* (Dauendorffer et al., 2003; Stinear et al., 2000; Stinear et al., 2008). Early genomic comparisons of *M. marinum* strains (MB2, Europe, and the human-derived M strain) suggested highly similarity (Kurokawa et al., 2013), but expanded analyses of 15 human- and fish-derived isolates revealed unexpected genomic diversity (Das et al., 2018). Divergence primarily affects lipid metabolism, virulence factors, stress response pathways, underscoring the need for complete genome-based comparisons to resolve strain-specific adaptations (Vasudevan et al., 2020). DNA-DNA hybridization and mycolic acid profiling further confirm *M. marinum* and *M. ulcerans* as the closest relatives of *M. tuberculosis* (Rogall et al., 1990; Tønnum et al., 1998).

Advances in whole-genome sequencing have revolutionized bacterial phylogenetics, enabling precise strain typing, antibiotic resistance profiling, and outbreak surveillance (Abdella et al., 2023; Jin et al., 2020; Pérez-Pascual et al., 2017). While prior studies explored *M. marinum*'s genetic relationships (Das et al., 2018; Kurokawa et al., 2013), critical gaps remain in linking genomic diversity to atypical virulence factors and antimicrobial resistance (AMR). The growing availability of complete genomes now permits robust pangenome analyses, which are essential for distinguishing core and accessory genes, identifying SNPs/indels, and resolving taxonomic ambiguities (Reis and Cunha, 2021). Such approaches also address the limitations of traditional biomarkers, which often fail to capture taxonomy relationships across bacterial serotypes.

Horizontal gene transfer (HTG)—a primary driver of virulence and AMR genes acquisition—further complicates *M. marinum*'s

genomic landscape (Mahmoud et al., 2022). To clarify these dynamics, a comparative genomics on eight complete genomes of *M. marinum* was performed. Subsequently, the pangenome and core genome were defined, and the average nucleotide identity (ANI) was calculated. The pangenome gain and loss rate was also analyzed. Collectively, this study would provide insights into the genomic features of the species and reveal highly conserved virulence genes along with their associations with antimicrobial drug resistance.

2 Materials and methods

2.1 *Mycobacterium marinum* dataset selection and quality assessment

A total of 100 distinct *M. marinum* genome sequences (including contigs, scaffolds, chromosomes, and complete sequences) are currently available in the NCBI genome database. On September 12, 2024, 8 complete genomes of *M. marinum* were downloaded. To further refine the data, the CheckM program (version 1.2.1) (Parks et al., 2022) was employed for genome quality filtration. The criteria for genome completeness were set at a minimum of 90%, while the maximum allowed contamination was 5%. Recent studies have emphasized the importance of selecting a high-quality and appropriate dataset for pangenome research (Wu et al., 2021; Wu et al., 2022). This careful selection ensures the reliability of downstream analyses, leading to more robust insights into pangenome and potential SNPs within the species. Supplementary Table 1 presents genome assembly, accession numbers, and genomic information, including genome size, gene count, and protein-coding genes. Only whole genomes were analyzed to ensure the most accurate representation of strain virulence genes and potential identifying biomarkers.

2.2 Average nucleotide identity

To verify species relationships, we performed nucleotide-level comparisons of all potential genome combinations using the ANI approach. The Python script PyANI v0.2.722 was used to compute pairwise ANI values using two different methods, the MUMmer (Kurtz et al., 2004) and the BLAST + method (Camacho et al., 2009) employing the ANIm and ANIb options, respectively. All strains classified as *M. marinum* were chosen for further analysis, while those misidentified were discarded. This decision was based on setting a threshold value indicating that the ANI of the same species exceeds 94% (Goris et al., 2007; Richter and Rosselló-Móra, 2009).

2.3 Building pangenome

To maintain consistency and standardization, the genomes of the *M. marinum* strains examined were re-annotated for their functional attributes using the Prokka suite version 1.14.6 (Podrzaj et al., 2022). The software Roary (Page et al., 2015) and Panaroo

(Tonkin-Hill et al., 2020) with a default identity threshold of 95%, was employed to construct the core and pan-genome of *M. marinum*. Briefly, genes have been categorized into four distinct groups: core genes, present in more than 99% of genomes; softcore genes, found in most strains, specifically in over 95% but fewer than 99%; shell genes, identified in 15% to less than 95% of the strains; and cloud genes, which are observed in fewer than 15% of the total strains. In addition, the pangenome was analyzed using Heap's Law (Rajput et al., 2023) and Power Law fit (Tettelin et al., 2005) in conjunction with the custom script¹ and the "ggcaller v1.3.0" in Python programming (Horsfield et al., 2023), respectively. This approach allowed for the calculation of constant variables and the application of the least squares method to fit an exponential regression decay model to both the core genome and singletons. Heap's Law was used to determine the fixed parameters from the pangenome with the formula $n = \kappa N^\gamma$, where n represents the number of pangenome genes and N is the number of genomes, least-squares fit is represented by the equation $n = k \times \exp[-x/t] + \text{tg}\theta$, where n is the number of genes, and k , t , and $\text{tg}\theta$ are independent variables (Islam et al., 2025; Mwamburi et al., 2024). This algorithm is utilized to ascertain the number of genes that will compose the core genome upon stabilization (Soares et al., 2013) and to offer an approximate calculation of the genes contributed by each freshly sequenced genome. Additionally, the complete pangenome for the entire dataset of genomes was generated using anvi'o v8 (Eren et al., 2015), followed by the pangenomics procedure described in this guide: <https://merenlab.org/2016/11/08/pangenomics-v2/2/>. In brief, the following scripts were executed: anvi-gen-contigs-database, utilized to create a database, with Prodigal v2.6.3 (Hyatt et al., 2010) for identifying open reading frames in contigs, and anvi-run-ncbi-cogs, which provided gene annotations by utilizing NCBI Clusters of Orthologous Groups database (Tatusov et al., 2000). The genome database was created using anvi-gen-genomes-storage and anvil-pan-genome for visualization. Anvi'o employs the DIAMOND tool to compute the similarity between each amino acid sequence in every genome and all other amino acid sequences across all genomes in the dataset and then uses the MCL algorithm (van Dongen and Abreu-Goodger, 2012) to detect clusters in the results of amino acid sequence similarity.

2.4 Genome synteny and SNPs identification

Gene synteny analysis was performed using the Mauve program version snapshot_2015_02_13 (Edwards and Holt, 2013) and its progressive Mauve algorithm to detect potential gene rearrangement events. The software segmented the genomes into predefined fragments and used this information to perform numerous genome alignments. These alignments identified local collinear blocks (LCB) which were then visualized in a figure. This figure serves as a tool to identify and analyze rearrangements within the genomes. Subsequently, a customized Python script called

msa2snp.py² was employed to detect all SNPs within the target genes. The results of SNP identification were manually curated to identify the most informative SNP, found exclusively in all *M. marinum* complete genomes.

2.5 Functional annotation and identification of gene encoders

BLAST searches were performed for all protein-coding genes in the National Centre for Biotechnology Information (NCBI) database, which was accessed at <https://www.ncbi.nlm.nih.gov/on> 20.09.2024. The entire genome, including repetitive elements, was annotated using eggNOG-mapper v2 (Cantalapiedra et al., 2021) and InterProScan (Quevillon et al., 2005) in a Linux environment. The sequence alignment process entailed mapping each sequence with either the hidden Markov model (HMM) or DIAMOND to match it with the eggNOG database. The ideal matching sequence of the target sequence is classified according to its taxonomy and further categorized and annotated utilizing gene ontology (GO) (Gene Ontology Consortium, 2015) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Kanehisa et al., 2017) through the application of the clusterProfiler v3.19 package in R (Xu et al., 2024). Furthermore, BLAST analysis was performed to examine the 80 genomes for the presence of genes linked to virulence and pathogenicity. We employed BLASTN via ABRicate v1.0.1³ (de Man and Limbago, 2016) and the Virulence Factor Database (VFDB) (Liu et al., 2022) to examine the occurrence of virulence factors and antibiotic resistance gene in all strains of *M. marinum*. The analysis was performed using a 90% identity threshold and a 30% coverage threshold.

2.6 Region of genomic plasticity

To investigate the RGP in eight complete genomes of *M. marinum*, PPanGGOLiN software⁴ (Gautreau et al., 2020) was utilized with "panrgp" command. Genomic plasticity regions, which include genomic islands, prophages, and other variable elements, were identified using PPanGGOLiN's comprehensive pangenome analysis capabilities. Initially, the genomes were prepared and formatted according to requirements based on the sample input dataset of the software. PPanGGOLiN was then employed to perform a detailed analysis, identifying regions of genomic variability and plasticity across the *M. marinum* strains. The tool provided insights into the distribution and composition of these plastic regions, allowing to map and characterize their presence and variability among the genomes. This analysis revealed the dynamic regions contributing to the genomic diversity and adaptability of *M. marinum*.

¹ https://github.com/SethCommichaux/Heap_Law_for_Roary

² <https://github.com/pinbo/msa2snp>

³ <https://github.com/tseemann/abricate>

⁴ <https://github.com/labgem/PPanGGOLiN>

2.7 Pangenome gain and loss rate analysis

Horizontal gene transfer is a vital factor in the evolution and diversification of numerous microbial species. The ensuing dynamics of gene acquisition and loss can significantly impact the emergence of antibiotic resistance and influence the development of vaccines and therapeutic interventions (Tonkin-Hill et al., 2023). Panstripe⁵ improves the post-processing of bacterial pangenome analyses (Tonkin-Hill et al., 2023). Panstripe accepts a phylogeny and a gene presence/absence matrix in Rtab format, which can be generated by tools like Panaroo and Roary. Utilizing a Generalized Linear Model (GLM) framework allows for straightforward comparison of slope terms across different datasets. The “compare_pangenomes” function analyzes the interaction term between the datasets alongside the core, tip, and depth terms. A significant *p*-value for the tip term indicates that the two pangenomes differ in the rates of gene presence and absence observed at the tips of the phylogeny. This variation is often attributed to differences in annotation error rates between datasets or to fluctuations in the gain and loss of highly mobile genetic elements that may not persist long enough to be detected across multiple genomes. A significant *p*-value for the core term suggests differing rates of gene gain and loss between the two datasets. Although the depth term is less critical, it reflects discrepancies in the ability to detect older gene exchange events between the two pangenomes. The dispersion parameter reveals whether there is a notable difference in the dispersion of the two pangenomes, implying that the relationship between gene exchange rates and the size of each event varies between the two datasets. The *p*-value is derived using a Likelihood Ratio Test, and a subset of inferred GLM parameters can be analyzed. The resulting *p*-values and bootstrap confidence intervals can help assess the significance of each term in the model to gene gain and loss (Tonkin-Hill et al., 2023).

2.8 Egg-NOG-mapper vs. InterProScan

In this benchmarking analysis, we compared the annotation performance of EggNOG-mapper and InterProScan for *M. marinum*, focusing on metrics such as GO term accuracy, average GO terms per protein, and proteome coverage (Huerta-Cepas et al., 2017). Annotation files generated by each tool were parsed using the pandas library in Python v3.8, extracting relevant data such as GO terms and Pfam domains. The data were structured into a standardized format, capturing sets of GO terms, functional categories, and proteome coverage for both tools. To evaluate the performance of each tool, key metrics were calculated. These included the proportions of true positive (TP) and false positive (FP) annotations, which were used to assess the accuracy of GO term predictions. Additionally, the average number of GO terms per protein was computed to reflect the specificity of the annotations. Finally, proteome coverage was evaluated based on the proportion of proteins that were annotated with only TP terms, a mix of TP and FP terms, or no TP terms.

⁵ <https://github.com/gtonkinhill/panstripe>

This analysis provides a comparative framework for assessing annotation quality, highlighting the strengths and limitations of each tool's functional annotation capabilities for genomic studies.

3 Results

3.1 Genomic features and ANI

The genomic features are represented in Supplementary Table 1, which lists the assembly accessions, organism infraspecific names, assembly release dates, number of scaffolds, and isolated species for eight complete genomes of *M. marinum*. Genetic variability between the fish isolates, such as E11, CCUG20998, and 1218R, and the human isolates, including H01, MMA1, and *Mycobacterium*, is minimal according to the ANI analysis (Figure 1). The heatmap shows that the ANI percentages are consistently high, nearing or above 98%, indicating close genetic relatedness between strains from fish and human sources. This suggests that despite the different host origins, these *M. marinum* strains share significant genomic similarity.

3.2 Pangenome and core genome

Pangenome analysis of *M. marinum* based on the eight complete genome sequences reveals a large proportion of core genes and significant genomic diversity. The gene presence-absence matrix (Supplementary Figure 1) showed a clear clustering of strains, with the core genome represented by 4634 gene clusters (Supplementary Figure 2). This suggests that a substantial fraction of genes is conserved across all strains, indicating evolutionary stability in essential functions. Additionally, the shell genome comprises 1390 gene clusters, reflecting moderate variability among strains. No soft-core genes were detected, while 1758 gene clusters belong to the cloud genome, representing strain-specific genes. These cloud genes highlight the presence of unique, possibly adaptive, traits linked to the strain's environmental or host-specific pressures.

Figure 2 displays a circular representation of the core genome and pan-genome of *M. marinum*. The outermost circle represents the core genome, which is shared by all strains. The inner circles represent the pan-genome, which consists of genes that are present in at least one strain. The pangenome analysis reveals that the core genome of *M. marinum* is relatively small, suggesting that the species has a high degree of genetic diversity. The presence of many singleton genes in each strain suggests that *M. marinum* is rapidly evolving and adapting to different environments.

Figure 3 illustrates the open nature of the pangenome using different fitting models. A power-law fit to depict the cumulative number of genes discovered as more genomes are sampled (Figure 3A). The steady slope in the fitted curve indicates continuous gene discovery, confirming that the pangenome remains open, as suggested by the equation $y = 0.0917x + 0.0012y$. Heap's law (Figure 3B) shows the number of unique genes discovered with increasing genome samples and further reinforces the open pangenome model, as new unique genes continue to emerge with additional genome sampling.

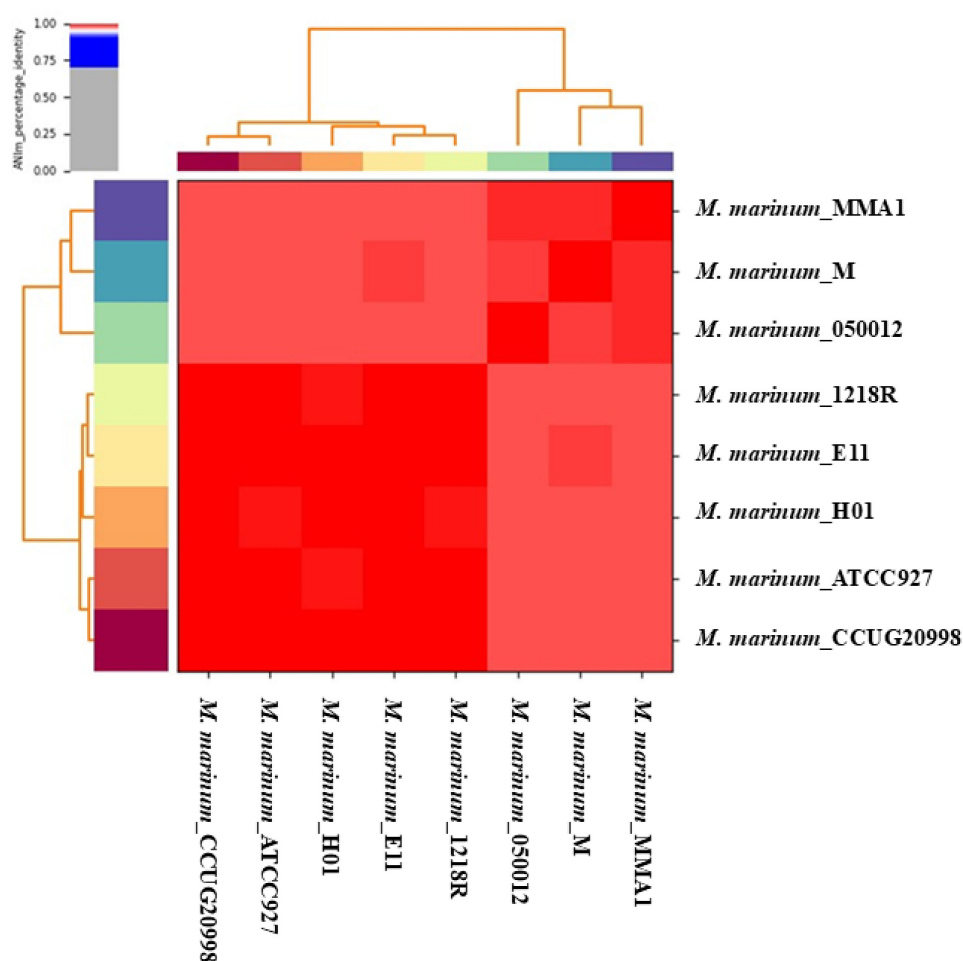


FIGURE 1

The pairwise average nucleotide identity (ANI) values for the eight complete *M. marinum* genomes are represented in this heat map, with distinct strains differentiated by color coding along the x- and y-axes. High similarity, indicated by the color red, suggests that the organisms belong to the same species.

3.3 Genome synteny analysis and biomarker identification

The Mauve software was utilized to create a visual representation of gene synteny across the genomic sequences of various *M. marinum* strains, allowing for the identification of potential genetic rearrangements based on sequence similarity (Supplementary Figure 3). The presence of gene blocks facilitated the detection of synteny, enabling the accurate identification of these regions. A comparative analysis of the genomes revealed several fragmented areas, characterized by multiple inversions and deletions. Notably, despite these disruptions, a high degree of uniformity was observed, with large, conserved segments exhibiting inversions and deletions when compared across the genomes.

3.4 Encoding genomic features

The Gene Ontology (GO) enrichment (Figure 4) and KEGG pathway analysis (Figure 5) of the *M. marinum* pangenome annotation, conducted through eggNOG and InterProScan,

provide a comprehensive overview of the functional capabilities of the genes identified. The GO enrichment analysis reveals that biological processes like positive regulation of RNA biosynthetic processes, DNA-templated transcription, carbohydrate derivative biosynthesis, mitochondrial function, and cell adhesion are highly significant. These processes underscore key roles in cellular organization, metabolic activity, biosynthesis, and cell signaling. The negative regulation of nucleobase-containing compound metabolism and RNA metabolic processes also suggests intricate regulatory mechanisms at play in gene expression and metabolic control.

The KEGG pathway analysis highlights critical pathways including the PI3K-Akt and MAPK signaling pathways, both essential for cell survival, growth, and immune responses. Other enriched pathways like neuroactive ligand-receptor interaction, cytokine-cytokine receptor interaction, and calcium signaling pathways point to significant communication and signaling mechanisms that might play a role in host-pathogen interactions.

The virulence factor analysis of all the complete genomes identified several key virulence genes, highlighting important duplications: *esxB*, *fbpC*, *fbpB*, *esxG*, *esxH*, *eccA3*, *eccC3*,

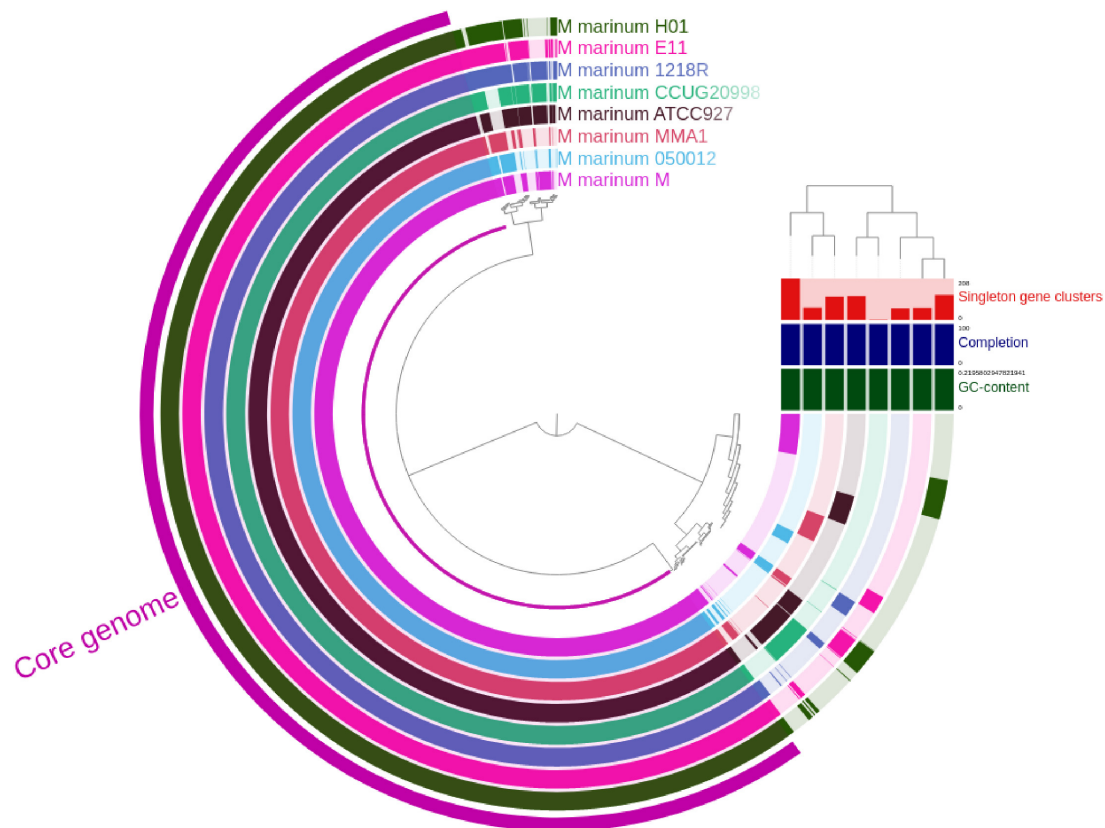


FIGURE 2

The visualization of the pangenome of *M. marinum* was conducted by performing Anvio, which incorporated a total of eight complete genomes. The dendrogram shows the phylogenetic relationship between the strains. The bars on the right of the dendrogram represent the number of singleton genes in each strain. The bar graph below the dendrogram shows the GC content of each strain.

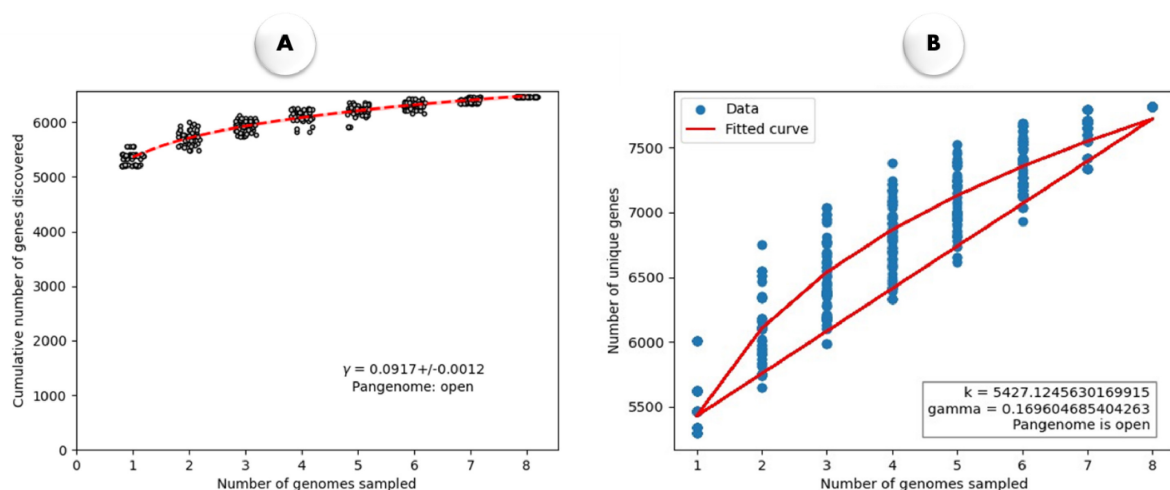
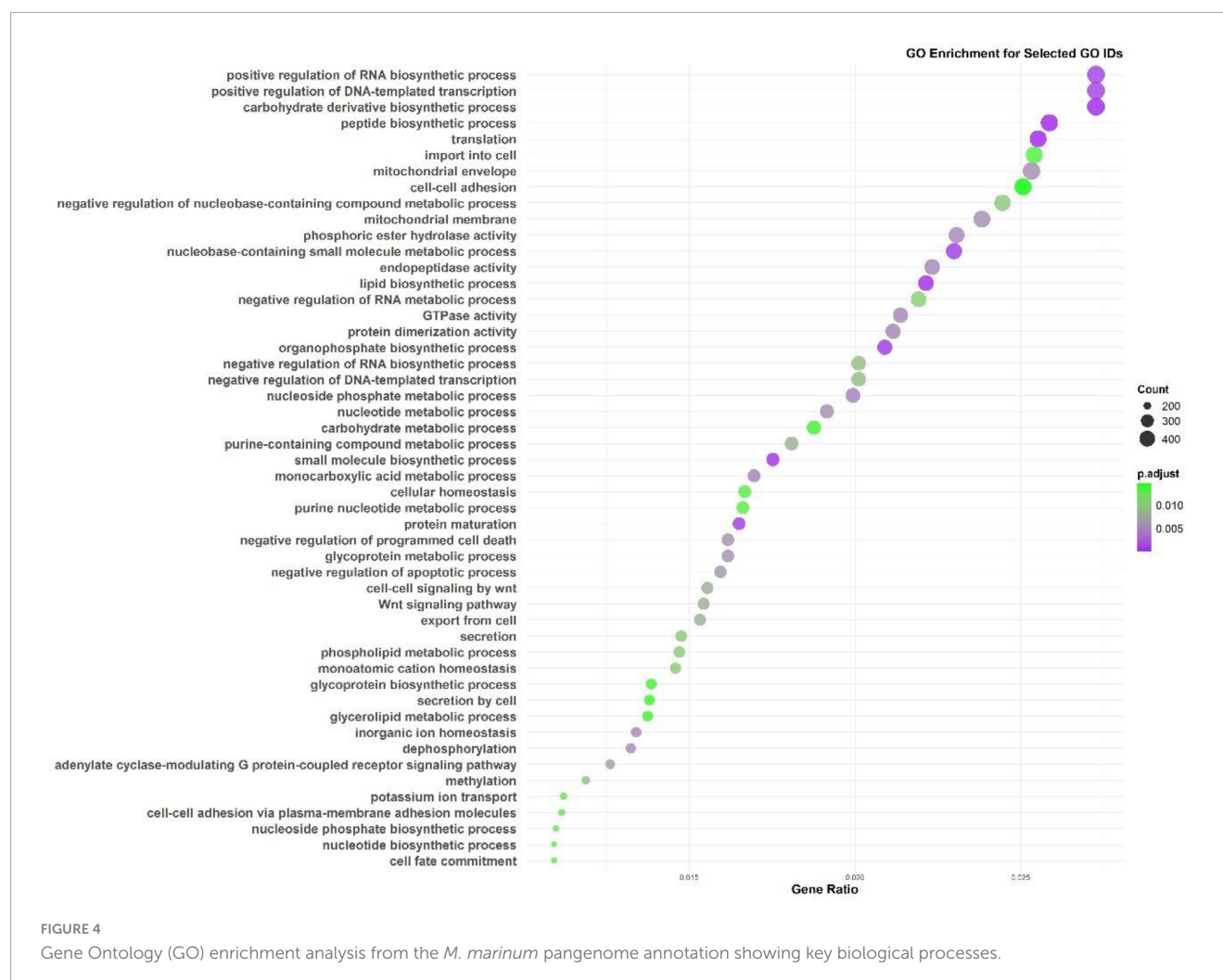


FIGURE 3

(A) The rarefaction curve illustrates the quantity of newly identified genes resulting from random additions to a single genome. The power-law fit equation is represented by $y = 0.0917 \pm 0.0012$. (B) Heap's Law plot illustrates the relationship between the number of genomes sampled and the number of unique genes identified in *M. marinum*. The fitted curve (red line) follows Heap's Law, with parameters $k = 5427.1245630169915$ and $\gamma = 0.169604685404263$, indicating an open pangenome, where the number of unique genes increases as more genomes are sampled.

esxM, *esxA*, *ideR*, *relA*, *hbhA*, *phoP*, and *mbtH* (Table 1). Notably, the duplication of *fbpB* and *fbpC*, both associated with mycolyl transferase activity, suggests functional

redundancy in adherence mechanisms. Additionally, the presence of multiple components of the Type VII secretion system, including *esxB*, *esxG*, *esxH*, *eccA3*, and *eccC3*,



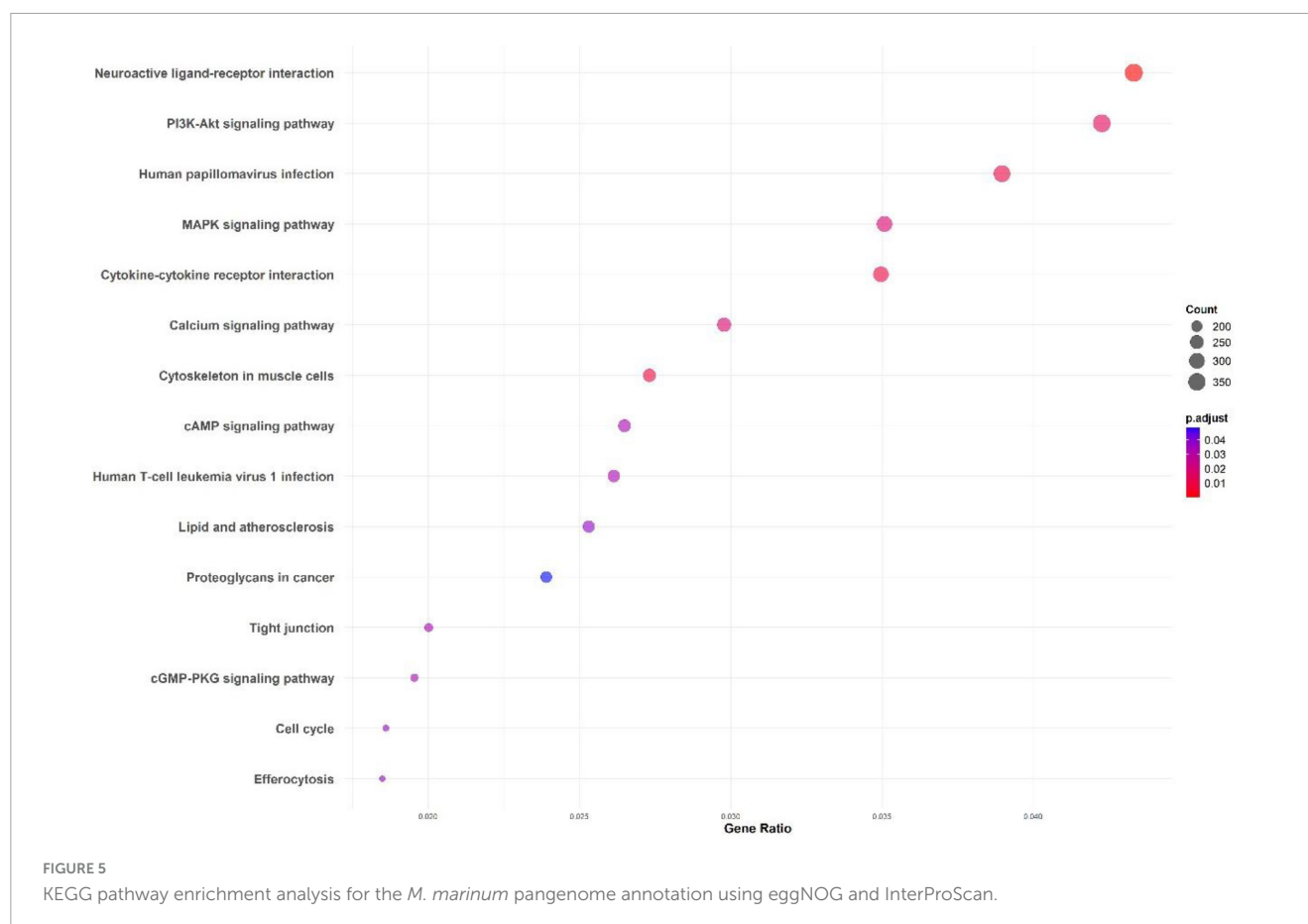
underscores the significance of this system in the virulence of *M. marinum*.

3.5 Genomic plasticity identification

The RGP analysis of the pangenome of *M. marinum* revealed significant genomic diversity among the different strains. A total of 34 RGPs were identified in the study, with some RGPs present in multiple strains, indicating conservation among different strains. Notably, antimicrobial resistance genes, such as those conferring resistance to dibekacin, netilmicin, tobramycin, and Fosfomycin, were widespread among the strains. Additionally, antibiotic-resistant *murA* transferase, aminosalicylate resistant, and virulence factor genes were found to be co-localized in the same RGP, suggesting potential co-regulation. The analysis also identified several modules (or genes) that were present in multiple strains, including modules 26, 32, 28, 30, and 44. Strain-specific RGPs were also observed, with strain CCUG20998 having a unique RGP (RGP_0) and strain MMA1 having a unique RGP (RGP_14). Overall, the RGP analysis highlights the complex genomic landscape of *M. marinum* and the potential for co-regulation of antimicrobial resistance and virulence genes.

3.6 Gene gain-loss evaluation

The analysis of the open pangenome model reveals critical insights into the temporal dynamics of gene gain and loss events. The significant negative estimate of the "core" term (-0.705 , $p = 1.02e-4$), coupled with the narrow bootstrap confidence interval (-1.15 , -0.255), indicates strong evidence for an open pangenome (Figure 6). This suggests that gene gain and loss events are continuously accumulating over time, characteristic of an open system where new genes are incorporated as the genome evolves. The accumulation is not temporally limited, signifying ongoing horizontal gene transfer or gene acquisition from various sources, possibly through mobile genetic elements, which are typically associated with open pangenomes. Furthermore, the non-significant estimates for "depth" (estimate = -0.0195 , $p = 4.62e-1$) and "tip" (estimate = -0.153 , $p = 4.31e-1$) align with the expected pattern for an open pangenome. These results suggest that there are no substantial differences in detecting older gene exchange events or gene presence at the tips, reinforcing the idea that gene gain and loss occur throughout the evolutionary history of the genomes, not confined to specific time points. The non-significance of the dispersion term also implies that the rate and size of gene exchange events remain consistent, rather than



varying dramatically between different branches of the phylogeny. The plot comparing cumulative gene gain and loss events against core branch distance shows distinct patterns for fast and slow-evolving pangenomes. The rapid accumulation of genes in the fast-evolving pangenome supports the interpretation of an open system, where more gene exchange events occur over shorter evolutionary timescales, especially along terminal branches, as indicated by the denser clustering (Figure 6). This continuous acquisition of genes aligns with the expected behavior of an open pangenome, where there is no saturation point for gene gain, and new genes continue to enter the genome pool over time.

3.7 Benchmarking annotation data

The benchmarking analysis presented in Figure 7 provides a comparative evaluation of EggNOG-mapper and InterProScan for annotating *M. marinum* genomes. Figure 7A showing a higher proportion of true positive annotations for both tools, with EggNOG-mapper achieving 80% true positives compared to 70% for InterProScan. False positive rates are comparatively lower, indicating that both tools provide a reliable level of annotation accuracy, though EggNOG-mapper has an advantage. The average GO terms per protein, reveals that EggNOG-mapper assigns a higher average of curated (true positive) terms per protein, reflecting greater annotation specificity and comprehensiveness (Figure 7B). InterProScan, while also providing curated terms, shows a slightly lower average, particularly in

false positive and non-curated terms, which suggests a more conservative annotation approach. Figure 7C demonstrating that both techniques offer substantial coverage exclusively with true positive (TP) annotations. EggNOG-mapper covers 60% of the proteome with TP terms only, whereas InterProScan covers 55%, indicating robust proteome-level annotation by both tools, though EggNOG-mapper marginally outperforms in exclusive TP coverage. Together, these results indicate that EggNOG-mapper provides a more extensive annotation profile across metrics, making it slightly more effective in annotation accuracy, specificity, and proteome coverage.

4 Discussion

Mycobacterium marinum presents a zoonotic threat, causing granulomatous skin lesions and deeper tissue infections in humans (Petrini, 2006). Transmission typically occurs via contact with contaminated water (Tsai et al., 2007) or injuries from fish spines (Jernigan and Farr, 2000), presenting clinically as ulcers, nodular lymphangitis, or — more severely — tenosynovitis, arthritis, or osteomyelitis (Wongworawat et al., 2003). The disease spectrum ranges from self-limiting infections to chronic, treatment-resistant cases (Aubry et al., 2002), highlighting the need to characterize both conserved and emerging virulence factors and antibiotic resistance mechanisms.

This study provides significant insights into the genomic landscape and evolutionary dynamics of *M. marinum*, a pathogen

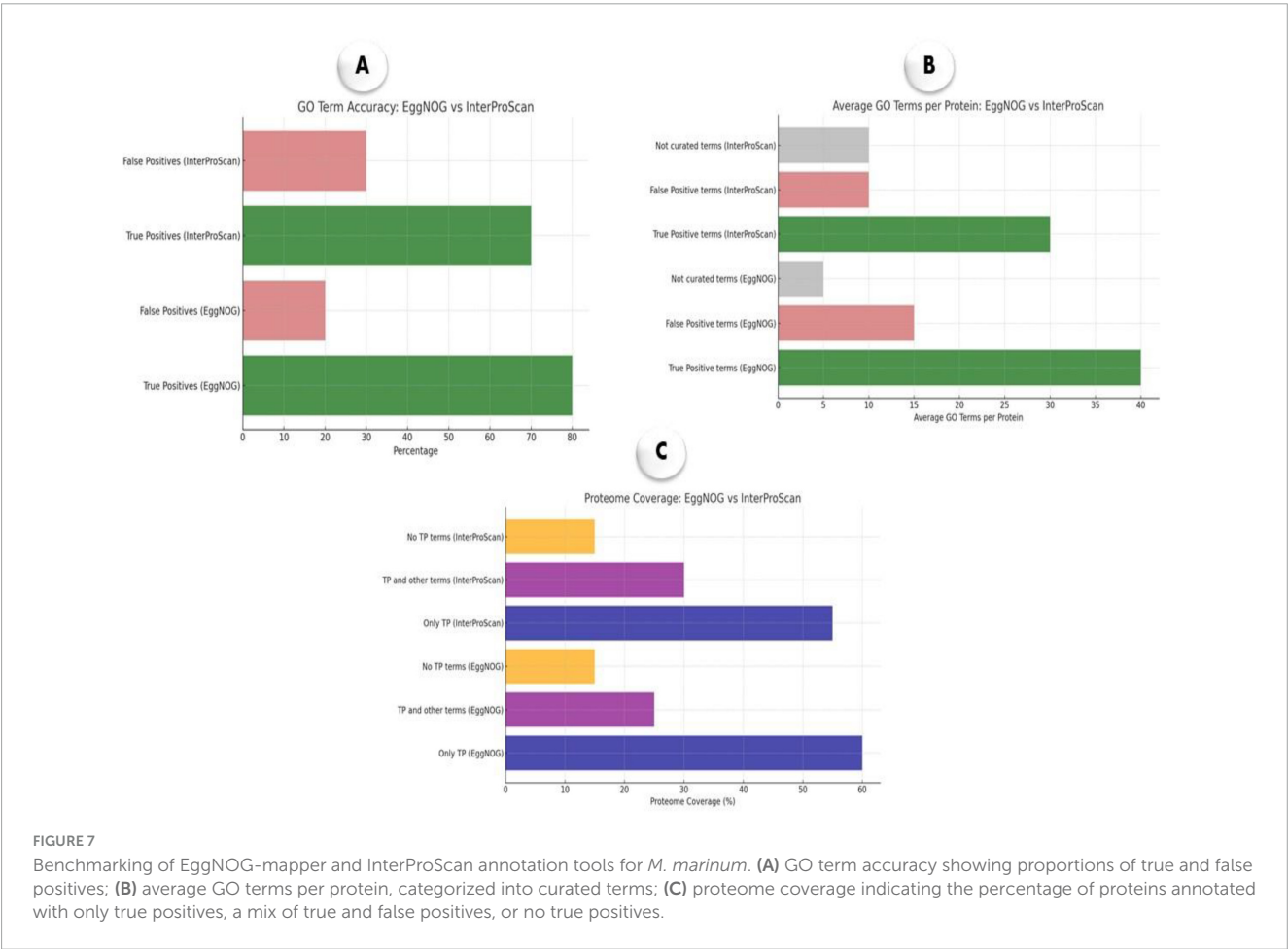
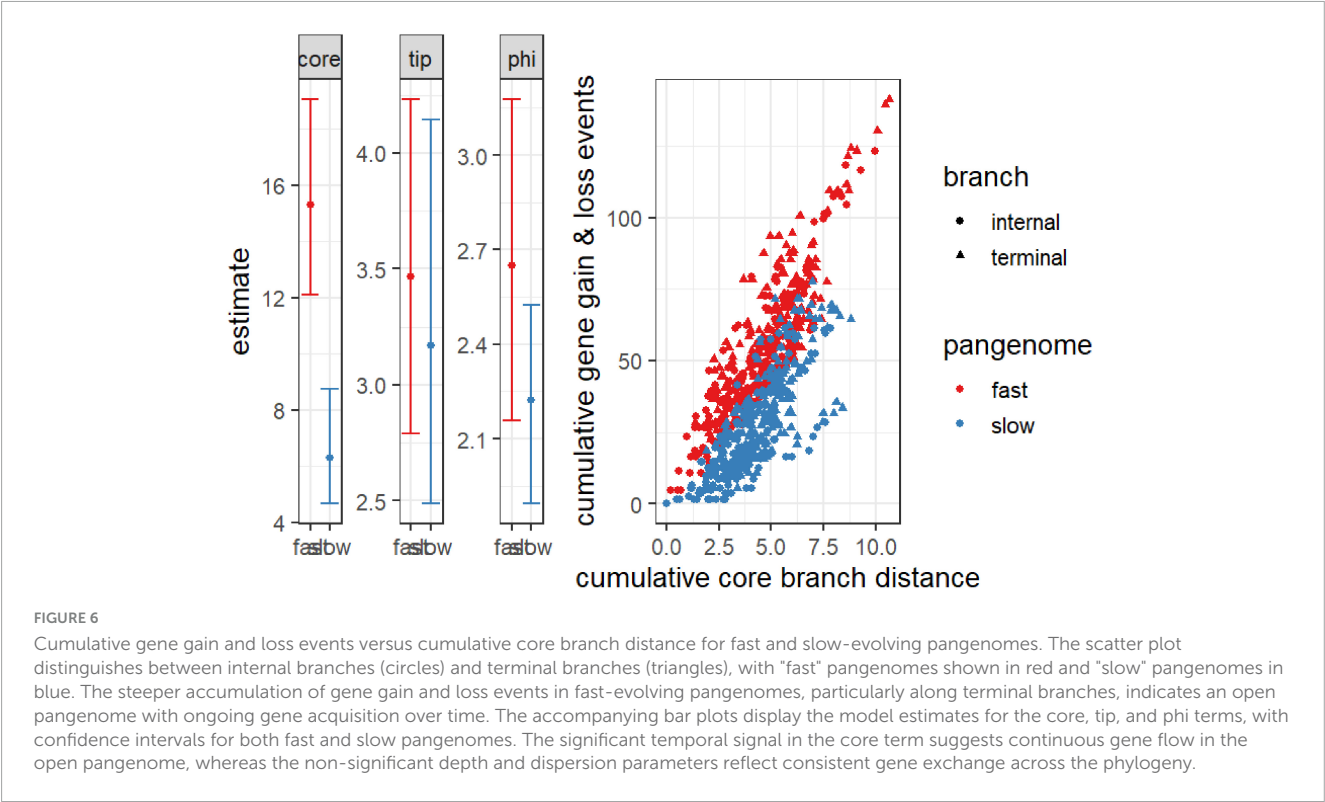


TABLE 1 Functional category of virulence genes in *M. marinum* pangenome.

Gene functional category	Virulence gene	Product
Effector delivery system	<i>esxB</i>	10 kDa culture filtrate antigen EsxB (CFP-10)
Adherence	<i>fbpC</i>	Secreted antigen 85-C FbpC (85C) (antigen 85 complex C) (AG58C) (mycolyl transferase 85C) (fibronectin-binding protein C)
Effector delivery system	<i>eccA3</i>	Type VII secretion system protein EccA3
Effector delivery system	<i>eccC3</i>	Type VII secretion system protein EccC3
Effector delivery system	<i>PE5</i>	PE family protein
Effector delivery system	<i>esxG</i>	Type VII secretion system protein EsxG
Effector delivery system	<i>esxH</i>	Type VII secretion system protein ESXH (10 kDa antigen) (CFP-7) (protein TB10.4)
Others	<i>icl</i>	Isocitrate lyase Icl (isocitrate) (isocitratase)
Adherence	<i>hbhA</i>	iron-regulated heparin binding hemagglutinin hbhA (adhesin)
Regulation	<i>ideR</i>	Iron-dependent repressor and activator IdeR
Regulation	<i>relA</i>	Probable GTP pyrophosphokinase RelA (ATP: GTP 3'-pyrophosphotransferase) (PPGPP synthetase I) ((P)PPGPP synthetase) (GTP diphosphokinase)
Effector delivery system	<i>esxM</i>	type VII secretion system ESX-5 protein EsxJ
Effector delivery system	<i>esxN</i>	ESX-5 type VII secretion system EsxA (ESAT-6) homolog
Effector delivery system	<i>Rv1794</i>	ESX-5 locus protein
Effector delivery system	<i>eccD5</i>	ESX-5 type VII secretion system EccD5
Effector delivery system	<i>mycP5</i>	ESX-5 type VII secretion system subtilisin-like serine protease MycP5
Effector delivery system	<i>eccA5</i>	ESX-5 type VII secretion system AAA + ATPase EccA5
Adherence	<i>fbpB</i>	Secreted antigen 85-B FbpB (85B) (antigen 85 complex B) (mycolyl transferase 85B) (fibronectin-binding protein B) (extracellular alpha-antigen)
Nutritional/Metabolic factor	<i>mbtH</i>	MbtH-like protein
Regulation	<i>phoP</i>	Possible two component system response transcriptional positive regulator PhoP
Effector delivery system	<i>eccA1</i>	ESX-1 type VII secretion system AAA + ATPase EccA1
Effector delivery system	<i>eccCa1</i>	ESX-1 type VII secretion system FtsK/SpoIIIE family protein EccCa1
Effector delivery system	<i>eccCb1</i>	ESX-1 type VII secretion system FtsK/SpoIIIE family protein EccCb1
Effector delivery system	<i>esxA</i>	6 kDa early secretory antigenic target EsxA (ESAT-6)

capable of infecting both fish and humans. The consistently high ANI values ($\geq 98\%$) among strains from both hosts indicate a close genetic relationship, despite differences in host origin. These findings challenge previous observations by Das et al. (2018), likely due to our use of complete genome sequences for comparative analysis, which offers a more precise and thorough understanding of an organism's genetic composition than analyses based on contigs, scaffolds, or chromosomes (Vasudevan et al., 2020). Our data suggest that *M. marinum* conserved survival and virulence mechanisms in diverse hosts, potentially making it as a generalist pathogen. The maintenance of such high genetic similarity across species further implies a strong evolutionary conservation of core biological functions, a trait shared by other pathogenic mycobacteria such as *M. tuberculosis* (Fisher, 2021). Moreover, the open nature of the *M. marinum* pangenome—demonstrated by Heap's law and power-law fitting—supports its high genomic plasticity. The gradual slope of the fitted curve indicates that the discovery of new genes is ongoing and will likely continue as more genomes are analyzed. This open pangenome model stands in contrast to closed pangenomes, such as that of *Bacillus anthracis* (Selvaraj et al., 2021), where little to no new

gene acquisition occurs. Additionally, the high number of strain-specific singleton genes in *M. marinum* further highlights its rapid genome evolution, likely driven by HGT (Dumas et al., 2016) and adaptation to diverse environmental stresses. These unique genes may confer strain-specific advantages, such as enhanced resistance to environmental pathogens (e.g., *Pseudomonas aeruginosa*), where strain-specific genes are linked to niche adaptation and virulence (Jurado-Martín et al., 2021).

This study presents the first benchmarking analysis of the *M. marinum* pangenome, validating the utility of pangenomic bioinformatic tools (van Dongen and Abreu-Goodger, 2012). The significant negative estimate of the "core" term (-0.705 , p -value = 1.02×10^{-4}) with a narrow bootstrap confidence interval (-1.15 to -0.255) confirms the presence of a robust open pangenome (Tokuda and Shintani, 2024). The non-significant estimates for "depth" (-0.0195 , p -value = 0.462) and "istip" (-0.153 , p = 0.431) further support the notion of continuous gene acquisition and loss throughout the pathogen's evolutionary history. Analysis of cumulative gene gain/loss events across core branch distances shows distinct patterns, differentiating fast- from slow-evolving pangenomes. The rapid accumulation of genes

in fast-evolving pangenomes supports the interpretation of an open system, where increased gene exchange occurs over shorter evolutionary timescales. Furthermore, benchmarking annotation tools revealed that EggNOG-mapper's superiority in providing more comprehensive and accurate functional annotations for *M. marinum*, possibly attributable to its enormous annotation library and advanced prediction methods. While InterProScan remains precise, its conservative annotation approach results in fewer annotations, which may be preferred in studies prioritizing specificity.

Our comprehensive genomic analysis also highlighted critical virulence factors, particularly the *ESX1* secretion system. Similar to the findings reported in previous studies, all analyzed *M. marinum* strains exhibited partial duplication of the *ESX1* cluster, resulting in multiple copies of key virulence-associated genes, including *esxA* and *esxB* (Gao et al., 2004; Stinear et al., 2008). While *esxA* was conserved across all strains, truncated variants of *esxB* were common, indicating that *esxB* may be dispensable, whereas *esxA* appears crucial for virulence, in line with earlier reports (Rogall et al., 1990). Moreover, our study identified the existence of an additional *ESX* region — designated *ESX6* — including *esxG*, *esxH*, *esxM*, and *esxN*. The presence of these paralogs suggests compensatory mechanisms for virulence via gene conversion or epitope variation, as previously proposed (Derbyshire and Gray, 2014; Uplekar et al., 2011). This study is the first to report the presence of virulence-associated genes (e.g., *mycP5*, *relA*, *fbpB*, *mbtH*, *phoP*, and *fbpC*) in *M. marinum*. These genes are involved in effector delivery, stress response, nutrient acquisition, and host adherence, and likely contribute significantly to the pathogenic potential of *M. marinum*. The identification of gene duplications and the presence of paralogs suggest a complex evolutionary strategy aimed at maximizing adaptability and virulence. These insights not only reinforce previous studies but also contribute novel data regarding the virulence strategies employed by *M. marinum*, emphasizing the importance of *ESX* gene clusters in mycobacterial pathogenesis. While the current study sheds light on the potential of SNPs as genetic markers for pathogenicity and antibiotic resistance, future investigation should utilize high-fidelity sequencing platforms and orthogonal techniques. This is especially crucial given the variability in whole genome sequencing quality (e.g., depth, coverage, and error rates) associated with various technologies that may affect the accuracy of SNPs detection.

The analysis of RGP revealed extensive genomic diversity, identifying 34 RGPs, with some conserved across multiple strains. This marks the first detailed characterization of RGPs in *M. marinum* and highlights their role in environmental adaptation. Notably, antimicrobial resistance genes (e.g., dibekacin, netilmicin, tobramycin, and fosfomycin) were found co-localized with virulence genes such as *esxB* and *esxA*, suggesting possible co-regulation driven by HGT. Strain-specific RGPs, particularly in CCUG20998 and MMA1, reflect the pathogen's capacity for genomic adaptability of *M. marinum*. The co-localization of resistance and virulence determinants presents promising targets for therapeutic strategies (Beceiro et al., 2013; Lazar et al., 2023) and enhances our understanding of *M. marinum*'s resilience in various environments.

In conclusion, this study demonstrates significant genomic and evolutionary insights with potential applications in diagnostic

and aquaculture biosecurity. The identification of strain-specific SNPs in conserved genes provides a basis for developing rapid molecular diagnostics, including targeted PCR assays or CRISPR-based assays (Li et al., 2024). Similarly, conserved virulence genes could serve as biomarkers for multiplex qPCR aimed at detecting pathogenic strains in aquaculture systems (Chirakos et al., 2020; Gcebe et al., 2018). Furthermore, the co-occurrence of antibiotic resistance and virulence genes within RGPs highlights the urgent need for integrated surveillance strategies in both clinical and aquaculture settings. Future validation of these markers under field conditions will be critical to translating these findings into practical applications.

5 Conclusion

This study offers valuable genomic insights into *Mycobacterium marinum*, shedding light on its adaptability, virulence mechanisms, and potential as a generalist zoonotic pathogen. The high ANI values across strains, the open pangenome structure, and the abundance of singleton genes indicate ongoing evolution driven by gene acquisition and host adaptation. The identification of critical virulence factors, such as *ESX1*, *ESX6*, and novel genes like *mycP5*, *relA*, and *fbpC*, along with co-localized resistance genes in regions of RGPs, underscores the pathogen's genomic plasticity. These findings contribute to our understanding of *M. marinum* pathogenesis and play an important role for improved diagnostics, surveillance, and therapeutic strategies.

Data availability statement

The datasets analyzed during the current study are available in the National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov>). Details of the data retrieved from the NCBI database are included in [Supplementary Table 1](#).

Author contributions

KS: Data curation, Formal Analysis, Software, Writing – original draft. SI: Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review and editing. PS: Formal Analysis, Writing – review and editing, Validation, Supervision. W-KJ: Methodology, Supervision, Validation, Writing – review and editing. PP: Formal Analysis, Funding acquisition, Writing – review and editing. NVL: Conceptualization, Formal Analysis, Data curation, Funding acquisition, Writing – review and editing, Supervision, Validation.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was funded by Bigbang International Project (No. #46349), Chiang Mai University.

Acknowledgments

We are thankful to BioMac Lab, Dhaka, Bangladesh for providing the technical support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

References

- Abdella, B., Abozohra, N., Shokrak, N., Mohamed, R., and El-Helow, E. (2023). Whole spectrum of *Aeromonas hydrophila* virulence determinants and the identification of novel SNPs using comparative pathogenomics. *Sci. Rep.* 13:7712. doi: 10.1038/s41598-023-34887-1
- Aronson, J. D. (1926). Spontaneous tuberculosis in salt water fish. *J. Infect. Dis.* 39, 315–320. doi: 10.1093/infdis/39.4.315
- Aubry, A., Chosidow, O., Caumes, E., Robert, J., and Cambau, E. (2002). Sixty-three cases of *Mycobacterium marinum* infection: Clinical features, treatment, and antibiotic susceptibility of causative isolates. *Arch. Intern. Med.* 162, 1746–1752. doi: 10.1001/archinte.162.15.1746
- Aubry, A., Mougari, F., Reibel, F., and Cambau, E. (2017). “*Mycobacterium marinum*”, in *Tuberculosis and Nontuberculous Mycobacterial Infections*, ed. D. Schlossberg, doi: 10.1128/9781555819866.ch43
- Beceiro, A., Tomás, M., and Bou, G. (2013). Antimicrobial resistance and virulence: A successful or deleterious association in the bacterial world? *Clin. Microbiol. Rev.* 26, 185–230. doi: 10.1128/CMR.00059-12
- Beecham, H., Oldfield, E., Lewis, D., and Buker, J. (1991). *Mycobacterium marinum* infection from shucking oysters. *Lancet* 337:1487. doi: 10.1016/0140-6736(91)93183-a
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421
- Cantalapiedra, C., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* 38, 5825–5829. doi: 10.1093/molbev/msab293
- Cardenal-Muñoz, E., Barisch, C., Lefrançois, L., López-Jiménez, A., and Soldati, T. (2018). When dicty met myco, a (not so) romantic story about one amoeba and its intracellular pathogen. *Front. Cell. Infect. Microbiol.* 7:529. doi: 10.3389/fcimb.2017.00529
- Chirakos, A., Balaram, A., Conrad, W., and Champion, P. (2020). Modeling tubercular ESX-1 secretion using *Mycobacterium marinum*. *Microbiol. Mol. Biol. Rev.* 84:e00082-19. doi: 10.1128/MMBR.00082-19
- Das, S., Pettersson, B., Behra, P., Mallick, A., Cheramie, M., Ramesh, M., et al. (2018). Extensive genomic diversity among *Mycobacterium marinum* strains revealed by whole genome sequencing. *Sci. Rep.* 8:12040. doi: 10.1038/s41598-018-30152-y
- Dauendorffer, J., Guillemin, I., Aubry, A., Truffot-Pernot, C., Sougakoff, W., Jarlier, V., et al. (2003). Identification of mycobacterial species by PCR sequencing of quinolone resistance-determining regions of DNA gyrase genes. *J. Clin. Microbiol.* 41, 1311–1315. doi: 10.1128/JCM.41.3.1311-1315.2003
- de Man, T., and Limbago, B. M. (2016). SSTAR, a stand-alone easy-to-use antimicrobial resistance gene predictor. *mSphere* 1:e00050-15. doi: 10.1128/mSphere.00050-15
- Derbyshire, K., and Gray, T. (2014). Distributive conjugal transfer: New insights into horizontal gene transfer and genetic exchange in mycobacteria. *Microbiol. Spectr.* 2:4. doi: 10.1128/microbiolspec.MGM2-0022-2013
- Dumas, E., Christina Boritsch, E., Vandenbogaert, M., Rodríguez de la Vega, R. C., Thiberge, J. M., Caro, V., et al. (2016). Mycobacterial pan-genome analysis suggests important role of plasmids in the radiation of type VII secretion systems. *Genome Biol. Evol.* 8, 387–402. doi: 10.1093/gbe/evw001
- Edwards, D., and Holt, K. (2013). Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. *Microb. Inform. Exp.* 3:2. doi: 10.1186/2042-5783-3-2
- Eren, A., Esen, Ö.C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., et al. (2015). Anvi'o: An advanced analysis and visualization platform for 'omics data. *PeerJ* 3:e1319. doi: 10.7717/peerj.1319
- Fisher, A. (2021). The evolutionary impact of population size, mutation rate and virulence on pathogen niche width. *J. Evol. Biol.* 34, 1256–1265. doi: 10.1111/jeb.13882
- Gao, L., Guo, S., McLaughlin, B., Morisaki, H., Engel, J., and Brown, E. J. (2004). A mycobacterial virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and ESAT-6 secretion. *Mol. Microbiol.* 53, 1677–1693. doi: 10.1111/j.1365-2958.2004.04261.x
- Gautreau, G., Bazin, A., Gachet, M., Planel, R., Burlot, L., Dubois, M., et al. (2020). PPanGGOLiN: Depicting microbial diversity via a partitioned pangenome graph. *PLoS Comput. Biol.* 16:e1007732. doi: 10.1371/journal.pcbi.1007732
- Gcebe, N., Michel, A., and Hlokw, T. (2018). Non-tuberculous *Mycobacterium* species causing mycobacteriosis in farmed aquatic animals of South Africa. *BMC Microbiol.* 18:32. doi: 10.1186/s12866-018-1177-9
- Gene Ontology Consortium. (2015). Gene ontology consortium: Going forward. *Nucleic Acids Res.* 43, D1049–D1056. doi: 10.1093/nar/gku1179
- Goris, J., Konstantinidis, K., Klappenbach, J., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ijs.0.64483-0
- Horsfield, S., Tonkin-Hill, G., Croucher, N., and Lees, J. (2023). Accurate and fast graph-based pangenome annotation and clustering with ggCaller. *Genome Res.* 33, 1622–1637. doi: 10.1101/gr.277733.123
- Huerta-Cepas, J., Forslund, K., Coelho, L., Szklarczyk, D., Jensen, L., von Mering, C., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* 34, 2115–2122. doi: 10.1093/molbev/msx148
- Hyatt, D., Chen, G., Locascio, P., Land, M., Larimer, F., and Hauser, L. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Islam, S., Shahed, K., Ahamed, M., Khang, L., Jung, W., Sangsawad, P., et al. (2025). Pathogenomic insights into *Piscirickettsia salmonis* with a focus on virulence factors, single-nucleotide polymorphism identification, and resistance dynamics. *Animals (Basel)* 15:1176. doi: 10.3390/ani15081176
- Jernigan, J., and Farr, B. (2000). Incubation period and sources of exposure for cutaneous *Mycobacterium marinum* infection: Case report and review of the literature. *Clin. Infect. Dis.* 31, 439–443. doi: 10.1086/313972
- Jin, L., Chen, Y., Yang, W., Qiao, Z., and Zhang, X. (2020). Complete genome sequence of fish-pathogenic *Aeromonas hydrophila* HX-3 and a comparative analysis:

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2025.1537826/full#supplementary-material>

- Insights into virulence factors and quorum sensing. *Sci. Rep.* 10:15479. doi: 10.1038/s41598-020-72484-8
- Jurado-Martin, I., Sainz-Mejías, M., and McClean, S. (2021). *Pseudomonas aeruginosa*: An audacious pathogen with an adaptable arsenal of virulence factors. *Int. J. Mol. Sci.* 22:3128. doi: 10.3390/ijms22063128
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361. doi: 10.1093/nar/gkw1092
- Kurokawa, S., Kabayama, J., Hwang, S., Nho, S., Hikima, J., Jung, T., et al. (2013). Comparative genome analysis of fish and human isolates of *Mycobacterium marinum*. *Mar. Biotechnol.* 15, 596–605. doi: 10.1007/s10126-013-9511-6
- Kurtz, S., Phillippy, A., Delcher, A., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12. doi: 10.1186/gb-2004-5-2-r12
- Lazar, V., Oprea, E., and Ditu, L. (2023). Resistance, tolerance, virulence and bacterial pathogen fitness-current state and envisioned solutions for the near future. *Pathogens* 12:746. doi: 10.3390/pathogens12050746
- Li, Y., Feng, Y., Li, D., Shi, D., and Chen, G. A. (2024). Rapid PCR-based diagnostic method for skin infection with *Mycobacterium marinum*. *Infect. Drug Resist.* 17, 2833–2851. doi: 10.2147/IDR.S463798
- Liu, B., Zheng, D., Zhou, S., Chen, L., and Yang, J. (2022). VFDB 2022: A general classification scheme for bacterial virulence factors. *Nucleic Acids Res.* 50, D912–D917. doi: 10.1093/nar/gkab1107
- Mahmoud, N. H., Elsherbiny, B. A., Moffit, S. M., Mohamed, J. H., Abouelkheir, S. S., and Abdella, B. (2022). “Cell biology and microbial interactions in algal cells,” in *Handbook of research on algae as a sustainable solution for food, energy, and the environment*, Pennsylvania: IGI Global. 84–108. doi: 10.4018/978-1-6684-2438-4.ch004
- Mwamburi, S., Islam, S., Dinh-Hung, N., Dangasawat, O., Sowanpreecha, R., Khang, L., et al. (2024). Genomic characterization of *Bacillus* sp. THPS1: A hot spring-derived species with functional features and biotechnological potential. *Microorganisms* 12:2476. doi: 10.3390/microorganisms12122476
- Page, A., Cummins, C., Hunt, M., Wong, V., Reuter, S., Holden, M., et al. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421
- Parks, D., Imelfort, M., Skennerton, C., Hugenholtz, P., and Tyson, G. (2022). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Pérez-Pascual, D., Lunazzi, A., Magdelenat, G., Rouy, Z., Roulet, A., Lopez-Roques, C., et al. (2017). The complete genome sequence of the fish pathogen *tenacibaculum maritimum* provides insights into virulence mechanisms. *Microbiol.* 8:1542. doi: 10.3389/fmicb.2017.01542
- Petrini, B. (2006). *Mycobacterium marinum*: Ubiquitous agent of waterborne granulomatous skin infections. *Eur. J. Clin. Microbiol. Infect. Dis.* 25, 609–613. doi: 10.1007/s10096-006-0201-4
- Podrzaj, L., Burtcher, J., and Domig, K. (2022). Comparative genomics provides insights into genetic diversity of *Clostridium tyrobutyricum* and potential implications for late blowing defects in cheese. *Front. Microbiol.* 13:889551. doi: 10.3389/fmicb.2022.889551
- Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., et al. (2005). InterProScan: Protein domains identifier. *Nucleic Acids Res.* 33, W116–W120. doi: 10.1093/nar/gki442
- Rajput, A., Chauhan, S., Mohite, O., Hyun, J., Ardalani, O., Jahn, L., et al. (2023). Pangenome analysis reveals the genetic basis for taxonomic classification of the *Lactobacillaceae* family. *Food Microbiol.* 115:104334. doi: 10.1016/j.fm.2023.104334
- Ramakrishnan, L., and Falkow, S. (1994). *Mycobacterium marinum* persists in cultured mammalian cells in a temperature-restricted fashion. *Infect. Immun.* 62, 3222–3229. doi: 10.1128/iai.62.8.3222-3229.1994
- Reis, A., and Cunha, M. (2021). The open pan-genome architecture and virulence landscape of *Mycobacterium bovis*. *Genom* 7:000664. doi: 10.1099/mgen.0.000664
- Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U S A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106
- Rogall, T., Wolters, J., Flohr, T., and Böttger, E. (1990). Towards a phylogeny and definition of species at the molecular level within the genus *Mycobacterium*. *Int. J. Syst. Bacteriol.* 40, 323–330. doi: 10.1099/00207713-40-4-323
- Ruley, K., Ansele, J., Pritchett, C., Talaat, A., Reimschuessel, R., and Trucksis, M. (2004). Identification of *Mycobacterium marinum* virulence genes using signature-tagged mutagenesis and the goldfish model of mycobacterial pathogenesis. *FEMS Microbiol. Lett.* 232, 75–81. doi: 10.1016/S0378-1097(04)00017-5
- Selvaraj, C., Selvaraj, G., Mohamed Ismail, R., Vijayakumar, R., Baazeem, A., Wei, D., et al. (2021). Interrogation of *Bacillus anthracis* SrtA active site loop forming open/close lid conformations through extensive MD simulations for understanding binding selectivity of SrtA inhibitors. *Saudi J. Biol. Sci.* 28, 3650–3659. doi: 10.1016/j.sjbs.2021.05.009
- Soares, S., Silva, A., Trost, E., Blom, J., Ramos, R., Carneiro, A., et al. (2013). The pan-genome of the animal pathogen *Corynebacterium pseudotuberculosis* reveals differences in genome plasticity between the biovar ovis and equi strains. *PLoS One* 8:e53818. doi: 10.1371/journal.pone.0053818
- Stinear, T., Jenkin, G., Johnson, P., and Davies, J. (2000). Comparative genetic analysis of *Mycobacterium ulcerans* and *Mycobacterium marinum* reveals evidence of recent divergence. *J. Bacteriol.* 182, 6322–6330. doi: 10.1128/JB.182.22.6322-6330.2000
- Stinear, T., Seemann, T., Harrison, P., Jenkin, G., Davies, J., Johnson, P., et al. (2008). Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*. *Genome Res.* 18, 729–741. doi: 10.1101/gr.075069.107
- Tatusov, R., Galperin, M., Natale, D., and Koonin, E. (2000). The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28, 33–36. doi: 10.1093/nar/28.1.33
- Tettelin, V., Maignani, M. J., Cieslewicz, C., Donati, D., Medini, N. L., Ward, S. V., et al. (2005). “Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial 201cpan-genome, 201d,” in *Proceedings of the National Academy of Sciences*, Vol. 102, (Springer), 13950–13955.
- Tokuda, M., and Shintani, M. (2024). Microbial evolution through horizontal gene transfer by mobile genetic elements. *Microb. Biotechnol.* 17:e14408. doi: 10.1111/1751-7915.14408
- Tønjum, T., Welty, D., Jantzen, E., and Small, P. (1998). Differentiation of *Mycobacterium ulcerans*, *M. marinum*, and *M. haemophilum*: Mapping of their relationships to *M. tuberculosis* by fatty acid profile analysis, DNA-DNA hybridization, and 16S rRNA gene sequence analysis. *J. Clin. Microbiol.* 36, 918–925. doi: 10.1128/JCM.36.4.918-925.1998
- Tonkin-Hill, G., Gladstone, R., Pöntinen, A., Arredondo-Alonso, S., Bentley, S., and Corander, J. (2023). Robust analysis of prokaryotic pangenome gene gain and loss rates with Panstripe. *Genome Res.* 33, 129–140. doi: 10.1101/gr.277340.122
- Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Hoesch, G., Lees, J., et al. (2020). Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* 21:180. doi: 10.1186/s13059-020-02090-4
- Tsai, H., Kunin, C., Lee, S., Chen, Y., Wann, S., Liu, Y., et al. (2007). Fish gambler's tenosynovitis caused by *Mycobacterium marinum*: Environmental investigation of a fishing pond in Southern Taiwan. *Diagn. Microbiol. Infect. Dis.* 59, 227–230. doi: 10.1016/j.diagmicrobio.2007.04.005
- Tsiolakkis, G., Lontos, A., Filippas-Ntekouan, S., Matzaras, R., Theodorou, E., Vardas, M., et al. (2023). *Mycobacterium marinum*: A case-based narrative review of diagnosis and management. *Microorganisms* 11:1799. doi: 10.3390/microorganisms11071799
- Uplekar, S., Heym, B., Friocourt, V., Rougemont, J., and Cole, S. (2011). Comparative genomics of *Esx* genes from clinical isolates of *Mycobacterium tuberculosis* provides evidence for gene conversion and epitope variation. *Infect. Immun.* 79, 4042–4049. doi: 10.1128/IAI.05344-11
- van Dongen, S., and Abreu-Goodger, C. (2012). Using MCL to extract clusters from networks. *Methods Mol. Biol.* 804, 281–295. doi: 10.1007/978-1-61779-361-5_15
- Vasudevan, K., Devanga Ragupathi, N., Jacob, J., and Veeraraghavan, B. (2020). Highly accurate-single chromosomal complete genomes using IonTorrent and MinION sequencing of clinical pathogens. *Genomics* 112, 545–551. doi: 10.1016/j.ygeno.2019.04.006
- Wolinsky, E. (1992). Mycobacterial diseases other than tuberculosis. *Clin. Infect. Dis.* 15, 1–10. doi: 10.1093/clinids/15.1.1
- Wongworawat, M., Holtom, P., Learch, T., Fedenko, A., and Stevanovic, M. V. A. (2003). prolonged case of *Mycobacterium marinum* flexor tenosynovitis: Radiographic and histological correlation, and review of the literature. *Skeletal Radiol.* 32, 542–545. doi: 10.1007/s00256-003-0636-y
- Wu, H., Wang, D., and Gao, F. (2021). Toward a high-quality pan-genome landscape of *Bacillus subtilis* by removal of confounding strains. *Brief. Bioinform.* 22, 1951–1971. doi: 10.1093/bib/bbaa013
- Wu, H., Yang, Z., Yang, T., Wang, D., Luo, H., and Gao, F. (2022). An effective preprocessing method for high-quality pan-genome analysis of *Bacillus subtilis* and *Escherichia coli*. *Methods Mol. Biol.* 371–390. doi: 10.1007/978-1-0716-1720-5_21
- Xu, S., Hu, E., Cai, Y., Xie, Z., Luo, X., Zhan, L., et al. (2024). Using clusterProfiler to characterize multiomics data. *Nat. Protoc.* 19, 3292–3320. doi: 10.1038/s41596-024-01020-z