中央研究院生物多樣性研究中心
Biodiversity Research Center, Academia Sinica
biodiv@gate.sinica.edu.tw
02-2789-9621

Evolutionary Genetics and Genomics

Evolutionary Genomics and Biosynthetic Potential of the Genus *Nocardia*

Mr. Kiran Kumar Eripogu
Ph. D. Candidate
伊奇朗 先生
博士候選人

TIGP Biodiversity Program, Academia Sinica
Biodiversity Research Center, Academia Sinica
Department of Life Science, National Taiwan Normal University

Time: 2025. 10. 23 Thu. 09:30

Venue: Auditorium, 1st Floor
Interdisciplinary Research Building
跨領域科技研究大樓1樓演講廳

Host: Dr. Wen-Hsiung Li 李文雄特聘研究員

~Doctoral Dissertation Defense Presentation~

中央研究院生物多樣性研究中心
Biodiversity Research Center, Academia Sinica
biodiv@gate.sinica.edu.tw
02-2789-9621

Evolutionary Genetics and Genomics

Abstract

Genome mining enables the identification and characterization of biosynthetic gene clusters (BGCs) from microbial genome sequences that encode natural products with therapeutic potential. Despite the success of genome mining in well-studied bacterial genera like *Streptomyces*, many microbial lineages remain underexplored. The genus *Nocardia*, belonging to the phylum Actinomycetota, is both ecological versatile and clinically significant, yet, its genomic architecture, evolutionary dynamics, and biosynthetic potential remain largely unexplored.

This study integrates phylogenomics, comparative pangenomics, and genome mining to assess the evolutionary and functional diversity of *Nocardia*. Phylogenomic reconstruction using 110 conserved single-copy orthologs identified six well-supported major clades. ANI-based classification highlighted the anomalies, notably *N. globerula*, which aligned more closely with the outgroup i.e., *Rhodococcus erythropolis*, suggesting a misclassification. Pangenome analysis revealed an open pangenome structure, with a compact core genome (~900 genes) and an expansive accessory genome encompassing over 205,000 species-specific genes. High genomic fluidity and elevated pairwise Jaccard dissimilarity indicated rapid gene turnover and extensive interspecies variation. Functional annotation revealed a conserved core enriched in essential cellular functions (e.g., réplication, transcription, translation), and a highly variable accessory genome enriched in genes associated with environmental adaptation, stress response, and secondary metabolism.

We identified 8,322 BGCs belonging to 46 classes, including well-known types such as type I polyketide synthases (T1PKS), non-ribosomal peptide synthetases (NRPS), and terpenes. Remarkably, over one-third of these clusters lacked similarity to any experimentally known BGCs, highlighting a rich source of novel natural products. For the first time, lanthipeptide-class-V and prodigiosin BGCs were identified in *Nocardia*, expanding the known chemical repertoire of the genus. Network analysis revealed over 4,000 gene cluster families (GCFs) and 3,220 singletons, indicating high biosynthetic modularity and specialization. Synteny analysis of key BGCs, such as those encoding Nocobactin NA, Nargenicin, and Brasilicardin A, revealed conserved core regions with structural rearrangements across species, suggesting vertical inheritance with modification. Evidence of cross-phylum gene homology and clade-specific cluster distribution further supports the role of horizontal gene transfer and genome expansion in driving biosynthetic innovation.

Collectively, this study presents the most comprehensive genomic analysis of *Nocardia* to date, integrating evolutionary, functional, and biosynthetic dimensions. It refines our understanding of *Nocardia* taxonomy, illuminates its complex evolutionary history, and establishes the genus as a promising reservoir of novel biosynthetic pathways. The findings lay the foundation for future research in natural product discovery, microbial evolution, functional genomics, and synthetic biology.