frontiers | Research Topics

Insights in evolutionary & genomic microbiology 2022

Edited by
Ernesto Perez-Rueda and Feng Gao

Published in
Frontiers in Microbiology

ISSN 1664-8714
ISBN 978-2-8325-3372-7
DOI 10.3389/978-2-8325-3372-7

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Insights in evolutionary & genomic microbiology: 2022

Topic editors

Ernesto Perez-Rueda — Universidad Nacional Autónoma de México, Mexico

Feng Gao — Tianjin University, China

Citation

Perez-Rueda, E., Gao, F., eds. (2023). *Insights in evolutionary & genomic microbiology: 2022*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-3372-7

Table of contents

04 Editorial: Insights in evolutionary & genomic microbiology: 2022
Ernesto Perez-Rueda and Feng Gao

06 Out of the Qinghai-Tibetan plateau: Origin, evolution and historical biogeography of *Morchella* (both Elata and Esculenta clades)
Qing Meng, Zhanling Xie, Hongyan Xu, Jing Guo, Yongpeng Tang, Ting Ma, Qingqing Peng, Bao Wang, Yujing Mao, Shangjin Yan, Jiabao Yang, Deyu Dong, Yingzhu Duan, Fan Zhang and Taizhen Gao

21 Early origin and evolution of the FtsZ/tubulin protein family
Carlos Santana-Molina, DMaría del Saz-Navarro and Damien P. Devos

35 Growth rate-associated transcriptome reorganization in response to genomic, environmental, and evolutionary interruptions
Yuichiro Matsui, Motoki Nagai and Bei-Wen Ying

47 The evolution of morphological development is congruent with the species phylogeny in the genus *Streptomyces*
Min Wang, Cong-Jian Li, Zhen Zhang, Pan-Pan Li, Ling-Ling Yang and Xiao-Yang Zhi

62 Whole-genome sequencing reveals high-risk clones of *Pseudomonas aeruginosa* in Guangdong, China
Yonggang Zhao, Dingqiang Chen, Boyang Ji, Xingju Zhang, Mikkel Anbo and Lars Jelsbak

72 Pan-genome wide association study of *Glaesserella parasuis* highlights genes associated with virulence and biofilm formation
You Zhou, Dike Jiang, Xueping Yao, Yan Luo, Zexiao Yang, Meishen Ren, Ge Zhang, Yuanyuan Yu, Aiping Lu and Yin Wang

94 *RpoZ* regulates 2,4-DAPG production and quorum sensing system in *Pseudomonas fluorescens* 2P24
Yarui Wei, Baozhu Dong, Xiaogang Wu, Mingmin Zhao, Dong Wang, Na Li, Qian Zhang, Liqun Zhang and Hongyou Zhou

105 Evolution of genetic architecture and gene regulation in biphenyl/PCB-degrading bacteria
Hidehiko Fujihara, Jun Hirose and Hikaru Suenaga

113 Genomic insights into *Aspergillus sydowii* 29R-4-F02: unraveling adaptive mechanisms in subseafloor coal-bearing sediment environments
Jun-Peng Jiang, Xuan Liu, Yi-Fan Liao, Jun Shan, Yu-Ping Zhu and Chang-Hong Liu

125 The fitness cost of horizontally transferred and mutational antimicrobial resistance in *Escherichia coli*
Marie Vanacker, Natacha Lenuzza and Jean-Philippe Rasigade

frontiers | Frontiers in Microbiology

TYPE Editorial
PUBLISHED 21 August 2023
DOI 10.3389/fmicb.2023.1269933

OPEN ACCESS

EDITED AND REVIEWED BY
Ludmila Chistoserdova,
University of Washington, United States

*CORRESPONDENCE
Ernesto Perez-Rueda
✉ ernesto.perez@iimas.unam.mx
Feng Gao
✉ fgao@tju.edu.cn

RECEIVED 31 July 2023
ACCEPTED 10 August 2023
PUBLISHED 21 August 2023

CITATION
Perez-Rueda E and Gao F (2023) Editorial: Insights in evolutionary & genomic microbiology: 2022.
Front. Microbiol. 14:1269933.
doi: 10.3389/fmicb.2023.1269933

Editorial: Insights in evolutionary & genomic microbiology: 2022

Ernesto Perez-Rueda[1]* and Feng Gao[2,3,4]*

[1]Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, Unidad Académica del Estado de Yucatán, Mérida, Mexico, [2]Department of Physics, School of Science, Tianjin University, Tianjin, China, [3]Frontier Science Center of Synthetic Biology (MOE), Key Laboratory of Systems Bioengineering (MOE), Tianjin University, Tianjin, China, [4]SynBio Research Platform, Collaborative Innovation Center of Chemical Science and Engineering (Tianjin), Tianjin, China

KEYWORDS

bacteria, fungi, evolution, phylogenetics, genomic epidemiology, gene regulation and expression

Editorial on the Research Topic
Insights in evolutionary & genomic microbiology: 2022

During the past decade, the field of evolutionary and genomic microbiology has witnessed significant advancements, with development of high-throughput "omics" technologies, bioinformatics, and artificial intelligence (Gao et al., 2022), shedding new light on the intricate mechanisms that shape microbial evolution and genomic diversity (Shu and Huang, 2022). As Topic Editors, Drs. Daniel Yero, Feng Gao, and Baolei Jia organized the Research Topic "*Insights in evolutionary and genomic microbiology: 2021*" for Frontiers in Microbiology (Yero et al., 2022), which has been recommended by the Chief Editors and journal team as 2022 outstanding Research Topic in terms of views and downloads. Given its previous success, this Research Topic was revisited by Dr. Ernesto Perez-Rueda and Dr. Feng Gao in 2022. After rigorous peer review, a total of 10 articles have been published in this topic, including 8 original research articles, one review, and one mini-review.

Majority of the articles focus on various studies on evolution in bacteria via the omics methods. Among the review articles, Vanacker et al. systematically reviewed the potential fitness cost of antimicrobial resistance (AMR) in *Escherichia coli* with the meta-analysis of related high-quality studies. Fujihara et al. summarized the evolutionary mechanisms regarding degradation gene systems by analyzing the genome sequences of isolated bacteria degrading xenobiotics.

Among the original research articles, Wang et al. discovered that the evolution of morphological development of *Streptomyces* is in accord with the species phylogeny by using a comparative phylogenetic approach. Through the genomic epidemiology analysis, Zhao et al. provided an overview of the prevalence of clinical multidrug-resistant *Pseudomonas aeruginosa* infections and the emergence of high-risk clones in Guangdong, China. Based on the growth rate-associated transcriptome, Matsui et al. found the occurrence of the environmental stressors remarkably decreased the growth rate of the wild-type instead of reduced-genome *Escherichia coli* when investigating whether and how bacterial growth was affected by the genomic, environmental, and evolutionary interruptions. By a pan-genome wide association study, Zhou et al. highlighted the genes associated with virulence and biofilm formation for further investigation in *Glaesserella parasuis*. Based on the comparative genomics and phylogenetic analysis, Santana-Molina et al. investigated early origin and evolution of the FtsZ/tubulin protein family, which would provide valuable insights into the diversification of the three domains of life.

Facilitated by the mutant obtained by Tn5 mutagenesis, Wei et al. found *rpoZ* was an important regulator of antibiotic 2,4-diacetylphloroglucinol (2,4-DAPG) production and quorum sensing system in *Pseudomonas fluorescens* 2P24.

In addition to the studies on bacteria, there are two original research articles about fungi in this topic. Through *de novo* sequencing, assembly, and comparative analyses, Jiang et al. found more carbohydrate-active enzymes (CAZymes) and distinct genes related to vesicular fusion and autophagy in an *Aspergillus sydowii* strain 29R-4-F02 compared to the terrestrial strain CBS593.65, which revealed the survival and environmental adaptation mechanism of this subseafloor fungus. Based on the genomic data newly generated or retrieved from GenBank, Meng et al. explored the origin, evolution, and historical biogeography of the *Morchella* fungi in the Qinghai-Tibet Plateau subkingdoms (QTPs), particularly focusing on the Elata and Esculenta clades, which provided strong evidence for the origin theory of the QTPs.

Author contributions

FG: Writing—original draft, Writing—review and editing. EP-R: Writing—original draft, Writing—review and editing.

Funding

FG was supported by the National Key Research and Development Program of China (grant number 2018YFA0903700) and the National Natural Science Foundation of China (grant numbers 32270692 and 31571358). EP-R was supported by the Dirección General de Asuntos del Personal Académico-Universidad Nacional Autónoma de México (IN-220523) and CONAHCYT (320012).

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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Gao, F., Huang, K., and Xing, Y. (2022). Artificial intelligence in omics. *Genom. Proteom. Bioinform.* 20, 811–813. doi: 10.1016/j.gpb.2023.01.002

Shu, W. S., and Huang, L. N. (2022). Microbial diversity in extreme environments. *Nat. Rev. Microbiol.* 20, 219–235. doi: 10.1038/s41579-021-00648-y

Yero, D., Jia, B., and Gao, F. (2022). Editorial: insights in evolutionary and genomic microbiology: 2021. *Front. Microbiol.* 13, 915593. doi: 10.3389/fmicb.2022.915593

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 28 December 2022
DOI 10.3389/fmicb.2022.1078663

OPEN ACCESS

EDITED BY
Qiang Lin,
University of Antwerp,
Belgium

REVIEWED BY
Peng Zhao,
Northwest University,
China
Rui Xing,
Northwest Institute of Plateau Biology
(CAS), China

*CORRESPONDENCE
Zhanling Xie
xiezhanling2012@126.com

SPECIALTY SECTION
This article was submitted to
Evolutionary and Genomic Microbiology,
a section of the journal
Frontiers in Microbiology

RECEIVED 24 October 2022
ACCEPTED 30 November 2022
PUBLISHED 28 December 2022

CITATION
Meng Q, Xie Z, Xu H, Guo J, Tang Y, Ma T, Peng Q, Wang B, Mao Y, Yan S, Yang J, Dong D, Duan Y, Zhang F and Gao T (2022) Out of the Qinghai-Tibetan plateau: Origin, evolution and historical biogeography of *Morchella* (both Elata and Esculenta clades).
Front. Microbiol. 13:1078663.
doi: 10.3389/fmicb.2022.1078663

Out of the Qinghai-Tibetan plateau: Origin, evolution and historical biogeography of *Morchella* (both Elata and Esculenta clades)

Qing Meng[1,2], Zhanling Xie[1,2]*, Hongyan Xu[1,3], Jing Guo[1,2], Yongpeng Tang[4], Ting Ma[1], Qingqing Peng[1], Bao Wang[1,2], Yujing Mao[1,2], Shangjin Yan[1], Jiabao Yang[1,2], Deyu Dong[1,2], Yingzhu Duan[4], Fan Zhang[5] and Taizhen Gao[4]

[1]College of Ecological and Environment Engineering, Qinghai University, Xining, Qinghai, China, [2]State Key Laboratory Breeding Base for Innovation and Utilization of Plateau Crop Germplasm, Qinghai University, Xining, Qinghai, China, [3]Academy of Agriculture and Forestry Sciences, Qinghai University, Xining, Qinghai, China, [4]State-owned Forest Farm of Tianjun County, Delingha, Qinghai, China, [5]Forestry and Grassland Station of Tianjun County, Delingha, Qinghai, China

Introduction: Morchella has become a research hotspot because of its wide distribution, delicious taste, and phenotypic plasticity. The Qinghai-Tibet Plateau subkingdoms (QTPs) are known as the cradle of Ice age biodiversity. However, the diversity of Morchella in the QTPs has been poorly investigated, especially in phylogenetic diversity, origin, and biogeography.

Methods: The genealogical concordance phylogenetic species recognition (GCPSR, based on Bayesian evolutionary analysis using sequences from the internal transcribed spacer (ITS), nuclear large subunit rDNA (nrLSU), translation elongation factor 1-α (EF1-α), and the largest and second largest subunits of RNA polymerase II (RPB1 and RPB2)), differentiation time estimation, and ancestral region reconstruction were used to infer Morchella's phylogenetic relationships and historical biogeography in the QTPs.

Results: Firstly, a total of 18 Morchella phylogenetic species are recognized in the QTPs, including 10 Elata clades and 8 Esculenta clades of 216 individuals Secondly, the divergences of the 18 phylogenetic species were 50.24–4.20 Mya (Eocene-Pliocene), which was closely related to the geological activities in the QTPs. Furthermore, the ancestor of Morchella probably originated in the Northern regions (Qilian Shan, Elata cade) and southwestern regions (Shangri-La, Esculenta clade) of QTPs and might have migrated from North America (Rufobrunnea clade) via Beringian Land Bridge (BLB) and Long-Distance Dispersal (LDD) expansions during the Late Cretaceous. Moreover, as the cradle of species origin and diversity, the fungi species in the QTPs have spread out and diffused to Eurasia and South Africa starting in the Paleogene Period.

Conclusion: This is the first report that Esculenta and Elata clade of Morchella originated from the QTPs because of orogenic, and rapid differentiation of fungi is strongly linked to geological uplift movement and refuge in marginal

areas of the QTPs. Our findings contribute to increasing the diversity of Morchella and offer more evidence for the origin theory of the QTPs.

KEYWORDS

Morchella, Qinghai-Tibet plateau subkingdoms, multigene phylogenetics, age estimation, phylogeographic structure

Introduction

As a famous edible mushroom, *Morchella* owns important ecological functions and has high commercial value around the world (Dissanayake et al., 2021; Wu H. et al., 2021; Yu et al., 2022). It was popular research in taxonomy, species diversity, distribution, ecological diversity, phylogeny, biogeography, and artificial cultivation of *Morchella* species (Annette et al., 1978; Dahlstrom et al., 2000; Hao et al., 2019; Liu et al., 2019; Ali et al., 2021; Cao et al., 2022; Deng et al., 2022). The distribution of *Morchella* exhibits a high level of cryptic speciation and provincialism due to phenotypic plasticity and unreliable morphological species recognition (O'Donnell et al., 2011; Du et al., 2012a, 2015, 2018; Richard et al., 2015;). There are 72 phylogenetically distinct species in the world that have been recognized in this genus based on GCPSR (Loizides et al., 2015, 2016; Yatsiuk et al., 2016; Baroni et al., 2018; Du et al., 2019). In China, which was known as the center of *Morchella* species diversification and rich floristic diversity, a total of 16 Elata clades and 27 Esculenta clades phylospecies have been recorded (Du et al., 2012a,b, 2019). The phylogenetic species diversity of the *Morchella* in the Qinghai-Tibet Plateau subkingdom (QTPs) is not yet known, though.

Great changes in crustal movement on earth occurred during the Phanerozoic Paleozoic (4.6 billion years ago), after occurring of biological explosions and forming of thick sedimentary limestone (Mittermeier et al., 2011; Kate, 2022). The QTPs have been uplifted steadily and rhythmically since the end of the Early Tertiary period, which is famous as the "Third Pole" of the earth (Mao et al., 2021; Spicer et al., 2021; Xiong et al., 2022); and were known as the ecological barrier of China and even Asia based on complex topography, variable plateau climate, and rich ecosystem (Sun et al., 2012; Qin et al., 2015; Miehe et al., 2019; Liu et al., 2022). Several fossil records of QTPs illustrated the cradle of mammalian fauna and mountain flora in the Ice age (Deng et al., 2020; Mao et al., 2021; Wu Y. D. et al., 2021). For fungus in the QTPs, the opportunities for genetic variation and speciation were strongly increased *via* the isolation of high-altitude geographic and geological complexity. Furthermore, the QTPs are also called a refugium created by microclimatic variations that provided some protection, and situ speciation and relic persistence in the early originated lineages (Shrestha et al., 2010; Yuan, 2015; Yan et al., 2017; Phonepaserd et al., 2019; Mao et al., 2021). The rapidly radiational differentiation of *Morchella* species was reported in North America, Asia, and Europe. *M. rufobrunnea* (Rufobrunnea clade), as the oldest taxon of the genus *Morchella* and might diverge into the basal lineage, originated in western North America in the late Jurassic (O'Donnell et al., 2011; Du et al., 2012a, 2015; Loizides et al., 2016, 2021). During the emergence of the Mid-Continental Seaway and the subsequent uplift of the Rocky Mountains, the ancestors of the Esculenta and Elata clades spread to eastern North America from western North America in the early Cretaceous (Sanmartín et al., 2001; Donoghue, 2008; Du et al., 2015). After that, *Morchella* experienced widespread extinction due to the new folding of the Rocky Mountains and the uplift of the Sierra Madre Oriental Range in central North America. It is also speculated that the *Morchella* species spread to Europe and Asia from North America *via* the Thulean North Atlantic Land Bridge and the Beringian Land Bridge (Du et al., 2015). During the middle Miocene to the Pleistocene, the *Morchella* species in East Asia and Europe rapidly evolved under the gradually cooling climate and environmental heterogeneity caused by the rise of the Qinghai-Tibetan Plateau (O'Donnell et al., 2011; Du et al., 2012a, 2015). However, further information regarding the differentiation, speciation, origination, and evolution of *Morchella* in the QTPs is still unclear.

In this study, ITS rDNA sequences of 174 *Morchella* individuals collected from QTPs were generated for the aims of (i) investigating phylogenetic species diversity and geographic distribution of *Morchella* in the QTPs by using genealogical concordance phylogenetic species recognition (GCPSR); (ii) estimating divergence times of *Morchella* species lineages in the QTPs; (iii) defining the geographic distributions of ancestor lineages of the *Morchella* in the QTPs; and (iv) estimating divergence times and reconstructing ancestral regions for world-widely distributed species of *Morchella*.

Materials and methods

Sampling

A total of 216 individuals were collected from three parts, among which 174 individuals were from the QTPs and 30 individuals were from Xinjiang and 12 individuals were from Jilin University. Total 7 natural localities approximately extending the whole distribution range of QTPs, ranging from 26.07°N–44.01°N to 81.60°E–105.05°E during the harvesting seasons (April–June).

We divided the sampling sites into the higher latitude (35°N–40°N), the lower latitude (25°N–30°N), and the middle latitude (30°N–35°N) regions according to the north latitude lines (Figure 1). Of these, there are 56, 64, and 92 individuals were collected from high, low, and middle latitudes of the northeastern, central-eastern, and southwestern in the QTPs, respectively. Micromorphological data were obtained from the dried specimens and observed under a light microscope following Baran and Boroń (2017). Voucher specimens were deposited in Extreme Environment Microbiology Laboratory, Qinghai University, Xining, China. The codes, locations, sampling year, and sample numbers of *Morchella* from the QTPs are shown in Supplementary Table S1.

DNA extraction, PCR, and DNA sequencing

Total extracted DNAs of each sample were extracted using the modified 2× CTAB buffer method (Doyle and Doyle, 1987), checked by 1% agarose gel electrophoresis with ethidium bromide staining, and measured the concentration by spectrophotometer (Biospec-nano; Shimadzu). Five DNA gene fragments were analyzed, including those coding for RNA polymerase II largest subunit (*RPB1*) and second largest subunit (*RPB2*), translation elongation factor-1a (*TEF1*), along with two non-protein coding regions: internal transcribed spacer (ITS), nuclear large subunit rDNA (nrLSU). The PCR amplifications was performed according to Du et al. (2019), Baran and Boroń (2017), and Loizides et al. (2021). The primers used for PCR are listed in Supplementary Table S2. A total of 894 sequences of QTP *Morchella* were newly generated, including 216 ITS, 201 nrLSU, 161 *EF1-α*, 166 *RPB1*, and 125 *RPB2*. The ITS sequences generated in this study were combined with the representative 117 ITS sequences retrieved from GenBank (including 46 representatives of *Morchella* species recorded in NCBI) to identify the relationships between all of our individuals and the known related samples in GenBank. All newly generated sequences were submitted to GenBank (Supplementary Table S1).

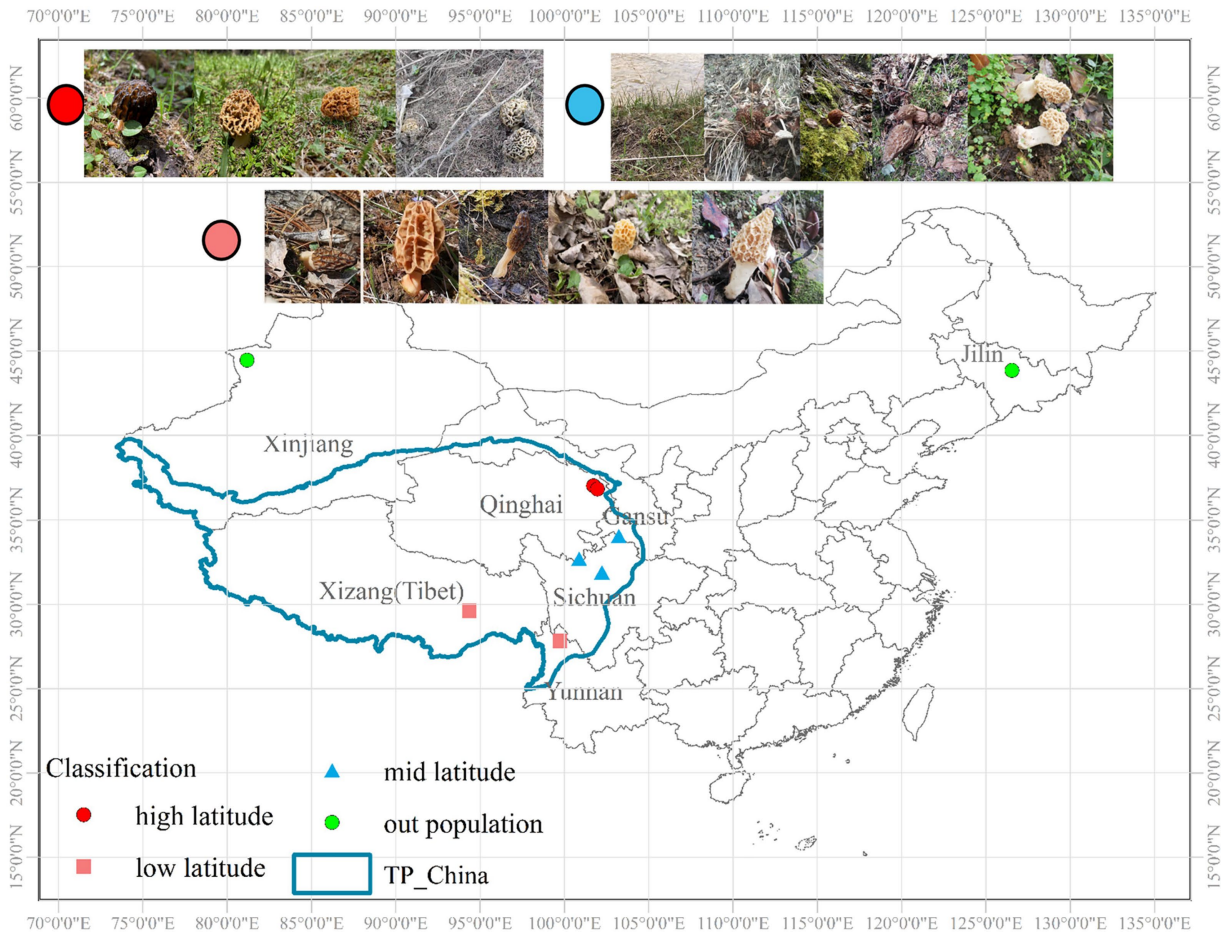


FIGURE 1
Geographical location and sampling site in Qinghai-Tibet Plateau subkingdom (QTPs) generated by ArcGIS v10.7. The red, pink, and blue sampling points are within the QTPs, where red is defined as the high-latitude region, blue is the middle-latitude region, and pink is the low-latitude region of the QTPs. The green sampling points are outside the QTPs.

Sequence alignments and phylogenetic analyses

For phylogenetic analyses, the ITS gene datasets were analyzed *via* Maximum likelihood (ML) and Bayesian inference (BI): a 110-taxon, 864 bp Esculenta Clade data set; and (ii) a 101-taxon, 612 bp Elata Clade dataset. Five-gene datasets were analyzed *via* Maximum parsimony (MP), ML, and BI: (i) a 14-taxon, 3,380 bp Esculenta Clade data set; and (ii) a 24-taxon, 2,517 bp Elata Clade dataset. The genes extracted and aligned were using the MAFFT (version 7; Katoh and Standley, 2013). The conservative region was selected in Gblocks 0.91b and the vacancy gap in the data were treated as missing data. We performed MP, ML, and BI based on the combined sequences of five genes to reconstruct the relationships of *Morchella* and related taxa were subsequently conducted on PAUP∗ version 4.0 beta 10 (Swofford, 2002) and PholySuite v1.2.2 (Zhang et al., 2020). For the MP analysis was performed in PAUP∗ version 4.0 beta 10 (Swofford, 2002). All characters were equally weighted, and gaps were treated as missing data. Trees were inferred using the heuristic search option with TBR branch swapping and 1,000 random sequence additions. The PholySuite v1.2.2 contains programs for sequence alignment and phylogenetic analysis, such as ModelFinder to find the best model. For the ML analyses, all parameters were kept at their default settings, the concatenated dataset was partitioned into five parts by sequence region, and 1,000 Ml searches under the K80 (K2P) model with all model parameters estimated using the ModelFinder program, IQ-TREE v1.6.8 web server[1] to carry out the ML searches. The MrBayes v3.2.6 for BI phylogenetic analyses also used ModelFinder to generate the best model (Ronquist et al., 2012; Du et al., 2012a; Zhu et al., 2019; Zhao et al., 2021). The phylogenetic trees were modified using FigTree v1.4.4 and the iTOL website.[2]

Divergence dating analysis

we used BEAST v.2.6.6 (Chen et al., 2015; Zhu et al., 2019; Kim and Kim, 2022) to estimate the divergence times of *Morchella* phylospecies in the QTPs. In this study, we used *Floccularia luteovirens*, which is the endemic basidiomycete fungus in the QTPs, instead of *P. devonicus* as the calibration point 1 in Basidiomycota and Ascomycota (Taylor et al., 2004; Chen et al., 2015; Zhu et al., 2019; Guo et al., 2022). Normal distribution was applied by setting the mean and the standard deviation to 582.5 and 50.15, respectively. Calibration points 2 for analysis were obtained by including sequences of the following two species: *Verpa* and *M. rufobrunnea* (O'Donnell et al., 2011; Du et al., 2012a, 2015; Loizides et al., 2016, 2021). The origin time of *Morchella* was estimated in BEAST v.2.6.6 (Drummond and Rambaut, 2007) with the molecular clock and substitution models unlinked but with the trees linked for each gene partition. Two nuclear ribosomal RNA genes (ITS and nrLSU) and three protein-coding genes (*EF1-α*, *RPB1*, and *RPB2*), were concatenated for molecular dating. PholySuite v1.2.2 was also used to select the best models of evolution using the hierarchical likelihood ratio test. The GTR + I + G model was used for the *EF1-α* + *RPB1* + *RPB2* and the HKY + I + G model for the ITS + nrLSU data, based on the results from the PholySuite v1.2.2. The uncorrelated lognormal relaxed molecular clock and the Yule speciation prior set were used to estimate the divergence time and the corresponding credibility intervals by BEAUti 2. The Markov chain Monte Carlo (MCMC) analysis was 100 million generations, sampling parameters for every 1,000 generations. After discarding the first 10,000 (10%) trees as burn-in, the samples were summarized in a maximum clade credibility tree in TreeAnnotator v2.6.6 using a PP limit of 0.50 and summarizing the mean node heights. The means and 95% higher posterior densities (HPDs) of age estimates were obtained from the combined outputs using Tracer. FigTree v1.4.2 and iTOL website[3] was used to visualize the resulting tree and to obtain the means and 95% HPD. A 95% HPD marks the shortest interval that contains 95% of the values sampled.

Biogeographic analysis

Ancestral area reconstruction and estimating spatial patterns of geographic diversification within *Morchella* in the QTPs were inferred using the Bayesian binary method (BBM) and statistical dispersal-vicariance analysis (S-DIVA) as implemented in Reconstruct Ancestral State in Phylogenies (RASP v3.1). The distribution range of the Elata clade in the QTPs was divided into five regions, consisting of A (Qinghai), B (Tibet), C (Gansu), D (Xinjiang), and E (Other). And the distribution range of the Esculenta clade in the QTPs was also divided into five regions, consisting of (A) Qinghai, (B) Gansu, (C) Sichuan, (D) Yunnan, and E (Other). For the BBM analysis, we used all post-burn-in trees obtained from the BEAST v2.6.6 analysis. The BBM was run using the fixed state frequencies model (Jukes-Cantor) with equal among-site rate variation for 50,000 generations, 10 chains each, and two parallel runs. In statistical dispersal-vicariance analysis (S-DIVA), the frequencies of an ancestral range at a node in ancestral reconstructions are averaged over all trees. In addition, the world-widely distribution of 8 *Morchella* species (Supplementary Table S2), *M. spongiola*, *M. esculenta*, *M. crassipes*, *M. eohespera*, *M. eximia*, *M. costata*, *Mel*-13, *Mel*-14, were downloaded in NCBI to estimate the differentiation time and reconstruction ancestral area using BEAST v2.6.6 and RASP v3.1, respectively. ArcGIS v10.7 was used to visualize the

1 http://iqtree.cibiv.univie.ac.at/

2 https://itol.embl.de/personal_page.cgi

3 https://itol.embl.de/personal_page.cgi

geographic distribution and possible dispersal routes of *Morchella* (Yu et al., 2015; Kim et al., 2019; Kim and Kim, 2022).

Results

The recognition of 18 phylospecies of *Morchella* in the QTPs

A total of 216 individuals of *Morchella* were classified into Esculenta and Elata clades, there is no Rufobrunnea clade. A total of 101 individuals clustered with 10 phylogenetic species, including *Mel*-14, *M. deliciosa*/*Mel*-13, *M. norvegiensis* = *M. eohespera*/*Mel*-19, *Morchella eximia*/*Mel*-5, *Morchella costata*, *Morchella sextelata*/*Mel*-6, *Morchella septimelata*/*Mel*-7, *Morchella purpurascens*/*Mel*-20, *Mel*-33, and *Morchella pulchella*/*Mel*-31 belongs to Elata clade (Figures 2A–C); and a total of 101 individuals clustered with 8 phylogenetic species, including *M. vulgaris* = *M. spongiola*/*Mes*-5, *Mes*-9, *Mes*-12, *Mes*-26, *Morchella crassipes*, *Morchella esculenta*/*Mes*-8, *Mes*-19, and *Mes*-6 belongs to Esculenta clade (Figures 3A–C). All *Morchella* species also exhibit extreme bradytelic morphological evolution as evidenced by the retention of the ancestral ascocarp body plan (Figures 2D, 3D). Overall, we identified 18 phylospecies that was widely distributed in the QTPs.

The estimation of the divergence time of *Morchella* in the QTPs

The divergence time of 18 *Morchella* phylospecies in the QTPs ranged from 50.24 to 4.20 Mya (Figure 4; Table 1). The earliest diverging branch of *Morchella* in the QTPs was represented by the monotypic *M. rufobrunnea* (originated in North America) lineage with an estimated divergence time of 154.15 Mya (95% HPD interval: 152.14–156.08); the second diverging branch of Esculenta and Elata clades at 98.63 Mya (95% HPD interval: 97.30–100.0); the third evolutionary diversification of the Elata Clade was dated at 57.94–68.97 Mya (95% HPD interval: 30.09–69.04; 40.17–98.45) and the Esculenta Clade at 20.41–20.83 Mya (95% HPD interval:

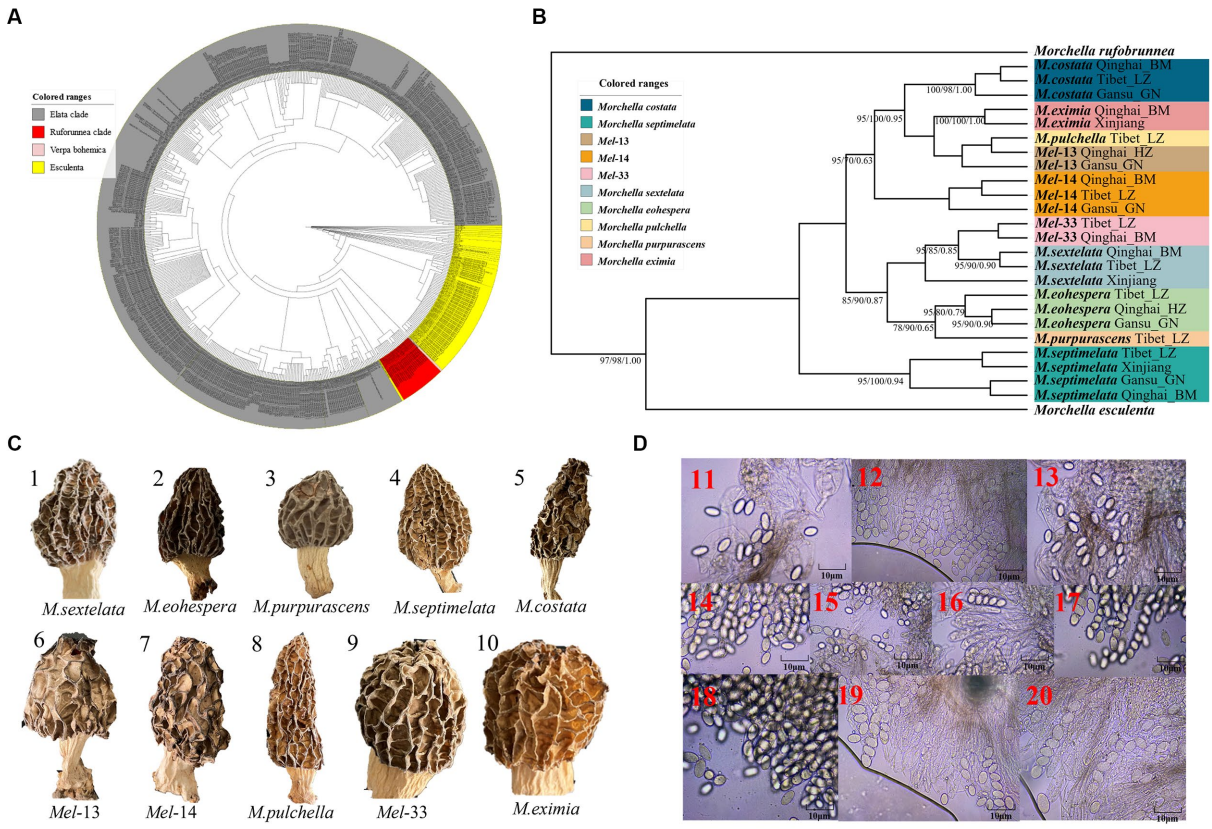


FIGURE 2
Species recognition of the *Morchella* in Elata clade from the QTPs. **(A)** Bayesian inference phylogenetic analyses of the Elata clade were inferred from 115 internal transcribed spacer (ITS) sequences representing a total of 10 phylospecies. **(B)** Phylogenetic analyses of the Elata clade were inferred from 120 (24*5) multi-genes (ITS+LSU+*EF1-α*+*RPB1*+*RPB2*) sequences representing a total of ten phylospecies. Branches are labeled where MP/ML support is greater than 60% and collapsed below that support threshold. BPP is labeled were greater than 0.95. **(C)** Morphological diversity of the 10 Elata clades' ascocarps from the QTPs: *M. sextelata*/*Mel*-6 (1), *M. norvegiensis* = *M. eohespera*/*Mel*-19 (2), *M. purpurascens*/*Mel*-20 (3), *M. septimelata*/*Mel*-7 (4), *M. costata* (5), *M. deliciosa*/*Mel*-13 (6), *Mel*-14 (7), *M. pulchella*/*Mel*-31 (8), *Mel*-33 (9), *M. eximia*/*Mel*-5 (10). **(D)** Micromorphological ascospores of the 10 Elata clades.

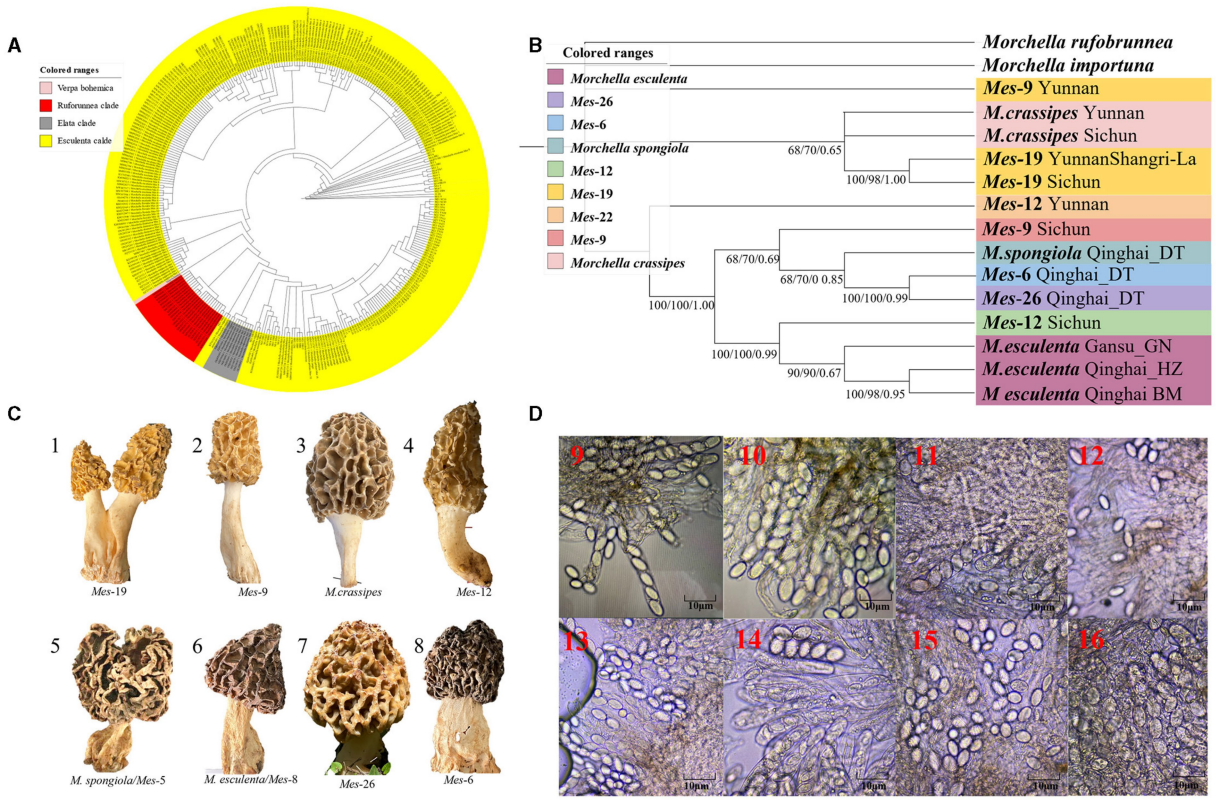


FIGURE 3
Species recognition of the *Morchella* in Esculenta clade from the QTPs. **(A)** Bayesian inference phylogenetic analyses of the Elata clade were inferred from 101 ITS sequences representing a total of 8 phylospecies. **(B)** Phylogenetic analyses of the Esculenta clade were inferred from 70 (14*5) multi-genes (ITS+LSU+*EF1-α*+*RPB1*+*RPB2*) sequences representing a total of 8 phylospecies. Branches are labeled where MP/ML support is greater than 60% and collapsed below that support threshold. BPP is labeled were greater than 0.95. **(C)** Morphological diversity of the 8 Esculenta clades' ascocarps from the QTPs: *Mes*-19 (1); *Mes*-9 (2), *M. crassipes* (3), *Mes*-12 (4), *M. vulgaris* = *M. spongiola*/*Mes*-5 (5), *M. esculenta*/*Mes*-8 (6), *Mes*-26 (7), *Mes*-6 (8). **(D)** Micromorphological ascospores of the 8 Escuenlta clades.

6.18–45.22; 2.62–52.67). The phylospecies of *M. norvegiensis* = *M. eohespera*/*Mel*-19, *M. deliciosa*/*Mel*-13, *Mel*-14, *Morchella eximia*, *M. costata*, *M. esculenta*/*Mes*-8, *M. crassipes*, and *Mes*-19, were estimated at 50.24 Mya (95% HPD interval: 40.17–98.45), 40.63 Mya (95% HPD interval: 18.95–63.83), 36.41 Mya (95% HPD interval: 2.83–66.27), 25.87 Mya (95% HPD interval: 4.77–48.44), 14.24 Mya (95% HPD interval: 6.18–45.22), 4.20 (95% HPD interval: 0.02–19.54), respectively. In summary, *Morchella* phylospecies in QTPs has maintained a very diversified evolutionary history during the Eocene and Pliocene, when the historical geological uplift and geological tectonic movement were experienced in the QTPs (Figure 5).

The reconstruction of the ancestral area and spatial patterns of *Morchella* in the QTPs

There were 19 dispersal events and 19 vicariance events that could explain the current distribution of the *Morchella* phylospecies in the QTPs. For the Elata clade, region A (Qilian Shan, Qinghai) located in the eastern part of QTPs, has the highest probability (50.08%) of being the ancestral area during the Eocene (Figure 6A). For the Esculenta clade, region D (Shangri-La, Yunnan), which was located in the southwestern part of QTPs, has the highest probability (77.48%) of being the ancestral area during the Miocene (Figure 6B). In general, based on the phylogeographic structures of the 8 subclades, 3 distribution patterns can be summarized: (i) the wide distribution around the QTPs, such as the *M. deliciosa*/*Mel*-13, *Mel*-14, *M. norvegiensis* = *M. eohespera*/*Mel*-19, suggests that not all of the *Morchella* species were narrowly distributed; (ii) the long-distance dispersal with latitude-based structure. (iii) Multi-origin: the two clades have different origins, the Easculenta clade originated from Shangri-La while the Esculenta clade originated from Qilian Shan (Figure 7).

The origination and evolution of worldwide *Morchella*

For the *Morchella* genus, there is no Rufobrunnea clade on any continent except North America, and most of the species that are

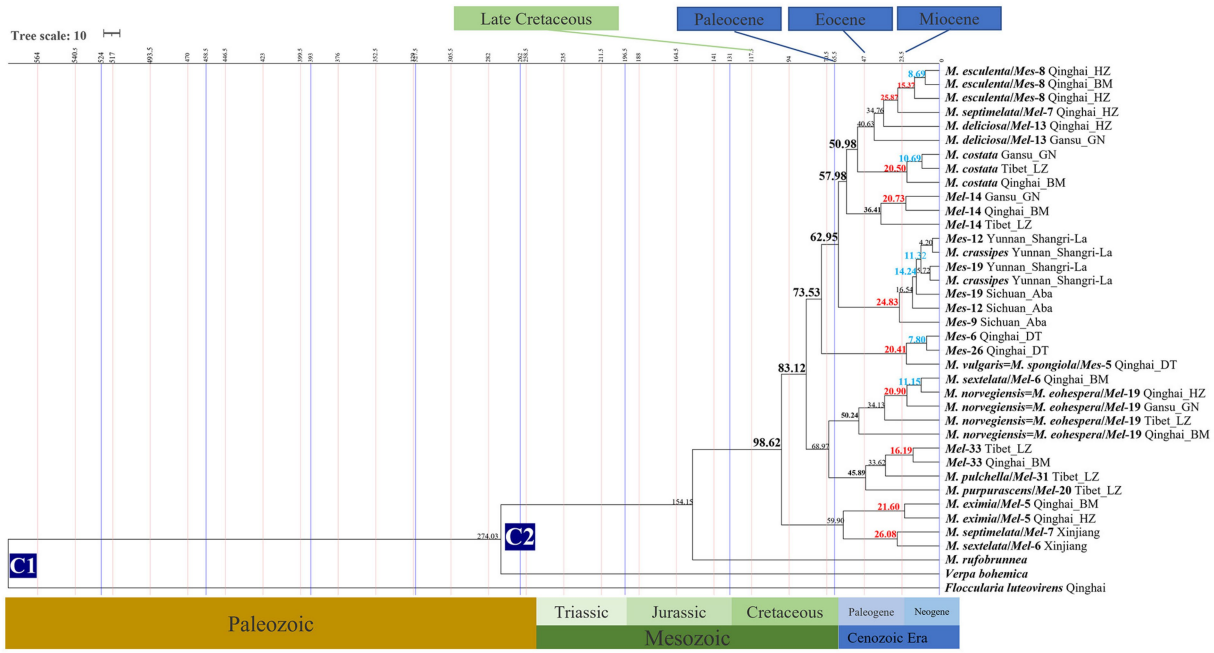


FIGURE 4
Chronogram and estimated divergence times of *Morchella* in QTPs generated by molecular clock analysis using the two concatenated datasets (ITS+LSU and *EF1-α+RPB1+RPB2*) dataset. The chronogram was obtained using the Ascomycota-Basidiomycota divergence time of 582.08 Mya as the calibration point 1. The *Morchella-Verpa bohenica* divergence time of 274.06 Mya as the calibration point 2. The calibration point and objects of this study are marked in the chronogram. The geological time scale is millions of years ago (Mya). The red font is defined as the first uplift of the QTPs, and the blue font is defined as the second uplift of the QTPs.

widely distributed in the continental region (besides Oceania) have the earliest divergence time in the QTPs, the ancestor of *Morchella* both the Esculenta and Elata probably originated in QTPs and initially covered Eurasian and South Africa in the Early Tertiary (Figure 8; Table 2).

For the Esculenta clade: (1) *M. vulgaris* = *M. spongiola/Mes*-5 there were distributed in central Europe and Asia, and the initial differentiation of *M. spongiloa* in these two areas was 28.88 Mya (95% HPD interval: 4.11–59.68) and 49.03 Mya (95% HPD interval: 26.48–68.63), respectively. They went through an extinction event between Asia and Europe. QTPs of Asia and Germany of Europe were most likely ancestral regions. However, the Asian phylospecies diverged 11 Mya earlier than the European species, and the QTPs were the ancestral region of *M. spongiloa* (Table 2; Supplementary Figure S1). (2) *M. esculenta/Mes*-8 was widely distributed in Europe, North America, and Asia, among them, the earliest divergences of this species occurred in China at 38.92 Mya (95% HPD interval: 19.26–56.02). That phylospecies experienced one extinction event in the distribution area and the highest probability (38.81%) of being a putative ancestral region is the QTPs (Table 2; Supplementary Figure S2). (3) *M. crassipes* is a widely distributed species in Asia, Europe, North America, and South Africa. The earliest differentiation of this species is Shanghai during the 68.97 Mya (95% HPD interval: 67.04–70.93), but North America has the highest probability (39.67%) of being a putative ancestral region (Table 2; Supplementary Figure S3).

For Elata clade: (4) *M. norvegiensis* = *M. eohespera/Mel*-19 (Elata clade), were distributed in Central Europe, Asia North America. The initial divergence time in the QTPs is 52.25 Mya (95% HPD interval: 24.37–74.33), which is earlier than in North America and Central Europe [15.81Mya (95% HPD interval: 0.13–49.54), 9.79Mya (95% HPD interval: 0–36.55)]. The QTPs have the highest probability (76.92%) of being the ancestral area of *M. eohespera* and the dispersal events and extinction events also occurred in the continental range (Table 2; Supplementary Figure S4). (5) *M. deliciosa/Mel*-13 and *Mel*-14 are widely distributed in Eurasia, especially in the QTPs. *M. deliciosa/Mel*-13 and *M. importuna* diverged at 35.67 Mya (95% HPD interval: 21.92–72.25), while *Mel*-14 and *M. importuna* diverged at 39.55Mya (95% HPD interval: 12.05–72.82). Europe is the putative ancestral region of *Mel*-13 (10.67%) while QTPs have the highest probability of being an ancestral region of *Mel*-14 (33.40%), and they all experienced one extinction event (Table 2; Supplementary Figure S5). (6) *M. eximia* are widely distributed in America, Europe, and Asia, which were estimated at 16.56 Mya (95% HPD interval: 10.23–67.80) in QTPs (Supplementary Figure S6). (7) *M. costata* were widely distributed in Eurasia. The initial differentiation in QTPs was 40.53 Mya (95% HPD interval: 42.31–71.50), which has the highest probability (81.35%) of being an ancestral region of *M. costata* (Table 2;

TABLE 1 The estimation of the divergence time of *Morchella* in the QTPs.

Node	Individual numbers	Mean ± standard error[a]	95% HPD[a]	Ancestors region[b]	Geological events[c]
Ascomycota/Basidiomycota	–	564.85 ± 0.11	467.24–666.82	–	Cambrian
Verpa bohemic/*Morchella*	–	274.03 ± 0.31	272.08–276	–	Triassic
Morchella rufobrunnea	–	154.15 ± 0.06	152.14–156.08	North America	Cretaceous
Esculenta/Elata	–	62.95 ± 0.38	51.24–69.76	–	Paleocene
Morchella eximia	18	21.6 ± 0.35	6.15–60.04	Qilian Mountains in the	the geological strike-slip
Morchella eohespera	19	50.24 ± 0.25	40.17–98.45	eastern part of the QTPs	The first stage of the uplift
Mel-33	3	16.19 ± 1.07	0.11–52.08		The second uplift
Morchella sextelata	8	11.15 ± 1.94	0.1–35.13		The third uplift
Morchella costata	11	20.05 ± 0.60	6.82–58.62		the geological strike-slip
Mel-14	11	26.41 ± 4.16	2.62–52.67		the geological strike-slip
Mel-13	21	40.63 ± 0.97	18.95–63.63		The first stage of the uplift
Morchella septimelata	3	26.08 ± 0.71	0.04–54.98		the geological strike-slip
M. purpurascens/*Mel*-20	5	45.89 ± 0.55	16.89–88.59		The first stage of the uplift
M. pulchella/*Mel*-31	3	33.62 ± 0.50	7.11–74.38		The second stage of uplift
Morchella esculenta	24	25.87 ± 0.87	4.77–48.44	Shangri-la in the	the geological strike-slip
Morchella crassipes	26	5.72 ± 0.15	0.16–16.35	southwestern QTPs	The third uplift
Morchella spongiola	11	20.41 ± 2.08	2.62–52.67		the geological strike-slip
Mes-26	9	7.8 ± 0.33	0.54–32.68		The third uplift
Mes-19	25	24.14 ± 0.31	0.09–28.99		the geological strike-slip
Mes-12	4	16.54 ± 0.15	0.02–19.54		The second uplift
Mes-9	4	24.83 ± 1.10	0.0–23.54		the geological strike-slip
Mes-6	11	7.8 ± 0.33	0.54–32.68		The third uplift

[a]The divergence times and 95% higher posterior densities (HPDs) were generated by molecular clock analysis using the two concatenated datasets (ITS + LSU and *EF1-α* + *RPB1* + *RPB2*) dataset.
[b]The ancestor region probability was obtained from the most likely states (MLS) using the Bayesian binary method (BBM) and statistical dispersal-vicariance analysis (S-DIVA) as implemented in Reconstruct Ancestral State in Phylogenies (RASP v3.1).
[c]The geological events were referenced in Dai et al. (2019).

Supplementary Figure S7). Overall, six of the eight worldwide *Morchella* originated in the QTPs, the ancestral region of 60% (3/5) of the Elata clade and 66.66% (2/3) of the Esculenta clade were the QTPs; therefore, it is possible that QTPs was the center of origin for the current *Morchella* species diversity. *Morchella* originated from the QTPs and expanded out and spread to the other continents in Paleogene Period (Figure 8).

Discussion

This is the first time report the phylospecies diversity of *Morchella* in the QTPs used GCSPR, and the results indicate that there is 18 phylospecies of *Morchella* in QTPs, which far exceeds the number of published taxa for these areas (Du et al., 2012b, 2016, 2019). Generally, both Esculenta and Elata clades have preferred habitats, known as phylogenetic niche conservation (PNC; Donoghue, 2008). We investigated the occurrence and preferred habitats of *Morchella* in the QTPs and found that occurred in primary forests with high vegetation coverage, deciduous forests, or mixed coniferous and broad-leaved forests, such as the primary forests of Qilian Shan, Belong River, Mote, and Shangri-la, all of which belong to the eastern or northwestern margin of the QTPs. In addition, about 90% of the species are found in temperate deciduous forests, while 80% of the species were found in coniferous forests (Figure 1). It can be inferred that the occurrence of *Morchella* in the QTPs was accompanied by the establishment of the temperate deciduous biome and coniferous biome in the late Cretaceous (O'Donnell et al., 2011; Du et al., 2012a, 2015; Chen et al., 2015).

The maximum crown age of *Morchella* in the QTPs was estimated to be around the Late Cretaceous (98.62 Mya). The majority of intracontinental range expansions within *Morchella* in the QTPs appear to have taken place relatively recently between the middle Oligocene and Miocene. During that time, the uplift of the southwest block (Yunnan-Guizhou Plateau) and the eastern margin of the QTP resulted in the distribution of plant diversity in this region. Previous studies have shown that the Beringian Land Bridge (BLB) plays a key role in the spread of *Morchella* species from North America to Asia in the early Cretaceous, which was a natural route linking Eurasia and North America (O'Donnell et al., 2011; Du et al., 2012a, 2015; Loizides et al., 2016, 2021). Palaeobotanical data indicate that the BLB route played a crucial role in plant dispersal for *Populus*, *Lonicera*, *Leibnitzia*, and *Linnaea* (Wang et al., 2009; Qiu et al., 2011; Wen et al., 2014; Du et al., 2015; Wang et al., 2015; Wu et al., 2017). The divergence events exhibited by the plant taxa

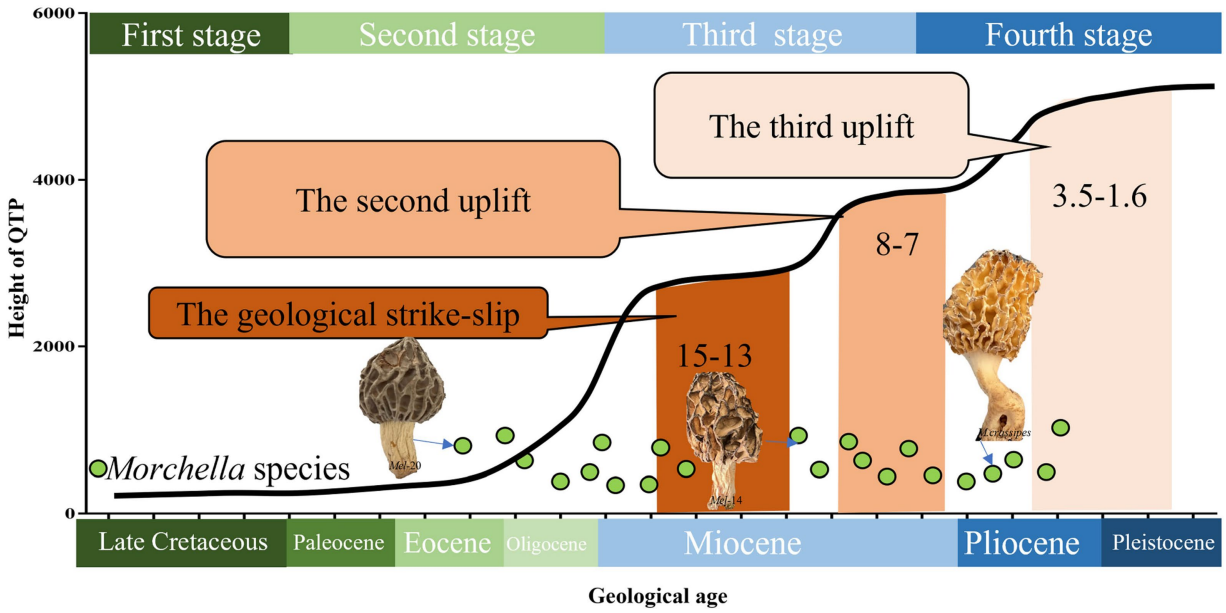


FIGURE 5
Species differentiation of *Morchella* related to the stage of plate movement and uplift on the QTPs from the Late Cretaceous to the present. The divergence time of *Morchella* phylospecies in QTPs has remained remarkably diverse during its long evolutionary history, ranging from 50.24 to 4.20 Mya (Eocene-Pliocene), when the historical geological uplift and geological tectonic movement were experienced in the QTPs. The plate activity and plateau uplift of QTP resulting from the collision are divided into four stages: (1) Late Cretaceous in the first stage marked as dark green; (2) Tertiary Period in the second stage marked as light green; (3) Neogene in the third stage marked as light blue; (4) Quaternary Period in the fourth stage marked as dark blue. The green dots represent the *Morchella* individual.

(*Pinus*, *Abies*, and *Picea*) confirm the geological events associated with species diversification, and those indigenous florae provide native secluded habitats for ancestral species of *Morchella* in QTPs (O'Donnell et al., 2011; Loizides et al., 2021). The Beringian Land Bridge as a species expansion channel is a critical driver of *Morchella* migrating to the QTPs between the Cretaceous and Paleocene when plant diversity was already established on QTPs, which means that the QTPs had formed a variety of suitable habitats for living organisms before in the Paleocene.

The differentiation of the *Morchella* is strongly linked to the geological movements of the QTPs (Figure 5). Palaeogeographical evidence suggests that plate activity and plateau uplift of QTP resulting from the collision are divided into four stages (Shi et al., 1998; Ding et al., 2022). (i) Late Cretaceous, the ancient Mediterranean crust subducted to the Eurasian continental crust, blowing the prelude to the plateau uplift. At this stage, *Morchella* of western North America is diverging from its closest relatives in the early Cretaceous, which are divided into Rufobrunnea, Esculenta, and Elata clades, and palaeogeological events played a driving role in the dispersal of *Morchella* (O'Donnell et al., 2011). (ii) Tertiary Period (Including the Paleocene, Eocene, and Oligocene), the Indian plate collided with the Eurasian plate, opening the prelude to the plateau uplift. The Indian plate rapidly drifted northward, the Indo-Pak subcontinent and the subcontinent were getting closer, and the ancient Mediterranean crust gradually disappeared during the Oligocene. In our data, *M. norvegiensis* = *M. eohespera*/*Mel*-19 was differentiated at 50.24 Mya with the new uplift belts of Tengchong-Bango formatted and the uplift area of Songpan-Ganzi shrank to the east during Eocene (Wang et al., 2008, 2022). Eocene (55.8 ± 0.2) was a key period for Asian paleoenvironmental changes and was characterized by a warmer climate than any other interval in the Cenozoic (Shi et al., 1998; Zachos et al., 2001; Hoorn et al., 2012). *M. norvegiensis* = *M. eohespera*/*Mel*-19 in the middle latitudes region were differentiated at 34.24 Mya with the further uplifted of Kunlun-Algin-Qilian during the Oligocene, which has been called the beginning of the present 'icehouse' epoch (Ling et al., 2021; Dieter et al., 2022). (iii) Neogene, the Indian and Tarim plates compressed and subducted to the QTPs with greater stress, and the plateau was greatly uplifted, forming the QTPs and the Himalayas. Our results indicate that 9 of the 18 QTPs *Morchella* species lineages (i.e., 50%) diversified between the middle Miocene and the present (Figure 4). (iv) Quaternary Period, the plateau was greatly uplifted, forming the present towering QTP. The Esculenta clade of *Morchella* in the QTPs was undergoing differentiation at this time. Tectonic activity and climate change in geological periods can form geographical isolation barriers to promote species differentiation and increase diversity, and can also reduce biological diffusion barriers to expand species distribution areas and increase biological exchanges between different regions (Che et al., 2010; Antonelli et al., 2018; Rahbek et al., 2019; Ding et al., 2020). Generally speaking, the formation and evolution of *Morchella* phylospecies in the QTPs were affected by the tectonic uplift of the QTPs and the geological movements of QTPs were an important force for the differentiation of the *Morchella* species

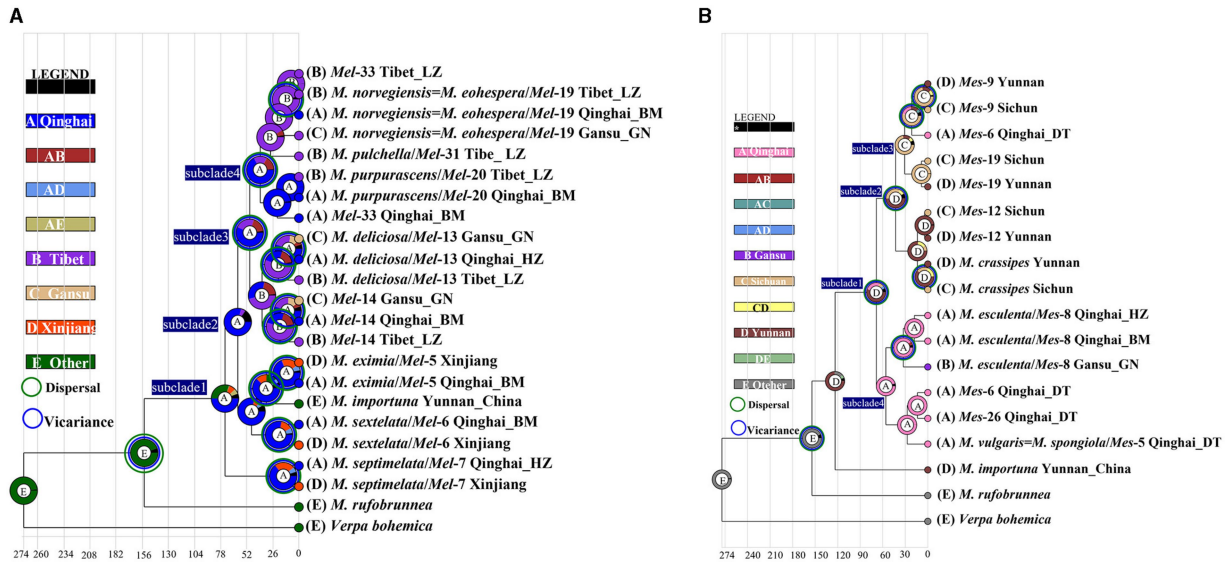


FIGURE 6
Ancestral area reconstruction of *Morchella* phylospecies in QTPs using the ITS dataset. The chronogram was obtained by molecular clock analysis using BEAST. The pie chart in each node indicates the possible ancestral distributions inferred from Bayesian Binary MCMC analysis (BBM) implemented in RASP. Bayesian credibility values (PP) over 0.85 are indicated near the pie chart of the tree. The green circle around the pie charts indicates possible dispersal events, the blue circle indicates possible vicariance events as suggested by BBM analysis. **(A)** Elata clade; **(B)** Esculenta clade.

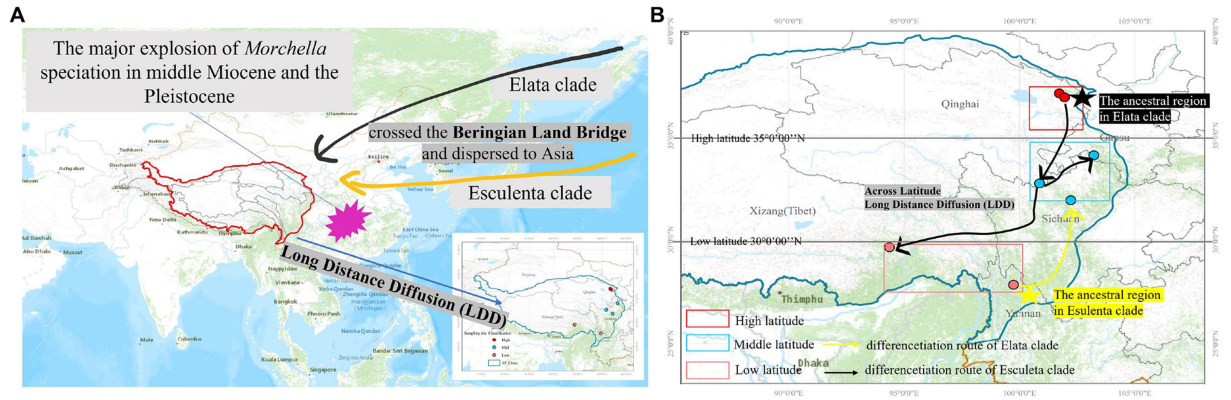


FIGURE 7
Map of the geographic distribution of *Morchella* and possible dispersal routes in QTPs generated by ArcGIS v10.1. **(A)** The major explosion of *Morchella* speciation in the middle Miocene and the Pleistocene crossed the Beringian Land Bridge and dispersed to Asia. **(B)** A hypothetical schematic depiction of the original locations, the migration routes the speciation of *Morchella* in QTPs.

(Merckx et al., 2015; Dai et al., 2019). The early marginal geological movement of the QTPs caused strong habitat fragmentation and rapid expansion of dry and cold habitats, which further strengthened the monsoon climate in East Asia and promoted the differentiation of microorganisms in the QTPs (An et al., 2001; Yang, 2005; Wan et al., 2014; Liu et al., 2016).

At least three geographic distribution patterns were discovered to correspond to the following phylogeographic structures. (i) the wide distributions around the QTPs: geographic locations of 4 endemic species, such as *Mel*-13, *Mel*-14, *Mel*-19, and *M. esculenta*, cross over three regions (Qinghai, Gansu, and Tibet in high, middle, and low latitudes respectively) of the QTPs. (ii) provincialism in the QTPs: the specific local distributions of two species in the Elata clade (*M. pulchella/Mel*-31, *M. purpurascens/Mel*-20) were unique (only in the Tibet region). (iii) the long-distance dispersal with latitude-based distributions: the divergence of *Mel*-13 and *Mel*-14 experienced two dispersal and vicariance events in three regions of the QTPs (Figure 6A). *Mel*-13 diverged in the mid-latitudes at 40.63

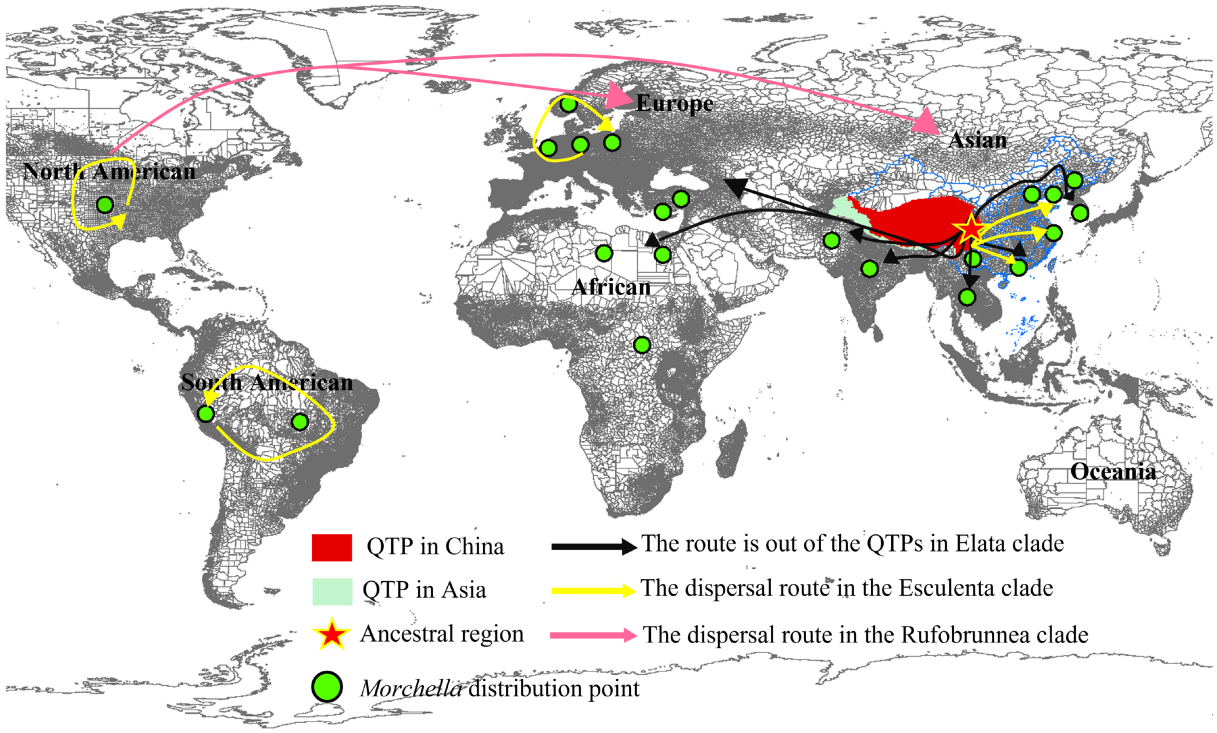


FIGURE 8
The origin of the QTPs and the route out of the QTPs in *Morchella*. *Morchella* originated from Qilian Shan and spread along the Himalayas on the edge of QTPs to surrounding Europe and South Africa in the Elata clade (Black route). Each continent has its ancestor in the Esculenta clade (Yellow route).

Mya and subsequently differentiated to the higher-latitudes at 34.76 Mya, which was based on latitude long-distance dispersal. *Mes*-19 also experienced dispersal and vicariance events in mid-latitudes and lower-latitudes, although this species originated from low latitudes (Yunnan) of the southwestern QTP (Figure 6B). To sum up, long-distance dispersals based on latitudes may have contributed to the current disjunct distribution ranges of *Morchella*, which is supported by biogeographic studies of plants and animals on the QTPs (Figure 7). The spread and differentiated way of *Morchella* in the global range, including intercontinental, intracontinental, and putative transoceanic long-distance dispersals, was slightly different from our result in QTPs (Du et al., 2012; Quan et al., 2014).

The complex topography and a special eco-climate of QTPs provide multiple periglacial microrefugia for *Morchella*. The Qilian Shan, located in the northeastern of the QTP, was defined as the ancestral region of the Elata clade as early as the Eocene. The Qilian Shan is an intraplate orogenic belt that experienced multiple episodes of fold and thrust deformation throughout the Mesozoic and Cenozoic periods, which controlled the evolution of regional climatic conditions in a broad region in inland Asia since the Miocene (Jia et al., 2022). The Shangri-La region in the southeast of the QTPs, which was the ancestors' region for the Esculenta clade of *Morchella*, not only served as an important glacial refugium but also as a center of diversification for a variety of plants and animals (Suzuki et al., 2003; Yang et al., 2008, 2009; Harrison and Noss, 2017). The QTPs contain several important biodiversity hotspots, particularly along its southeastern margin (the Henduan Mountains and the eastern Himalayas), which were proposed to be glacial refugia for alpine hepialid species that effectively avoided extinction during the Quaternary glacial period (Hewitt, 2000; Dai et al., 2019).

Multiple pieces of evidence suggest that the Elata clade originated from QTPs, whereas the origin of the Esculenta clade was diverse. Our results show that the worldwide distribution species was an initial differentiation time of 73.78 Mya–35.67 Mya earlier than Central Asia, Europe, and South Africa (Figure 8). During the Early Tertiary (65.5–23.03 Mya), the Indian plate collided with the Eurasian plate opening the prelude to the plateau uplift, which is the emergence time of modern organisms (Liu and Hu, 2003). For example, in *M. norvegiensis* = *M. eohespera*/*Mel*-19, the divergence time was estimated at 52.25 Mya, which was earlier than that of Europe (Supplementary Figure S4). The *Morchella* was out of QTPs probably because of Eurasia's collision and plate migration. On the other hand, the probability of QTPs as the most likely ancestral region for these widely distributed species is greater than that of Europe, North America, and South Africa (Table 2). All world-widely phylospecies have experienced historical extinction events that may be caused by the palaeogeology events and tectonic movement of the QTPs during

TABLE 2 Species initiation differentiation time and ancestor region reconstruction of worldwide *Morchella* species.

No	Phylospecies	Distribution[a]	Extinction event time (95% HPD interval)[b]	Initiation divergence time (95% HPD interval)[c]	Probability of ancestor region[d]		
					QTPs (%)	Europe (%)	North America (%)
1	*M. spongiloa*	Asia, Europe, and North America	68.70 Mya (66.70–70.63)	49.03 Mya (26.48–68.63)	16.23	6.93	none
2	*M. esculenta*	Asia and North America	59.98 Mya (26.48–68.63)	38.92 Mya (19.26–56.02)	38.81	20.10	5.91
3	*M. crassipes*	Asia, Europe, North America, South America, South Africa	65.72 Mya (59.87–70.12)	68.97 Mya (44.12–66.38)	23.45	25.86	39.67
4	*M. eohespera*	Asia, Central Europe, North America	28.00 Mya (4.11–59.68)	52.25 Mya (24.37–74.35)	76.92	6.50	7.50
5	*Mel*-13	Asia, Europe	35.67 Mya (10.54–62.42)	19.59 Mya (0.50–44.58)	0	10.67	none
6	*Mel*-14	Asia, Europe	39.55 Mya (12.55–42.72)	39.55 Mya (12.55–42.72)	33.40	8.30	none
7	*M. eximia*	Asia, Europe, and North America	61.89 Mya (43.31–74.34)	37.78 Mya (10.24–67.87)	32.64	39.96	17.22
8	*M. costata*	Asia, Europe	72.64 Mya (71.81–75.89)	72.64 Mya (71.81–75.89)	81.35	1.87	none

[a]Searched in NCBI.
[b]The divergence times were generated by molecular clock analysis using the ITS dataset.
[c]The means and 95% higher posterior densities (HPDs) of age estimates were obtained from the combined outputs using Tracer.
[d]The ancestor region probability was obtained from the most likely states (MLS) using RASP.

the Eocene, followed by the differentiation and expansion of the two clades of *Morchella* (Figure 8). For the Elata clade, *Morchella* originated from Qilian Shan and spread along the Himalayas on the edge of QTPs to surrounding Europe and South Africa. However, after the Esculenta clade originated from Shangri-La and spread to other regions of Asia, it was subsequently blocked by the geographical isolation formed in the Quaternary, which led to the diversification of the origin of the Esculenta clade. This result further enriches the theory of species origin in QTPs, except for the reported species such as *Floccularia luteovirens*, *Ophiocordyceps sinensis*, *Saccharomyces pastorianus*, etc. (Quan et al., 2014; Dai et al., 2019; Wu Y. et al., 2021; Bai, 2022; Guo et al., 2022).

Conclusion

In this study, a total of 216 individuals were identified 18 *Morchella* phylospecies of two clades (Elata and Escuelnta clade), and the results indicate that the *Morchella*'s phylogenetic diversity within the QTPs far exceeds the number of published taxa for these areas. The divergence time of the 18 *Morchella* phylospecies occurred in the Eocene-Pliocene period (50.24–4.2 Mya), it was strongly related to the geological strike-slip and the uplift movements of the QTPs, suggesting that the geographical movements had a large influence on the differentiation of the *Morchella*. Furthermore, the reconstructed ancestral areas of the Elata and Esculenta clades indicate that the northwestern and southeastern regions of QTPs are the likely ancestral area, which also has been suggested to coincide with the glacial refugia in the Quaternary. Moreover, we confirmed that Elata and Esculenta of *Morchella* originated from QTPs and spread out with plateau geological uplift, transoceanic or transcontinental long-distance dispersal, orogeny, or geographical movements during the Cenozoic Era. The origin of the Esculenta clade is not unique. These results offer strong evidence for the theory of the origin of species in the QTPs.

Data availability statement

The data presented in the study are deposited in the NCBI repository, accession number of those data was list in Supplementary Table S1.

Author contributions

QM: conceptualization, methodology, and writing-original draft. ZX: supervision, validation, and writing review and editing. HX, JG, and QP and TM: writing-review and editing. YT, BW, YM, SY, JY, YD, FZ, and TG: investigation. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Natural Science Planning Project of Qinghai Province (Grant no. 2021-HZ-802), and the first batch of central forestry and grassland ecological protection and restoration funds in 2021 (2021-87).

Acknowledgments

The authors appreciate Pan Tong (Chief Engineer of Qinghai Provincial Bureau of Geology and Mineral Exploration and Development) for the technical assistance and theoretical guidance in the Qinghai-Tibet Plateau uplift.

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.

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.2022.1078663/full#supplementary-material

References

Ali, S., Imran, A., Fiaz, M., Khalid, A. N., and Khan, S. M. (2021). Molecular identification of true morels (*Morchella* spp.) from the Hindu Kush temperate forests leads to three new records from Pakistan. *Gene Rep.* 23:101125. doi: 10.1016/j.genrep.2021.101125

An, Z. S., Kutzbach, J. E., Prell, W. L., and Porter, S. C. (2001). Evolution of Asian monsoons and phased uplift of the Himalaya-Tibetan plateau since late Miocene times. *Nature* 411, 62–66. doi: 10.1038/35075035

Annette, H., George, B., and Ina, L. (1978). Cultural studies of single Ascospore isolates of *Morchella esculenta*. *Mycologia* 70, 1269–1274. doi: 10.2307/3759331

Antonelli, A., Kissling, W. D., Flantua, S. G. A., Bermúdez, M. A., Mulch, A., Muellner-Riehl, A. N., et al. (2018). Geological and climatic influences on mountain biodiversity. *Nat. Geosci.* 11, 718–725. doi: 10.1038/s41561-018-0236-z

Bai, F. Y. (2022). Out of the Tibetan plateau: origin and evolution of lager yeast. *Acta Microbiol. Sin.* 62, 4188–4201. doi: 10.13343/j.cnki.wsxb.20220451

Baran, J., and Boroń, P. (2017). Two species of true morels (the genus *Morchella*, Ascomycota) recorded in the Ojców National Park (South Poland). *Acta Mycol.* 52:1094. doi: 10.5586/am.1094

Baroni, T. J., Beug, M. W., Cantrell, S. A., Clements, T. A., Iturriaga, T., Læssøe, T., et al. (2018). Four new species of 12 *Morchella* from the Americas. *Mycologia* 110, 1205–1221. doi: 10.1080/00275514.2018.1533772

Cao, Y. T., Lu, Z. P., Gao, X. Y., Liu, M. L., Sa, W., Liang, J., et al. (2022). Maximum entropy modeling the distribution area of *Morchella* dill. Ex Pers. species in China under changing climate. *Biology* 11:1027. doi: 10.3390/biology11071027

Che, J., Zhou, W. W., Hu, J. S., and Zhang, Y. P. (2010). Spiny frogs (Paini) illuminate the history of the Himalayan region and Southeast Asia. *Proc. Natl. Acad. Sci.* 107, 13765–13770. doi: 10.1073/pnas.1008415107

Chen, J. J., Bao, K. C., Zhou, L. W., Korhonen, K., and Dai, Y. C. (2015). Phylogeny, divergence time estimation, and biogeography of the genus *Heterobasidion* (Basidiomycota, Russulales). *Fungal Divers.* 71, 185–200. doi: 10.1007/s13225-014-0317-2

Dahlstrom, J. L., Smith, J. E., and Weber, N. S. (2000). Mycorrhiza-like interaction by *Morchella* with species of the Pinaceae in pure culture synthesis. *Mycorrhiza* 9, 279–285. doi: 10.1007/PL00009992

Dai, Y. D., Wu, C. K., Wang, Y. B., Wang, Y., Huang, L. D., Dang, X. J., et al. (2019). Phylogeographic structures of the host insects of *Ophiocordyceps sinensis*. *Zoology* 134, 27–37. doi: 10.1016/j.zool.2019.03.003

Deng, K., Lan, X., Chen, Y., Wang, T., Li, M., Xu, Y., et al. (2022). Integration of transcriptomics and metabolomics for understanding the different vegetative growth in *Morchella sextElata*. *Front. Genet.* 12:829379. doi: 10.3389/fgene.2021.829379

Deng, T., Wu, F., Zhou, Z., and Su, T. (2020). Tibetan plateau: An evolutionary junction for the history of modern biodiversity. *Sci. China Earth Sci.* 63, 172–187. doi: 10.1007/s11430-019-9507-5

Dieter, U., Rafael, S., Michael, W., Markus, J., and Poschmann, A. S. (2022). Wildfires during the Paleogene (late Eocene-late Oligocene) of the Neuwied Basin (W-Germany). *Rev. Palaeobot. Palynol.* 297:104565. doi: 10.1016/j.revpalbo.2021.104565

Ding, L., Kapp, P., Cai, F. C. N., Garzione, N., Xiong, Z. Y., Wang, H. Q., et al. (2022). Timing and mechanisms of Tibetan plateau uplift. *Nat. Rev. Earth Environ.* 3, 652–667. doi: 10.1038/s43017-022-00318-4

Ding, W., Ree, R. H., Spicer, R. A., and Xing, Y. W. (2020). Ancient orogenic and monsoon-driven assembly of the world's richest temperate al▪ pine flora. *Science* 369, 578–581. doi: 10.1126/science.abb4484

Dissanayake, A. A., Mills, G. L., Bonito, G., Rennick, B., and Nair, M. G. (2021). Chemical composition and anti-inflammatory and antioxidant activities of extracts from cultivated morel mushrooms, species of genus *Morchella* (Ascomycota). *Int. J. Med. Mushrooms* 23, 73–83. doi: 10.1615/IntJMedMushrooms.2021039297

Donoghue, M. J. (2008). A phylogenetic perspective on the distribution of plant diversity. *Proc. Natl. Acad. Sci. U. S. A.* 105, 11549–11555. doi: 10.1073/pnas.0801962105

Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry* 19, 11–15.

Drummond, A. J., and Rambaut, A. (2007). BEAST: bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214. doi: 10.1186/1471-2148-7-214

Du, X. H., Wu, D. M., He, G. Q., Wei, W., Xu, N., and Li, T. L. (2019). Six new species and two new records of *Morchella* in China using phylogenetic and morphological analyses. *Mycologia* 111, 857–870. doi: 10.1080/00275514.2019.1640012

Du, X. H., Zhao, Q., O'Donnell, K., Rooney, A. P., and Yang, Z. L. (2012a). Multigene molecular phylogenetics reveals true morels (*Morchella*) are especially species-rich in China. *Fungal Genet. Biol.* 49, 455–469. doi: 10.1016/j.fgb.2012.03.006

Du, X. H., Zhao, Q., Xu, N., and Yang, Z. L. (2016). High inbreeding, limited recombination and divergent evolutionary patterns between two sympatric morel species in China. *Sci. Rep.* 6:22434. doi: 10.1038/srep22434

Du, X. H., Zhao, Q., Yang, Z. L., Hansen, K., Taşkın, H., Büyükalaca, S., et al. (2012b). How well do ITS rDNA sequences differentiate species of true morels (*Morchella*)? *Mycologia* 104, 1351–1368. doi: 10.3852/12-056

Du, X. H., Zhao, Q., and Zhu, L. Y. (2015). A review on research advances, issues, and perspectives of morels. *Mycology* 6, 78–85. doi: 10.1080/21501203.2015.1016561

Guo, J., Xie, Z., Jiang, H., Xu, H., Liu, B., Meng, Q., et al. (2022). The molecular mechanism of yellow mushroom (*Floccularia luteovirens*) response to strong ultraviolet radiation on the Qinghai-Tibet plateau. *Front. Microbiol.* 13:918491. doi: 10.3389/fmicb.2022.918491

Hao, H. B., Zhang, J. J., Wang, H., Wang, Q., Chen, M. J., Juan, J. X., et al. (2019). Comparative transcriptome analysis reveals potential fruiting body formation mechanisms in *Morchella importuna*. *AMB Express* 9:103. doi: 10.1186/s13568-019-0831-4

Harrison, S., and Noss, R. (2017). Endemism hotspots are linked to stable climatic refugia. *Ann. Bot.* 119, 207–214. doi: 10.1093/aob/mcw248

Hewitt, G. M. (2000). The genetic legacy of quaternary ice ages. *Nature* 405:907913. doi: 10.1038/35016000

Hoorn, C., Straathof, J., Abels, H. A., Xu, Y., Utescher, T., and Dupont-Nivet, G. (2012). A late Eocene palynological record of climate change and Tibetan plateau uplift (Xining Basin, China). *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 344-345, 16–38. doi: 10.1016/j.palaeo.2012.05.011

Jia, J., Zheng, W., Zhang, Y., Wei, S., Liang, S., Feng, C., et al. (2022). Tectonic deformation of an intraplate Orogenic Belt: Mesozoic sedimentary basins in the northeastern Qilian Shan. *China. Front. Earth Sci.* 10:847921. doi: 10.3389/feart.2022.847921

Kate, T. (2022). Exceptional preservation of organs in Devonian placoderms from the Gogo lagerstätte. *Science* 337, 1311–1314. doi: 10.1126/science.abf3289Katoh

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kim, T. H., and Kim, J. H. (2022). Molecular phylogeny and historical biogeography of Goodyera R. Br. (Orchidaceae): a case of the vicariance between East Asia and North America. *Front. Plant Sci.* 10:396. doi: 10.3389/fpls.2022.850170

Kim, C., Kim, S. C., and Kim, J. H. (2019). Historical biogeography of melanthiaceae: a case of out-of-North America through the Bering land bridge. *Front. Plant Sci.* 10:396. doi: 10.3389/fpls.2019.00396

Ling, C. C., Ma, F. J., Dong, J. L., Zhou, G. H., Wang, Q. J., and Sun, B. N. (2021). A mid-latitude area in southwestern China experienced a humid subtropical climate with subtle monsoon signatures during the early Oligocene: evidence from the Ningming flora of Guangxi, Palaeogeography, Palaeoclimatology. *Palaeoecology.* 579:110601. doi: 10.1016/j.palaeo.2021.110601

Liu, W., Cai, Y. L., He, P. X., Chen, L. F., and Bian, Y. B. (2019). Comparative transcriptomics reveals potential genes involved in the vegetative growth of *Morchella importuna*. *Biotech* 3:81. doi: 10.1007/s13205-019-1614-y

Liu, Z. F., and Hu, X. M. (2003). Extreme climate events from the cretaceous and Paleogene. *Adv. Earth Sci.* 5, 681–690. doi: 1001-8166(2003)05-0681-10

Liu, H., Song, X., Wen, W., Jia, Q., and Zhu, D. (2022). Quantitative effects of climate change on vegetation dynamics in alpine grassland of Qinghai-Tibet plateau in a county. *Atmos.* 13:324. doi: 10.3390/atmos13020324

Liu, X. H., Xu, Q., and Ding, L. (2016). Differential surface uplift: Cenozoic paleoelevation history of the Tibetan plateau. *Sci. China Earth Sci.* 59, 2105–2120. doi: 10.1007/s11430-015-5486-y

Loizides, M., Alvarado, P., Clowez, P., Moreau, P. A., and Osa, L. R. D. L. (2015). *Morchella tridentina*, *M. rufobrunnea*, and *M. kakiicolor*: a study of three poorly known Mediterranean morels, with nomenclatural updates in section *Distantes*. *Mycol. Prog.* 14, 1–18. doi: 10.1007/s11557-015-1030-6

Loizides, M., Bellanger, J. M., Clowez, P., Richard, F., and Moreau, P. A. (2016). Combined phylogenetic and morphological studies of true morels (Pezizales, Ascomycota) in Cyprus reveal significant diversity, including *Morchella arbutiphila* and *M.disparilis* spp. *Nov. Mycol. Progress.* 15:39. doi: 10.1007/s11557-016-1180-1

Loizides, M., Gonou, Z. Z., Fransuas, G., Drakopoulos, P., Sammut, C., Martinis, A., et al. (2021). Extended phylogeography of the ancestral *Morchella anatolica* supports preglacial presence in Europe and the Mediterranean origin of morels. *Mycologia* 113, 559–573. doi: 10.1080/00275514.2020.1869489

Mao, K. S., Wang, Y., and Liu, J. Q. (2021). Evolutionary origin of species diversity on the Qinghai-Tibet plateau. *J. Syst. Evol.* 59, 1142–1158. doi: 10.1111/jse.12809

Merckx, V. S., Hendriks, K. P., Beentjes, K. K., Mennes, C. B., Becking, L. E., Peijnenburg, K. T., et al. (2015). Evolution of endemism on a young tropical mountain. *Nature* 524, 347–350. doi: 10.1038/nature14949

Miehe, G., Schleuss, P. M., Seeber, E., Babel, W., Biermann, T., Braendle, M., et al. (2019). The *Kobresia pygmaea* ecosystem of the Tibetan highlands-origin, functioning and degradation of the world's largest pastoral alpine ecosystem: Kobresia pastures of Tibet. *Sci. Total Environ.* 648, 754–771. doi: 10.1016/j.scitotenv.2018.08.164

Mittermeier, R. A., Turner, W. R., Larsen, F. W., Brooks, T. M., and Gascon, C. (2011). "Global biodiversity conservation: the critical role of hotspots" in *Biodiversity Hotspots: Distribution and Protection of Conservation Priority Areas*. eds. F. E. Zachos and J. C. Habel (Berlin: Springer), 3–22.

O'Donnell, K., Rooney, A. P., Mills, G. L., Kuo, M., Weber, N. S., and Rehner, S. A. (2011). Phylogeny and historical biogeography of true morels (*Morchella*) reveals an early cretaceous origin and high continental endemism and provincialism in the Holarctic. *Fungal Genet. Biol.* 48, 252–265. doi: 10.1016/j.fgb.2010.09.006

Phonepaserd, P., Yu, Z. D., and Li, T. (2019). High diversity of *Morchella* and a novel lineage of the Esculenta clade from the north Qinling Mountains revealed by GCPSR-based study. *Sci. Rep.* 9, 9–18. doi: 1001-8166(2003)05-0681-10

Qin, Y., Chen, J., and Yi, S. (2015). Plateau pikas burrowing activity accelerates ecosystem carbon emission from alpine grassland on the Qinghai-Tibetan plateau. *Ecol. Eng.* 84, 287–291. doi: 10.1016/j.ecoleng.2015.09.012

Qiu, Y. X., Fu, C. X., and Comes, H. P. (2011). Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of quaternary climate and environmental change in the world's most diverse temperate flora. *Mol. Phylogenet. Evol.* 59, 225–244. doi: 10.1016/j.ympev.2011.01.012

Quan, Q. M., Chen, L. L., Wang, X., Li, S., Yang, X. L., Zhu, Y. G., et al. (2014). Genetic diversity and distribution patterns of host insects of Caterpillar fungus *Ophiocordyceps sinensis* in the Qinghai-Tibet plateau. *PLoS One* 9:92293. doi: 10.1371/ journal.pone.0092293

Rahbek, C., Borregaard, M., Antonelli, A., Colwell, R., Holt, B., Nogués-Bravo, D., et al. (2019). Building mountain biodiversity: geological and evolutionary processes. *Science* 365, 1114–1119. doi: 10.1126/science.aax0151

Richard, F., Sauvé, M., Bellanger, J. M., Clowez, P., Hansen, K., O'Donnell, K., et al. (2015). True morels (*Morchella*, Pezizales) of Europe and North America: evolutionary relationships inferred from multilocus data and a unified taxonomy. *Mycologia* 107, 359–382. doi: 10.3852/14-166

Ronquist, F., Teslenko, M., Mark, P. V. D., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

Sanmartín, I., Enghoff, H., and Ronquist, F. (2001). Patterns of animal dispersal, vicariance and diversification in the Holarctic. *Biol. J. Linn. Soc.* 73, 345–390. doi: 10.1006/bijl.2001.0542

Shi, Y. F., Li, J. J., and Li, B. Y. (1998). *Uplift and Environmental Changes of Qinghai-Tibetan Plateau in the Late Cenozoic*. Guangzhou: Guangdong Science and Technology Press.

Shrestha, B., Zhang, W. M., Zhang, Y. J., and Liu, X. Z. (2010). What is the Chinese caterpillar fungus *Ophiocordyceps sinensis* (*Ophiocordy cipitaceae*)? *Mycology* 1, 228–236. doi: 10.1080/21501203.2010.536791

Spicer, R. A., Su, T., Valdes, P. J., Farnsworth, A., Wu, F. X., Shi, G. l., et al. (2021). Why 'the uplift of the Tibetan plateau' is a myth. *Natl. Sci. Rev.* 8:nwaa091. doi: 10.1093/nsr/nwaa091

Sun, H., Zheng, D., Yao, T., and Zhang, Y. (2012). Protection and construction of the national ecological security shelter zone on Tibetan plateau. *Acta Geograph. Sin.* 67, 3–12. doi: 10.11821/xb201201001

Suzuki, H., Sato, J. J., Tsuchiya, K., Luo, J., Zhang, Y. P., Wang, Y. X., et al. (2003). Molecular phylogeny of wood mice (Apodemus, Muridae) in East Asia. *Biol. J. Linn. Soc.* 80, 469–481. doi: 10.1046/j.1095-8312.2003.00253.x

Swofford, D. L. (2002). PAUP*: Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4.0b10. *Sinauer Associates.* doi: 10.1111/ j.0014-3820.2002.tb00191.x

Taylor, T. N., Hass, H., Kerp, H., Krings, M., and Hanlin, R. T. (2004). Perithecial ascomycetes from the 400 million-year-old Rhynie chert: an example of ancestral polymorphism. *Mycologia* 96, 1403–1419. doi: 10.1080/15572536.2006.11832862

Wan, W., Xiao, P. F., Feng, X. Z., Ma, R. H., Duan, H. T., and Zhao, L. M. (2014). Monitoring lake changes of Qinghai-Tibetan plateau over the past 30 years using satellite remote sensing data. *Chin. Sci. Bull.* 59, 1021–1035. doi: 10.1007/s11434-014-0128-6

Wang, L. Y., Abbott, R. J., Zheng, W., Chen, P., Wang, Y. J., and Liu, J. Q. (2009). History and evolution of alpine plants endemic to the Qinghai-Tibetan plateau: *Aconitum gymnandrum* (Ranunculaceae). *Mol. Ecol.* 18, 709–721. doi: 10.1111/j.1365-294X.2008.04055.x

Wang, M. S., Li, Y., Peng, M. S., Zhong, L., Wang, Z. J., Li, Q. Y., et al. (2015). Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. *Mol. Biol. Evol.* 32, 1880–1889. doi: 10.1093/molbev/msv071

Wang, L., Yuan, Q., Shen, L., and Ding, L. (2022). Middle Eocene Paleoenvironmental reconstruction in the Gonjo Basin, Eastern Tibetan Plateau: evidence from palynological and evaporite records. *Front. Earth Sci.* 10, 1–14.doi: 10.3389/feart.2022.818418

Wang, C., Zhao, X., Liu, Z., Lippert, P. C., Graham, S. A., Coe, R. S., et al. (2008). Constraints on the early uplift history of the Tibetan plateau. *PNAS* 105, 4987–4992. doi: 10.1073/pnas.0703595105

Wen, J., Zhang, J. Q., Nie, Z. L., Zhong, Y., and Sun, H. (2014). Evolutionary diversification of plants on the Qinghai-Tibetan plateau. *Front. Genet.* 5:4. doi: 10.3389/fgene.2014.00004

Wu, H., Chen, J., Li, J., Liu, Y., Park, H. J., and Yang, L. (2021). Recent advances on bioactive ingredients of *Morchella esculenta*. *Appl. Biochem. Biotechnol.* 193, 4197–4213. doi: 10.1007/s12010-021-03670-1

Wu, Y. D., Li, L., Fan, Y. L., Ni, X. W., Ohiolei, J. A., Li, W. H., et al. (2021). Genetic evolution and implications of the mitochondrial genomes of two newly identified *Taenia* spp. in rodents from Qinghai-Tibet plateau. *Front. Microbiol.* 12:647119. doi: 10.3389/fmicb.2021.647119

Wu, F. X., Miao, D. S., Chang, M. M., Shi, G. L., and Wang, N. (2017). Fossil climbing perch and associated plant megafossils indicate a warm and wet Central Tibet during the late Oligocene. *Sci. Rep.* 7:878. doi: 10.1038/s41598-017-00928-9

Xiong, Z. Y., Liu, X. H., Ding, L., Alex, F., Robert, A. S., Xu, Q., et al. (2022). The rise and demise of the Paleogene Central Tibetan Valley. *Sci. Adv.* 8, 1–15. doi: 10.1126/sciadv.abj0944

Yan, Y. J., Li, Y., Wang, W. J., He, J. S., Yang, R. H., and Yao, Y. J. (2017). Range shifts in response to climate change of *Ophiocordyceps sinensis*, a fungus endemic to the Tibetan plateau. *Biol. Conserv.* 206, 143–150. doi: 10.1016/j.biocon.2016.12.023

Yang, P. (2005). The main environmental effect of Qinghai-Tibet plateau uplift. *Qinghai Env.* 3, 93–95. doi: 1007-2454(2005)03-0093-03

Yang, S. J., Dong, H. L., and Lei, F. M. (2009). Phylogeography of regional fauna on the Tibetan plateau: a review. *Prog. Nat. Sci.* 19, 789–799. doi: 10.1016/j.pnsc.2008.10.006

Yang, F. S., Li, Y. F., Ding, X., and Wang, X. Q. (2008). Extensive population expansion of *Pedicularis longiflora* (Orobanchaceae) on the Qinghai-Tibetan plateau and its correlation with the quaternary climate change. *Mol. Ecol.* 17, 5135–5145. doi: 10.1111/j.1365-294X.2008.03976.x

Yatsiuk, I., Saar, I., Kalamees, K., Sulaymonov, S., Gafforov, Y., and O'Donnell, K. (2016). Epitypification of *Morchella steppicola* (Morchellaceae, Pezizales), a morphologically, phylogenetically and biogeographically distinct member of the Esculenta clade from Central Eurasia. *Phytotaxa.* 284, 31–40. doi: 10.11646/phytotaxa.284.1.3

Yu, Y., Harris, A. J., Blair, C., and He, X. J. (2015). RASP (reconstruct ancestral state in phylogenies): a tool for historical biogeography. *Mol. Phylogenet. Evol.* 87, 46–49. doi: 10.1016/j.ympev.2015.03.008

Yu, F. M., Jayawardena, R. S., Thongklang, N., Lv, M. L., Zhu, X. T., and Zhao, Q. (2022). Morel production associated with soil nitrogen-fixing and nitrifying microorganisms. *J. Fungi.* 8:299. doi: 10.3390/jof8030299

Yuan, F. (2015). Study on Phylogeography and suitable distribution of Ophiocordycepssinensispopulations. Ph. D. dissertation. Yunnan: Yunnan University.

Zachos, J., Pagani, M., Sloan, L., Thomas, E., and Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292, 686–693. doi: 10.1126/science.1059412

Zhang, D., Gao, F., Jakovlić, I., Zou, H., Zhang, J., Li, W. X., et al. (2020). PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.* 20, 348–355. doi: 10.1111/1755-0998.13096

Zhao, H., Zhu, J., Zong, T., Liu, X., Ren, L., Lin, Q., et al. (2021). Two new species in the family Cunninghamellaceae from China. *Mycobiology* 49, 142–150. doi: 10.1080/12298093.2021.1904555

Zhu, L., Song, J., Zhou, J. L., Si, J., and Cui, B. K. (2019). Species diversity, phylogeny, divergence time, and biogeography of the genus *Sanghuangporus* (Basidiomycota). *Front. Microbiol.* 10:812. doi: 10.3389/fmicb.2019.00812

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 10 January 2023
DOI 10.3389/fmicb.2022.1100249

OPEN ACCESS

EDITED BY
Feng Gao,
Tianjin University, China

REVIEWED BY
Pratick Khara,
University of Texas Health Science Center at Houston, United States
Shun Adachi,
Independent Researcher, Uji, Japan
Preeti Srivastava,
Indian Institute of Technology Delhi, India

*CORRESPONDENCE
Damien P. Devos
✉ damienpdevos@gmail.com
Carlos Santana-Molina
✉ csantmol@gmail.com

†PRESENT ADDRESS
DMaría del Saz-Navarro,
Intelligent Data Analysis Group (DATAi), Pablo de Olavide University, Seville, Spain

SPECIALTY SECTION
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 16 November 2022
ACCEPTED 16 December 2022
PUBLISHED 10 January 2023

CITATION
Santana-Molina C, Saz-Navarro D and Devos DP (2023) Early origin and evolution of the FtsZ/tubulin protein family.
Front. Microbiol. 13:1100249.
doi: 10.3389/fmicb.2022.1100249

Early origin and evolution of the FtsZ/tubulin protein family

Carlos Santana-Molina[1,2]*, DMaría del Saz-Navarro[1†] and Damien P. Devos[1]*

[1]Centro Andaluz de Biología del Desarrollo, Consejo Superior de Investigaciones Científicas, Junta de Andalucía, Universidad Pablo de Olavide, Seville, Spain, [2]Department of Marine Microbiology and Biogeochemistry, Royal Netherlands Institute for Sea Research (NIOZ), University of Utrecht, Utrecht, Netherlands

The origin of the FtsZ/tubulin protein family was extremely relevant for life since these proteins are present in nearly all organisms, carrying out essential functions such as cell division or forming a major part of the cytoskeleton in eukaryotes. Therefore, investigating the early evolution of the FtsZ/tubulin protein family could reveal crucial aspects of the diversification of the three domains of life. In this study, we revisited the phylogenies of the FtsZ/tubulin protein family in an extensive prokaryotic diversity, focusing on the main evolutionary events that occurred during its evolution. We found evidence of its early origin in the last universal common ancestor since FtsZ was present in the last common ancestor of Bacteria and Archaea. In bacteria, *ftsZ* genes are genomically associated with the bacterial division gene cluster, while in archaea, *ftsZ* duplicated prior to archaeal diversification, and one of the copies is associated with protein biosynthesis genes. Archaea have expanded the FtsZ/tubulin protein family with sequences closely related to eukaryotic tubulins. In addition, we report novel CetZ-like groups in Halobacterota and Asgardarchaeota. Investigating the C-termini of prokaryotic paralogs basal to eukaryotic tubulins, we show that archaeal CetZ, as well as the plasmidic TubZ from Firmicutes, most likely originated from archaeal FtsZ. Finally, prokaryotic tubulins are restricted to Odinarchaeaota and *Prosthecobacter* species, and they seem to belong to different molecular systems. However, their phylogenies suggest that they are closely related to α/β-tubulins pointing to a potential ancestrality of these eukaryotic paralogs of tubulins.

KEYWORDS

FtsZ, tubulin, CetZ, evolution, tree of life

Introduction

Members of the FtsZ/tubulin protein family are small GTPases that orchestrate essential cellular processes such as cell division in prokaryotes and eukaryotes or establishing the cytoskeleton in eukaryotes (Wagstaff and Löwe, 2018). These proteins are present in most organisms in the three domains of life, with few exceptions.

Therefore, their omnipresence and fundamental functions in living organisms indicate the relevance of the FtsZ/tubulin protein family for the origin and diversification of life.

In prokaryotes, FtsZ polymerizes to form a ring in the septal region of dividing cells, recruiting and cooperating with other proteins to carry out the process of division (Bi and Lutkenhaus, 1991). In Bacteria, FtsZ cooperates with the so-called bacterial divisome, which comprises proteins also found in Archaea such as FtsA/Mreb or SepF, and other proteins exclusively found in bacteria, which are mostly involved in peptidoglycan metabolism (whose genes form the *dcw* gene cluster; Ayala et al., 1994). Archaea usually present two ancestral copies of FtsZ that appear to have non-overlapping functions (Liao et al., 2021). The cell division process in Archaea is not as well understood as in Bacteria, but there are some similarities with the bacterial machinery, for example, FtsA and SepF (Pende et al., 2021; van Wolferen et al., 2022). Alternatively, some archaea that do not have FtsZ appear to have a cell division system based on ESCRT-III homologs (Caspi and Dekker, 2018; Pulschen et al., 2020). However, there are specific cases in prokaryotes that do not have FtsZ (nor ESCRT-III), and whose cell division mechanisms remain unknown (Lluch-Senar et al., 2010; Rivas-Marín et al., 2016; Ithurbide et al., 2022). On the other side, the tubulins constitute one of the main components of the eukaryotic cytoskeleton, which has been expanded in different paralogs with specific functions (Janke and Magiera, 2020). α- and β-tubulins heterodimerize, forming the canonical microtubules, while γ-tubulin is involved in their nucleation. δ-, ε-, and ζ-tubulins constitute the centriole controlling chromosome segregation during cell division. Thus, the members of the FtsZ/tubulin protein family have diversified functionally across the three domains of life, but they retain essential functions in cell division.

To provide a global overview of the evolution of the FtsZ/tubulin protein family, we reconstructed a phylogeny of representative sequences representing the diversity of subfamilies, their relationships, and their domain architecture (**Figure 1**). The FtsZ/tubulin protein family is mainly composed of prokaryotic FtsZ [including mitochondrial (Leger et al., 2015) and chloroplastic (TerBush et al., 2013) ones], eukaryotic tubulins [including prokaryotic sequences from Asgardarchaeota (Akıl et al., 2022) and Verrucomicrobia (Schlieper et al., 2005)], archaeal paralogs closely related to tubulins [including CetZ (Duggin et al., 2015), artubulins (Yutin and Koonin, 2012), and other bacterial and archaeal sequences not included in this tree such as the plasmidic TubZ found in Firmicutes (Larsen et al., 2007)], and divergent homologs bearing the tubulin_2 Pfam domain, which were used to root the tree. This latter group, here called FtsZ-like, displayed diverse domain composition and poorly aligned positions, and given its sparse distribution in prokaryotes, domain architecture, and genomic synteny, its evolutionary history is unlikely to be directly related to the one of the FtsZ/tubulin protein family (Makarova and Koonin, 2010). Similarly, the divergent eukaryotic proteins related to tubulins called Misato (not included in this tree) appear to be a eukaryotic subfunctionalization given that they show various motif extensions (Mottier-Pavie et al., 2011; Palumbo et al., 2015).

The domain architecture of these proteins shows a conserved N-terminus, while the C-termini have diverged between the subfamilies: FtsZ sequences have the FtsZ_C domain, and tubulins have tubulin_C domain (not in all paralogs). Remarkably, the archaeal paralogs basal to tubulins, including CetZ, did not present any recognizable Pfam domain. These observations on the C-terminus domains suggest a dynamic evolutionary region whose evolution could clarify the relationships between the members of these protein families. This is important because the evolutionary relationships between some of these prokaryotic proteins and eukaryotic tubulins are still debated.

Given the relevance of the FtsZ/tubulin protein family, a comprehensive understanding of its evolution would provide crucial insights into the diversification of the three domains of life. FtsZ is expected to have originated in the last universal common ancestor (LUCA) since it is in nearly all Bacteria and Archaea (Davis, 2002), both forming the two primary domains of life. On the other hand, tubulins are expected to have originated from multiple duplications during eukaryogenesis between the first eukaryotic common ancestor (FECA) and the last eukaryotic common ancestor (LECA), resulting in the contemporary α, β, δ-, ε-, and ζ-tubulins [possibly among others; (Findeisen et al., 2014)]. However, to the best of our knowledge, questions about the origin of tubulins and the order of their duplications are still limited. The identification of artubulins found in the *Nitrosoarcheaum* genus led to the suggestion of the archaeal origin of tubulins (Yutin and Koonin, 2012). However, posterior analyses suggested alternative scenarios given the phylogenetic relationships of artubulins with TubZ (Findeisen et al., 2014). Therefore, there are some controversies around the origin of some of these proteins that need to be addressed. In addition, greater taxonomic diversity and improved knowledge about the tree of life (Coleman et al., 2021; Moody et al., 2022) will help to decipher the main events during the evolution of the FtsZ/tubulin protein family.

In this study, we revisit the evolutionary history of the FtsZ/tubulin protein family, focusing on the main events in its evolution: (i) the early evolution of FtsZ in Bacteria and Archaea, (ii) the origins of archaeal paralogs found basal to tubulins based on their C-termini, and (iii) the relationships between eukaryotic tubulins and prokaryotic tubulins. Addressing these conundrums, we provide evidence for the early origin and evolution of this protein family and clarify some relationships of archaeal paralogs, showing that the C-termini of CetZ and TubZ relate their origin with archaeal FtsZ.

Results

FtsZ was present in the last bacterial common ancestor and was genomically associated with the bacterial divisome

Two datasets were built, one including bacterial and archaeal FtsZs (BA reconstruction) and the other including only the bacterial sequences (B reconstruction). After removing redundancy by taxonomic classes and spurious sequences, the final multiple sequence alignments (MSAs) were treated with three different trimming methods, referred to as Manual-, BMGE-, and trimAl-gt70-trimming (refer to section "Materials and methods"). Subsequently, we performed phylogenetic reconstructions using empirical models automatically selected by Modelfinder (Kalyaanamoorthy et al., 2017) and complex models (LG4X and C20 + R + F). We used an ultrafast bootstrap for manual-trimmed reconstructions and nearest neighbor interchange (NNI) search analyses (Minh et al., 2020) for all the MSAs.

In the resulting reconstructions, most bacterial phyla show a monophyletic pattern, although the topology was sensitive to different parameters, such as taxonomic sampling, sequence redundancy threshold, and phylogenetic

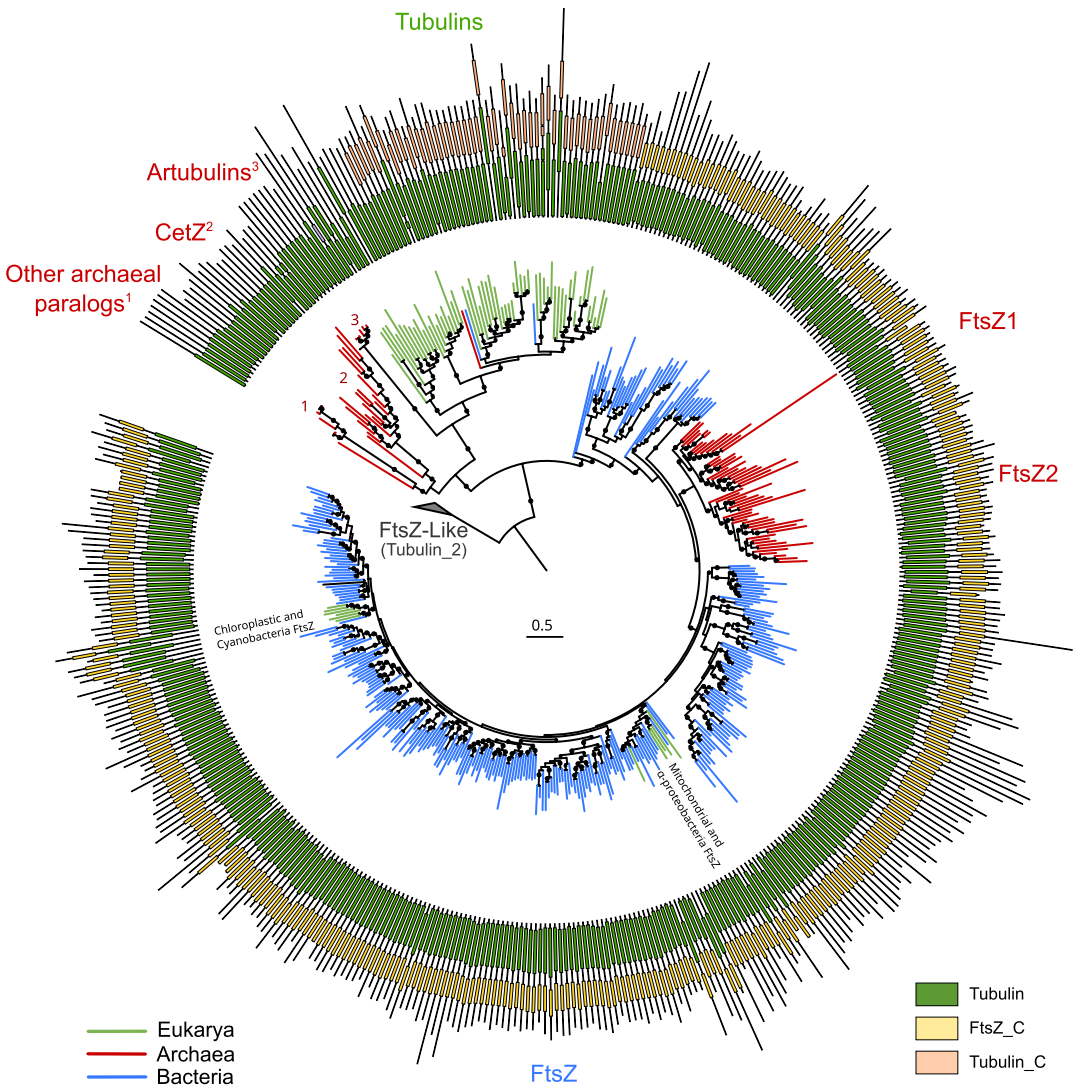


FIGURE 1
Overview of the FtsZ/tubulin protein family. Schematic phylogeny of representative sequences showing their Pfam domain architectures. Branches are colored according to the taxonomic domain, and black circles indicate bootstraps higher than 90. The tree was rooted at the divergent FtsZ-like sequences containing the Tubulin_2 Pfam domain. Note that some sequences discussed in this study are not shown in this tree (e.g., TubZ and other archaeal paralogs).

methods (Supplementary Figure 1). Basal nodes were weakly supported (<90%) in most reconstructions, except for UFBoot2 reconstructions. This low support in basal nodes can be explained by the antiquity and short length of these proteins (MSAs of 350 amino acids applying inclusive trimming). Therefore, despite the fact that UFBoot2 values lower than 95% should not be considered reliable nodes (Hoang et al., 2018), we interpreted these reconstructions with a relaxed threshold of confidence as the biological relationship between species is limited when collapsing nodes with 95% UFBoot2 values (Supplementary Figure 2). On the contrary, collapsing at 85% resulted in biologically meaningful clades (Figure 2 and refer to Supplementary Figures 2, 3 for an extended version).

Taking the Manual + UFBoot2 + MF reconstruction as a reference, the phylogeny was better resolved in the B reconstruction than in the BA reconstruction since the basal nodes better resolved the relationships between bacterial phyla (specifically in Gracilicutes, Supplementary Figure 2). This suggests that the addition of archaeal sequences promotes long branch attraction (LBA) artifacts, providing weaker topologies. This was observed in our reconstructions where archaeal FtsZ sequences attracted different bacterial FtsZ using various trimming methods and evolutionary models (refer to BA reconstructions in Supplementary Figure 1). Thus, B reconstruction was selected as a reference (Figure 2).

The bacterial FtsZ tree recovered the two main bacterial supergroups, Terrabacteria and Gracilicutes (Figure 2; Coleman et al., 2021). Between these two supergroups, intermediary branches are found, including Elusimicrobiota, Spirochaetota, Fusobacteriota, Synergistota, Armatimonadota, and Deinococcota phyla, or other phyla such as Firmicutes and Bacteroidota, forming paraphyletic groups and showing unstable locations in our reconstructions (Supplementary Figure 1). The FtsZ of Candidate Phyla Radiation (CPR, denoted as Patescibacteria in this GTDB version) is phylogenetically associated with Chloroflexota, consistent with their proposed origin (Coleman et al., 2021); however, this relationship was only observed in Manual-trimming reconstructions. Some CPR genomes also contain a divergent copy of FtsZ, which usually branches with Verrucomicrobiota, which are the most divergent groups (Figure 2 and Supplementary Figure 1). The verrucomicrobial FtsZ and the second copy of CPR had different topologies in the BA and B reconstructions (Supplementary Figures 1, 2), demonstrating their phylogenetic instability and suggesting strong divergence and artifacts usually associated with it, such as LBA. The divergent CPR FtsZ (together with that of Verrucomicrobiota) is nested in the canonical CPR FtsZ group in the BMGE-B-bnni reconstruction (Supplementary Figure 1), suggesting that it originated from the original FtsZ of CPR. In contrast, in the B reconstruction (refer to Figure 2 for UFBoot2 and Supplementary Figure 1 for NNI analysis), Verrucomicrobiota branches with other Gracilicutes near to Bacteroidetes (and other close relatives such as Cloacimonadota and WOR-3), consistent with their species relationship and supporting the idea that reconstructions without archaeal sequences solve more adequately bacterial evolution. Therefore, despite the phylogenetic instabilities or strong sequence divergence, the separation of Gracilicutes and Terrabacteria in most reconstructions (Supplementary Figures 1, 2 and in Manual-B-UFboot2 reconstruction supported by 95%; Figure 2) strongly suggests that FtsZ was present in the last bacterial common ancestor (LBCA).

We then inspected the genomic association of all these prokaryotic FtsZs (Figure 3). The most conserved Pfam domains present in the genes surrounding bacterial *ftsZs* are related to cell division, for example, FtsA, FtsW, or FtsQ, or peptidoglycan synthesis, for example, murein ligase, glycosyl transport, MreB, or MraY (Figure 3A), which form the *dcw* cluster. This genome context is well conserved in Gracilicutes, Terrabacteria, and CPR, showing that the *dcw* gene cluster, was already established in the LBCA. In the few bacteria having two copies of the *ftsZ* gene, such as Bipolaricuta, the additional copy is located outside the *dcw* cluster (some of them in conserved *loci*), indicating possible subfunctionalization due to gene duplication.

FtsZ duplicated in the last archaeal common ancestor and was genomically associated with protein biosynthesis genes

The genomes of most archaeal species present two copies of *ftsZ*, coding for FtsZ1 and FtsZ2, with the exception of Crenarchaeota archaea that mostly retained FtsZ2, or present divergent FtsZs branching outside of the canonical archaeal FtsZ groups (Figure 4 and Supplementary Figure 4, for an extended view). Alternatively, other Crenarchaeota classes do not present any FtsZ, and in the case of a few Euryarchaeota and a few Asgardarchaeota, only one copy of the gene was found, as previously reported (Pende et al., 2021).

The Crenarchaeota and Asgardarchaeota sequences are closely related (Figure 4), in agreement with their species relationships. The Euryarchaeota sequences (according to GTDB) tend to form monophyletic groups, but they are not monophyletic with other close relatives such as Halobacterota and Thermoplasmatota. However, Hydrothermarchaeota and Hadarchaeota sequences are closely related to Euryarchaeota (Figure 4), which is in agreement with the species' relationships. The DPANN supergroup branches paraphyletically: Micrarchaeota and Altiarchaeota branch close to the Euryarchaeota supergroup but distally to other DPANN such as Nanoarchaeota and Huberarchaeaota that show different topologies of the FtsZ1 and FtsZ2 branches.

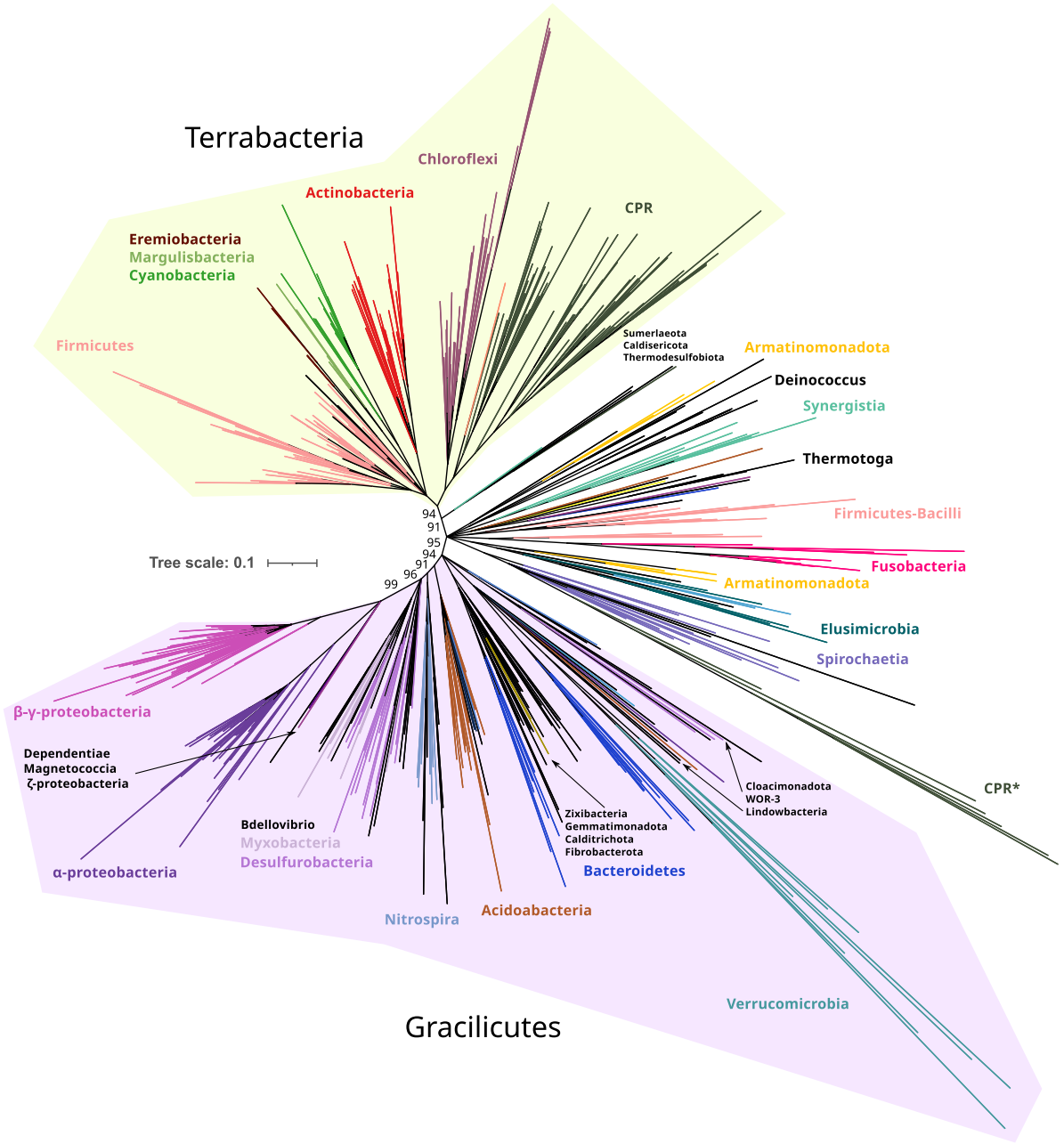


FIGURE 2
Phylogeny of only bacterial FtsZs. Branches are colored according to taxonomic phyla and GTDB nomenclature (release bac120, arc122). The dataset was reduced by redundant filtering by taxonomic classes and then selecting one sequence for each taxonomic order. The final MSA consisted of 534 sequences and 334 positions, denoted as manual trimming. The phylogeny was built using IQ-TREE, automatic model selection (LG + R10), and ultrafast bootstrap (UFBoot2). Refer to **Supplementary Figure 1**, for corrected bootstrap trees with NNI searches, different trimming methods, and evolutionary models. To ease the visualization of the most basal supported nodes, nodes with lower support than 85% were removed and only UFBoot2 values at basal nodes of interest were shown. Terrabacteria and Gracilicutes supergroups are highlighted in yellow and purple, respectively (refer to the extended tree in **Supplementary Figure 3**). Original and trimmed analyses and Newick trees are provided in **Supplementary Data**.

Acknowledging the low resolution in some basal nodes, and as previously suggested (Pende et al., 2021), the monophyletic branching of most archaeal phyla, combined with topological similarities with the archaeal species tree, suggests that FtsZ1 and FtsZ2 are derived from an ancestral duplication in the last archaeal common ancestor (LACA). In addition, FtsZ evolution in archaea is irregular as some archaeal groups have lost one or both FtsZ copies, for example, Crenarchaeota. However,

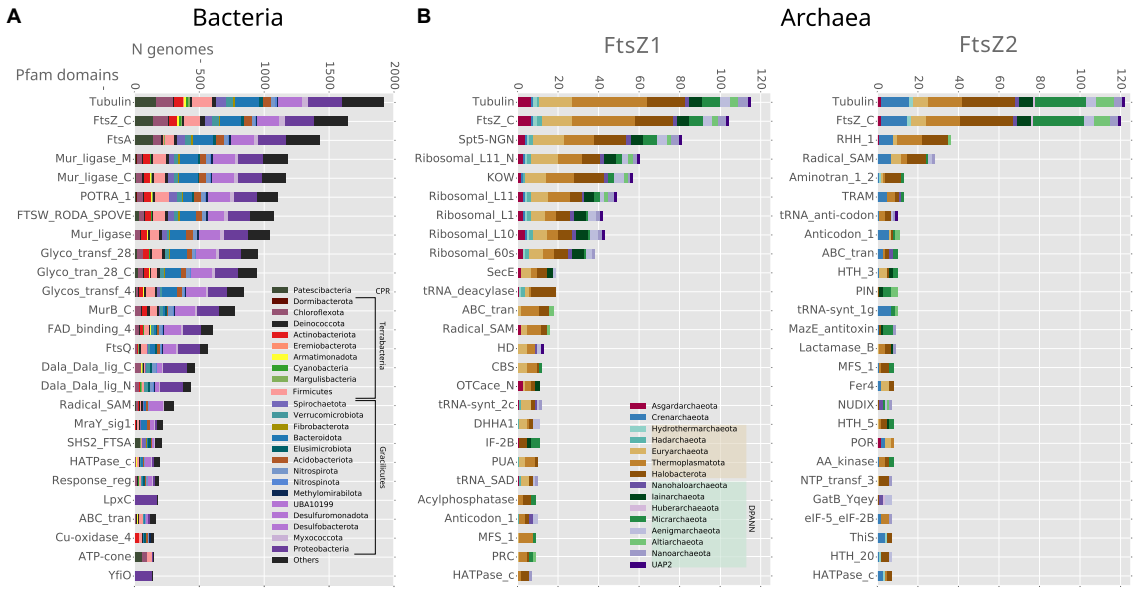


FIGURE 3
Genomic context analyses of the prokaryotic *ftsZ* genes. Analyses of **(A)** bacterial and **(B)** archaeal *ftsZ* genes, showing the 25 Pfam domains most abundant in their genome context; 10 Kb up and downstream of the *ftsZ* gene. Bars denote the number of genomes of each phylum, having at least one Pfam domain in the respective genome context. Note that some Pfam domains belong to the same proteins, such as Tubulin-FtsZ_C, Potra_1-FtsQ, and Mur_ligase_M/C.

others might have regained FtsZ by lateral gene transfer (LGT) from other archaea, as shown by suspicious phylogenetic groups in FtsZ2, such as Micrarchaeota and Altiarchaeota branching within Euryarcheaota, or the monophyletic branching of Crenarchaeota and Halobacterota sequences (refer to question marks in **Figure 4**), both of which could represent phylogenetic artifacts (assuming that FtsZ has been inherited vertically between the species) or actual LGT events (as suggested by the latter case with full UFBoot2 support). Therefore, although vertical inheritance seems to be the main mode of evolution of FtsZ in archaea, LGT events between them should not be discarded. However, it is important to remark that we did not detect notorious LGT of *ftsZ* genes between archaea and bacteria.

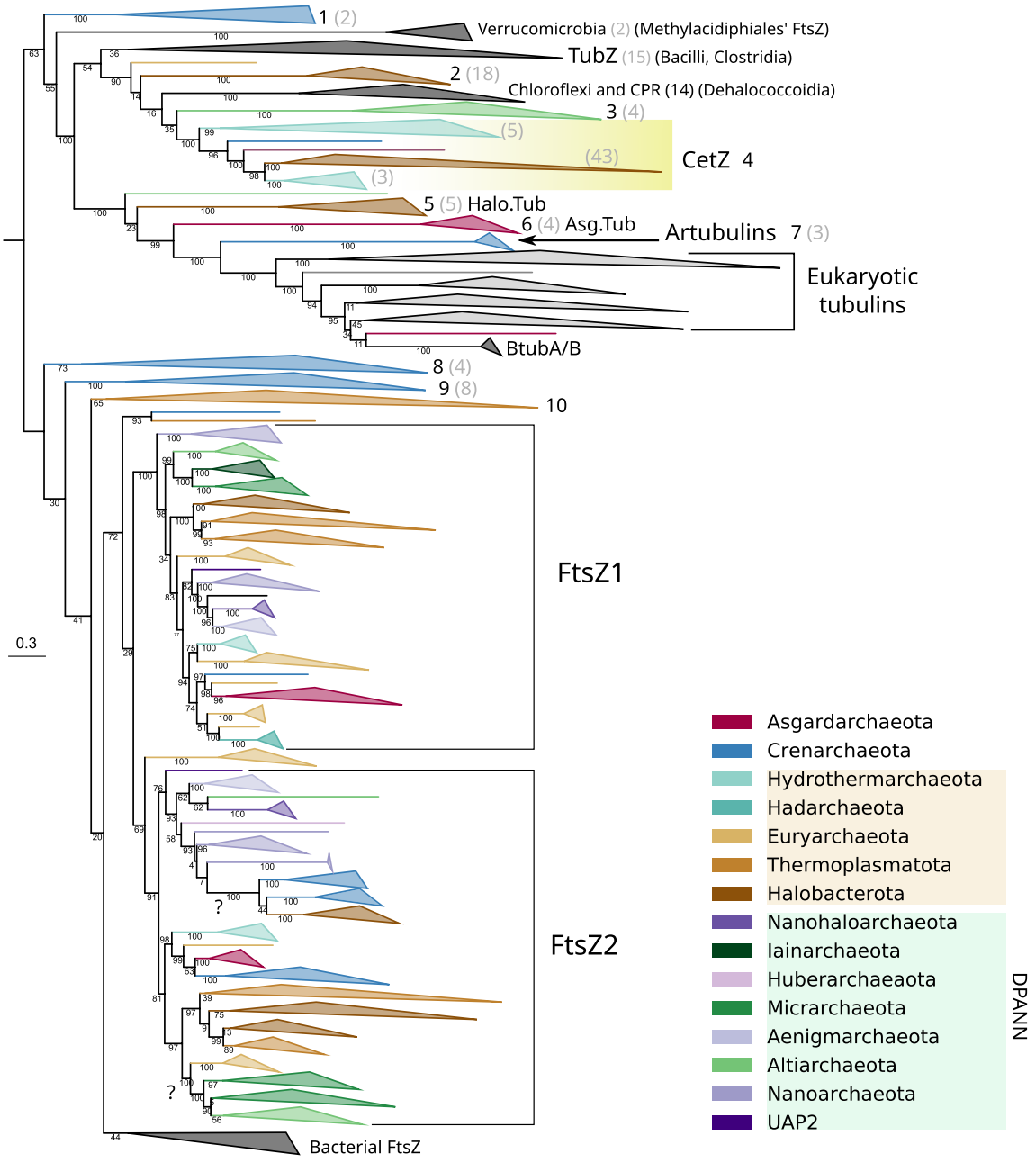
The genomic contexts of the archaeal *ftsZ1* and *ftsZ2* genes are different from those of their bacterial counterparts, but they are also different from each other (**Figure 3B**). The gene *ftsZ1* has a conserved genome context, whereas the one of the *ftsZ2* is not as well conserved (**Figure 3B**). In contrast to what is observed in bacteria, the domains associated with genes found in the genome context of archaeal *ftsZ* are related to informational housekeeping genes, mainly related to protein biosynthesis, including ribosomal and t-RNA genes (especially for the genome context of *ftsZ1* but also in Crenarchaeota *ftsZ2*; **Figure 3B**). The genomic contexts of *ftsZ2* usually contain a gene bearing the RHH_1 Pfam domain, which probably functions as a transcriptional regulator of the CopG family (Acebo et al., 1998), suggesting that the transcriptional expression of the *ftsZ2* gene cluster could trigger a specific transcriptional regulation response.

Homology at the C-termini suggests that CetZ and TubZ originated from archaeal FtsZ

Archaeal genomes often reflect functional expansions of the FtsZ/tubulin proteins (refer to arcs in **Figure 5**) as exemplified by the well-known CetZ and others, such as artubulins (Yutin and Koonin, 2012) and OdinTubulin from Odinarchaeota (Zaremba-Niedzwiedzka et al., 2017). In addition, we also identified other archaeal paralogs found in Asgardarchaeota (Asg.Tub) and another one in Halobacterota (Halo.Tub); although the phylogenetic positions of both were unstable, branching with tubulins (**Figure 4**) or with CetZ sequences (**Figure 6**; further discussed later).

CetZ was likely present in the Euryarchaeota ancestor as it is found in most classes, such as Archaeoglobi, Methanomicrobia, Halobacteria, Syntrophoarchaeia, and Hydrothermarchaeota (**Figure 5**). We found other CetZ-like proteins, including sequences from Halobacterota, Altiarchaeaota (DPANN), and sparse paraphyletic groups of Chloroflexota–Dehalococcoidia and Firmicutes–Bacilli/Clostridia (the latter was identified as TubZ; **Figure 4**).

As no recognizable domains were detected in the C-termini of the CetZ sequences (**Figure 1**), we attempted to clarify



1) Nitrososphaeria-Bathyarchaeia
2) Methanonatronarchaeia-Methanosarcinia
3) Altiarchaeia
4) CetZ: Altiarchaeia; Hydrothermarchaeia; Syntrophoarchaeia, Archaeoglobi, Methanomicrobia, Halobacteria; Bathyarchaeia
5) Halobacteria
6) Heimdallarchaeia, Lokiarchaeia
7) Artubulin: Nitrososphaeria
8) Bathyarchaeia
9) Korarchaeia, Nitrososphaeria
10) Thermoplasmata, Poseidoniia

FIGURE 4
Phylogeny of all members of the FtsZ/tubulin protein family centered on archaeal sequences. The tree was rooted at the mid-point. Branches are colored according to taxonomic phyla following GTDB nomenclature (release bac120, arc122). The dataset was constructed by a selection of representative bacterial FtsZs and eukaryotic tubulins, plus an archaeal selection based on redundancy filtering and adding back removed paralogs from the remaining taxa. The final MSA consisted of 1,637 sequences and 410 positions, trimmed with trimAl -gt 0.2. The phylogeny was built using IQ-TREE, automatic model selection (LG + R10), and ultrafast bootstrap (UFBoot2). Numbers indicate the taxonomic composition of each phylogenetic group shown at the bottom, and gray numbers between parenthesis are the number of sequences for the respective group. Refer to the extended tree in **Supplementary Figure 4**.

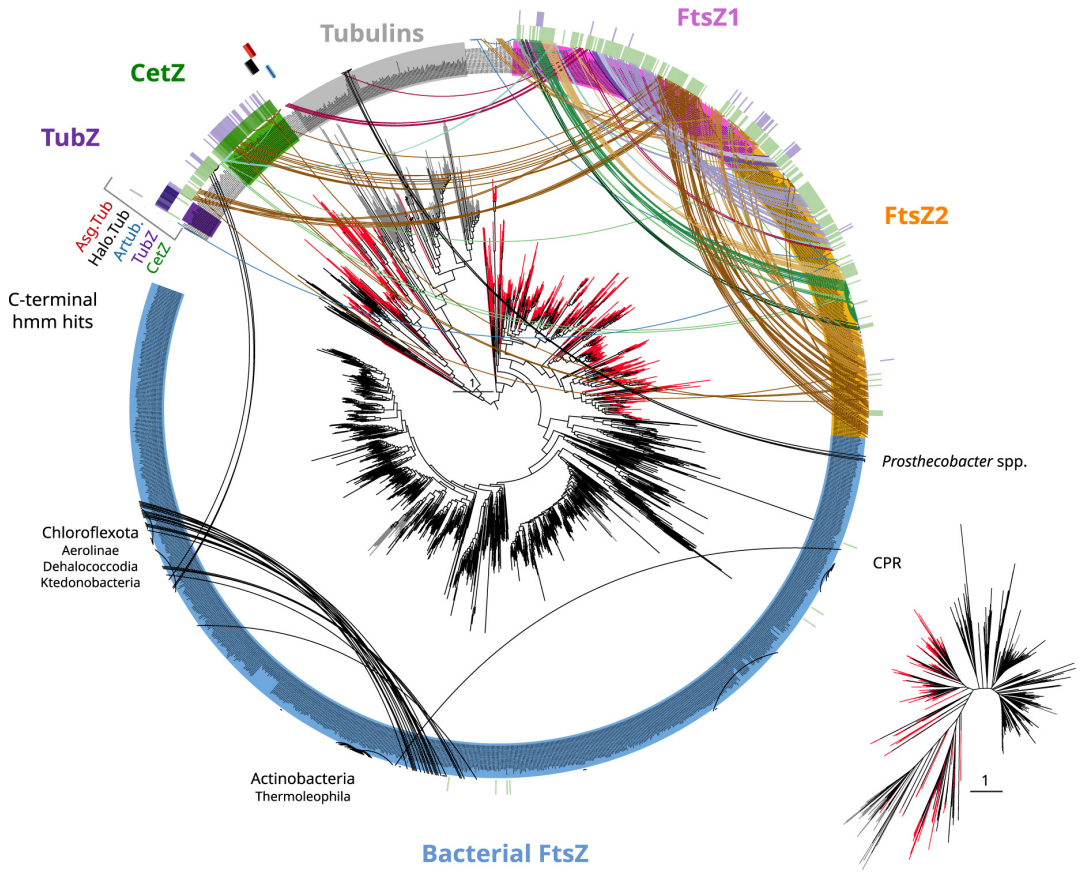


FIGURE 5
Tubulin/FtsZ-like paralogs in archaea and C-termini domain homologies. The phylogeny of **Figure 4** was used to map the hits of HMMSEARCH of the C-termini of different phylogenetic groups (external colored heatmaps): CetZ, TubZ, artubulins, Halo.Tub, and Asg.Tub, showing the relationship between CetZ and archaeal FtsZ, as well as the overlap between CetZ and TubZ. Branches are colored by domain: black bacteria, red archaea, and gray eukaryota. Arcs connect sequences from the same organism showing the abundance and taxonomic diversity of FtsZ paralogs in archaea. Arc colors follow the legend of **Figure 4**. Inset represents an unrooted tree of the same phylogeny. Refer to the extended tree for annotations in **Supplementary Figure 4**.

their evolutionary relationships by constructing protein models (hidden Markov model, HMM) of the C-termini of these proteins and performing hmmsearches against all our tubulin/FtsZ protein family collection (**Supplementary Data**). This CetZ_C HMM showed significant sequence similarities (e-value approximately $1e^{-10}$) with the other CetZ-like proteins (including bacterial Chloroflexota but not Firmicutes TubZ sequences) but also weaker similarities (e-value approximately $1e^{-3}$) with FtsZ, in particular with the archaeal sequences (see external green heatmap in **Figure 5**). CetZ_C HMM does not match with tubulins or archaeal tubulin-like proteins nor bacterial FtsZ. This possibility is more evident when visualizing the same tree unrooted (inset **Figure 5**). In this view, the closer proximity between the archaeal FtsZ sequences and tubulin-CetZ sequences is observed. Therefore, this phylogenetic signal, together with the homology at C-terminus, suggests that CetZ originated from the archaeal FtsZ.

The C-terminus of TubZ is substantially divergent, even between TubZ sequences. The HMM of CetZ_C did not match with that of Firmicutes TubZs, but the C-terminus HMM of TubZ matched with CetZ and archaeal FtsZ sequences (refer to external purple heatmap in **Figure 5**, e-value in the range of $1e^{-2}$), providing consistent support for the evolutionary relationship observed in the phylogenetic reconstructions (**Figure 4**). Thus, considering the potential antiquity of archaeal FtsZ compared with TubZ, it is likely that TubZ originated from CetZ or archaeal FtsZ sequences.

Unclear origins of other archaeal tubulin-related proteins

Next, we inspected the C-termini of the archaeal groups branching basally to the tubulins, one of which is formed

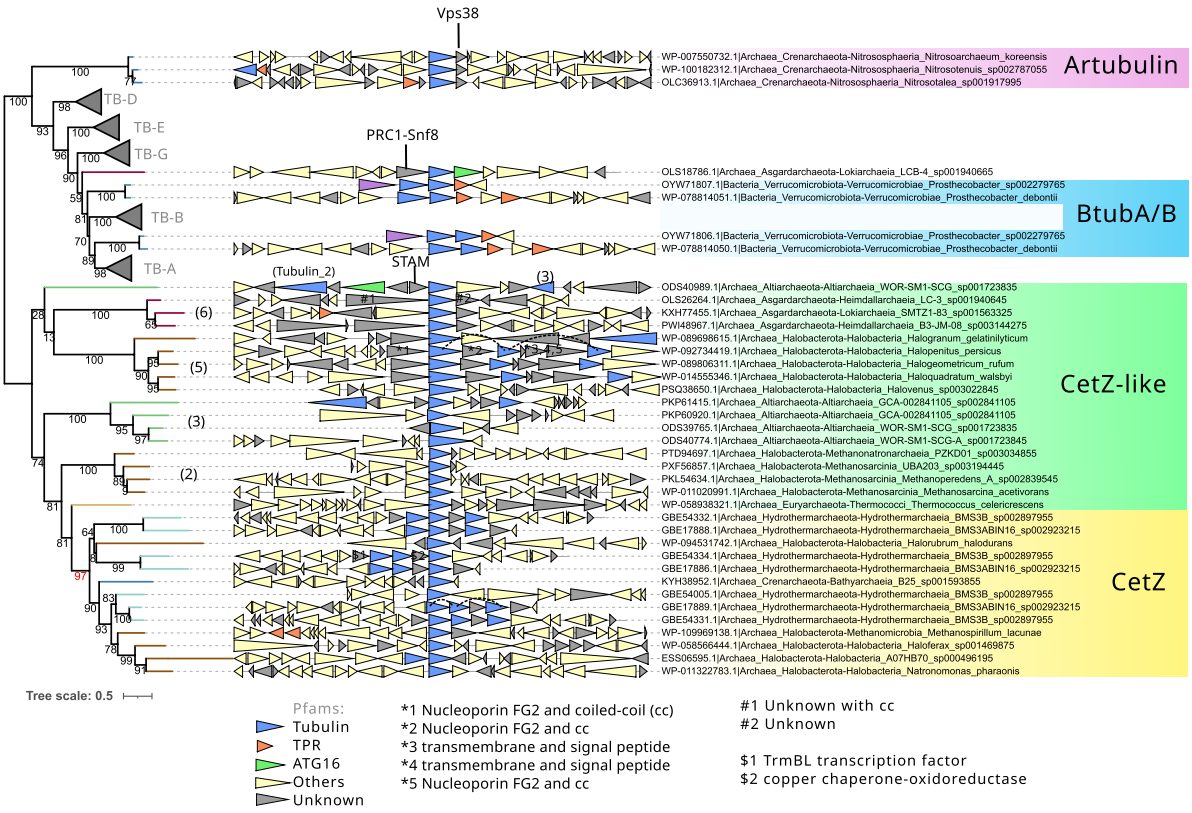


FIGURE 6
Phylogeny and genome context of the prokaryotic tubulin and CetZ-related sequences. The tree was rooted at the mid-point. The dataset for the phylogeny was selected manually, applying a gap trimming of 10%, resulting in a final MSA of 81 sequences and 485 positions. The tree was built using IQ-TREE, automatic model selection (LG + R5), and ultrafast bootstrap (UFBoot2). The number between parenthesis indicates the phylogenetic groups in **Figure 4**. Red bootstrap denotes the supported node defining the canonical CetZ proteins. Genes are colored according to the presence of Pfam domains, and genes without recognizable domains are colored in gray. Some of these latter were annotated with HHPred (indicated with #, *, $, and arrows). Discontinuous lines in the genome context panel show copies of the same CetZ(−like) that were omitted in this reconstruction.

by Asgardarchaeota (Asg.Tub), another one by Halobacterota (Halo.Tub), and the third one by artubulins (**Figure 4** and **Supplementary Figure 4**). The intermediary positions of these groups between CetZ and eukaryotic tubulins were stable when different phylogenetic methods were applied (**Supplementary Figure 5**). The monophyletic branching of Heimdallarchaeia and Lokiarchaeia is remarkable since it could illustrate the deep evolution of these paralogs in Asgardarchaeota. Here, it is important to mention that the scarcity of these groups translates into strict models that do not contain divergent sequences (note that the sequence conservation at the C-terminus is usually weak). The searches of their respective C-termini HMMs did not match with any other tubulin/FtsZ protein (refer to external red, black, and blue heatmaps in **Figure 5** and **Supplementary Data**), even with each other. However, by inspecting the MSA of the selected sequences, we observed that their C-termini show limited similarity to CetZ sequences (**Supplementary Figure 6**). Conversely, we found that artubulins are more closely related to eukaryotic tubulins than to Halo.Tub and Asg.Tub since these contain important indels that characterize archaeal CetZ (and CetZ-like) sequences from eukaryotic tubulins (black highlight in **Supplementary Figure 6**), which is also supported by the mid-point rooting in the respective phylogeny (**Figure 6**). However, it is important to note that some of these CetZ-like sequences presented few conserved blocks at the N-terminus that resemble eukaryotic sequences (particularly in Altiarchaeota and different Halobacterota classes such as Methanosarcina, and Halobacteria, Halo.Tub, 5). Therefore, although our analyses suggest that Halo.Tub and Asg.Tub seems to be phylogenetically associated with the tubulins (**Figure 4**); inspections of the MSA and mid-point rooting of the selected set of sequences (**Figure 6**) argue against an apparent tubulin-like affiliation.

The genome context of these archaeal "tubulins like" is characterized by unknown genes that we manually annotated with HHPRED (Söding et al., 2005). We observed that most of these contain coiled-coil regions with different functions (refer to gray triangles in **Figure 6**). In cases such as Halo.Tub,

their genome context contains several copies in a quasi-tandem disposition. In contrast, artubulin from *Nitrososphaera korensis* is associated with a potential Snf7 protein upstream, as previously observed (Yutin and Koonin, 2012). This abundance of unknown and unrelated genes could illustrate the diversity of molecular systems and functions in which these archaeal CetZ/Tubulin homologs are involved.

Therefore, the divergence of their C-termini, together with the sparse distribution of each phylogenetic group, shows the phylogenetic positions of these Halo.Tub and Asg.Tub in relation to each other and with eukaryotic tubulins must be interpreted cautiously. These proteins appear to be more closely related to CetZ sequences than eukaryotic tubulins. The genomic context between the CetZs and these CetZ-like sequences is not related (**Figure 6**), hindering inferences regarding the origin of these proteins. Indeed, the idea of LGTs between archaea, followed by neofunctionalization, should not be discarded for these specific proteins. Nevertheless, our results suggest that artubulins are the only ones of these proteins that could be related to ancestral eukaryotic tubulins, as previously pointed out (Yutin and Koonin, 2012).

Prokayotic tubulins are closely related to α/β-tubulins

The monophyletic branching of eukaryotic tubulins (**Figure 6**) confirms that eukaryotic tubulins had a single origin and suggests that its paralogs (α, β, γ, δ, ε, and ζ) were already established in the LECA as previously proposed (Findeisen et al., 2014). Other divergent tubulin-related proteins such as Misato, were not included in this analysis because of their strong sequence divergence. Tubulins have a very restricted distribution in prokaryotes and are only found in one bacterial species and one archaeal metagenome. OdinTubulin is only found in Odinarchaeota from the Asgardarchaeota phylum and a pair of tubulins, BtubA and BtubB, is present in various *Prostecobacter* species from the Verrucomicrobiota phylum. These prokaryotic sequences branch basally and paraphyletically to the α- and β-tubulins, although BtubA and BtubB can also branch basally to α- and β-tubulins, respectively (**Figures 4**, **6**). These phylogenetic instabilities of BtubA/B have been observed previously (Findeisen et al., 2014). In our reconstructions, their monophyletic branching was exclusively observed in the inclusive trimming, removing only gap positions (**Supplementary Figure 5** right panel). In contrast, using a stricter trimming with BMGE, BtubA and BtubB were paraphyletic (**Supplementary Figure 5** right panel; note that the topology of eukaryotic tubulins also differed and branched monophyletically to the β/γ group in BMGE trimming). In contrast, using an inclusive trimming (trimAl -gt 0.10) and considering only tubulin and CetZ(–like) sequences, the paraphyly of BtubA/B was again recovered (**Figure 6**). Therefore, the mono/paraphyly of BtubA/B seems to be sensitive to the trimming methods and the diversity of sequences included in the MSA. Similarly, the locations of the OdinTubulin in our reconstructions (**Figures 4**, **6**) differed from those obtained in previous studies, in which the Odinarchaeota sequence branched basally to the whole eukaryotic tubulin family (Zaremba-Niedzwiedzka et al., 2017), instead of basally to α- and β-tubulins. These differences could be due to different parameters of the reconstructions (taxonomic sampling, the use of different outgroups, or different methods for the MSA trimming), showing the phylogenetic instability of these prokaryotic tubulins and, therefore, the difficulty in inferring their evolutionary placement. Nevertheless, from our reconstructions, it is clear that OdinTubulins and BtubA/B are closer to α- and β-tubulins, in agreement with recent analyses (Akıl et al., 2022).

In eukaryotes, the α- and β-tubulins generally form 13 subunits filaments as imposed by the nucleation of the γ-tubulin ring composed of 13 γ-tubulins (Chaaban and Brouhard, 2017; Cavalier-Smith and Chao, 2020). The *Prosthecobacter*'s BtubA/B microtubules are formed by four or five protofilaments, presenting heterodimer polarity which can polymerize without the need of the microtubule processing centers, tubulin γ, never found in prokaryotes. In addition, *btubA/B* genes form an operon together with the *btubC* gene, a "bacterial kinesin light chain" that has been shown to stabilize the filament (Deng et al., 2017). Therefore, the functional similarities between BtubA/B and α- and β-tubulin supports their close relationships.

In contrast to the dual BtubA/B, there is a unique OdinTubulin in the Odinarchaeota metagenome sequenced from hydrothermal environments (Zaremba-Niedzwiedzka et al., 2017). Despite the possible incompleteness of the Asgardarchaeota metagenomes, the sequencing of several Asgardarchaeota clades (Liu et al., 2021) confirmed that this OdinTubulin is still uniquely found in this archaeon. *In vitro* and at high temperatures, this OdinTubulin forms tubules with short curved protofilaments coiling around the tubule circumference, more similar to FtsZ, rather than running parallel to its length as in eukaryotic microtubules (Akıl et al., 2022). Nevertheless, the genome context of this OdinTubulin displays interesting features related to eukaryotes. A hypothetical protein showing similarity to Prc1 at the N-terminus and Snf8 at the C-terminus (containing coiled-coil segments) is located upstream of the OdinTubulin gene. Prc1 is a key regulator of cytokinesis that cross-links antiparallel microtubules, while Snf8 is a member of the ESCRT-II system. Downstream of the OdinTubulin, we found a gene containing coiled-coil segments weakly related to the ATG16 Pfam domain, which belongs to the eukaryotic autophagy system. However, according to the NCBI annotation, this gene is annotated as chromosome partitioning protein Smc. The two genes surrounding the OdinTubulin are specific to this genome as protein searches do not retrieve any homolog sequences in the three domains. Together, these observations

suggest that this OdinTubulin is specific to species related to Odinarchaeota and that it could be functionally related to a eukaryotic-like system involved in endomembrane system and cell division (exemplified by the hypothetical prc1-snf8 gene), in agreement with its phylogenetic position.

Therefore, even though BtubA/B and OdinTubulin are closely related at the sequence level, as they are associated with α- and β-tubulins, they seem to be involved in different molecular systems or complexes, where the BtubA/B pair functions as eukaryotic-like microtubule filaments and OdinTubulin could function as monomeric filaments in combination with the Pcr1-Snf8 protein.

Discussion

In this study, we investigated the early origin and diversification of the main members of the FtsZ/tubulin protein family, which was already present in the LUCA. We inferred that FtsZ was present in the LBCA and LACA, although, in the latter, FtsZ was duplicated before the diversification of the Archaea domain. The differences in the gene organization around the *ftsZ* gene in archaeal and bacterial genomes reveal different evolutionary histories. Bacterial *ftsZ* genes are associated with cell division and peptidoglycan synthesis genes, while one of the archaeal *ftsZ*s is mainly associated with genes involved in protein biosynthesis (t-RNA, ribosomal proteins, etc.). The latter suggests that the expression of the archaea FtsZ gene (mainly FtsZ1) is concurrent with an increase in protein biosynthesis, perhaps necessary for preparing the cell division. In contrast, peptidoglycan biosynthesis genes are nearly absent in archaea, except for a few Euryarchaeota, such as Methylobacteriales and Methylopyrales, that are able to produce peptidoglycan-like cell walls made up of pseudomurein, and whose biosynthesis has a common origin with the bacterial murein, but not necessarily reflecting an LGT from bacteria (Subedi et al., 2021). Nevertheless, the gene cluster involved in the biosynthesis of pseudomurein does not contain *ftsZ* (manually checked), suggesting that the duplication of FtsZ and acquisition of pseudomurein in these Euryarchaeaota orders were independent. In fact, although intradomain LGT of FtsZ can not be discarded, it is important to note that we did not detect any case of FtsZ LGT between bacteria and archaea. Therefore, this scenario of different genome landscapes around *ftsZ* genes is possibly a reflection of the evolutionary pressures that happened during the divergence between Bacteria and Archaea, such as the loss or acquisition of peptidoglycan biosynthesis.

Our analyses also demonstrate that the C-termini of these proteins contain enough evolutionary signals to solve their relationships. In particular, we show that the archaeal CetZ (mainly found in the Euryarchaeota supergroup) and the plasmidic TubZ (found in Firmicutes), most likely originated from archaeal FtsZ. This last observation regarding TubZ could clarify previous claims regarding the affiliation of TubZ to prokaryotic FtsZ (Findeisen et al., 2014). In contrast, the exclusiveness of the different C-termini domains in CetZ-like proteins, together with their unrelated genome context, suggests a complex evolutionary history of these proteins. In the MSAs, we observed consistent evidence to at least differentiate Halo.Tubs and Asg.Tubs from artubulins, and to define artubulins as members of the eukaryotic tubulin protein family—arguing thus against previous assumptions stating a close relationship between TubZ and artubulins (Findeisen et al., 2014). Given that the evolution of these sequences is irregular in archaea (and in bacteria), LGT events between them could be the most plausible explanation for such a distribution. However, given the close proximity of archaeal paralogs to eukaryotic tubulins, which probably originated from archaeal FtsZ, our data still suggest an archaeal origin of tubulins, as previously suggested (Yutin and Koonin, 2012).

Prokaryotic tubulins, such as BtubA/B and OdinTubulin, seem to be closely related to α- and β-tubulins. This could be an indication of the origin and diversification of eukaryotic tubulins from α- and β-tubulin paralogs, and consequently, the basal position of the rest of the tubulins could be the result of strong divergence due to subfunctionalization. In fact, the monophyletic branching of BtubA/B in some of our reconstructions could reflect a reminiscence of the early diversification of tubulins by gene duplications, although further analyses are required to test this assumption. Nevertheless, while prokaryotic tubulins are probably involved in different cellular processes, there is no doubt that they are similar to the ancestral eukaryotic molecular systems or complexes. Altogether, our analyses revisit the evolution of the main members of the FtsZ/tubulin protein family, providing evidence of its early origin and diversification across the three domains of life.

Materials and methods

Collection of FtsZ/tubulin homologs

To collect homologous sequences from the tubulin/FtsZ protein family, we performed HMMSEARCH (HMMER 3.1b2) (e-value threshold $1e^{-3}$; Prakash et al., 2017) against a local database comprising all the NCBI genomes taxonomically annotated with GTDB (release bac120 and arc122; Parks et al., 2018), plus 36 diverse eukaryotes (24,664 taxa in total). As a query, we used the Tubulin Pfam model (PF00091) spanning the N-terminus as it is the commonly shared domain of this family, for both FtsZ, tubulins, and derivatives. This search also detected FtsZ-like sequences (Makarova and Koonin, 2010), containing the Tubulin_2 Pfam model (PF13809) instead. The scattered and sparse taxonomic distribution of this FtsZ-like (Tubulin_2), together with their irregular domain architecture, that is, they

do not have the FtsZ_C nor Tubulin_C Pfam domains but have important N/C-terminus extensions (in Planctomycetes a protein kinase domain is found at the N-terminus). Their sequence divergence (to tubulin/FtsZ family members but also between them) suggested that these proteins had a different evolutionary history to the canonical tubulin/FtsZ protein family (Makarova and Koonin, 2010). Thus, to obtain more accurate multiple sequence alignments (MSAs) and phylogenies, we exclude these sequences from our main reconstructions.

Dataset construction, MSA, and phylogenies

To define the main tubulin/FtsZ subfamilies, we aligned all the sequences (45,929 sequences) with MAFFT v7.310 (Katoh and Standley, 2013), applied gap trimming using trimAl 1.4.22 (–gt 0.7; Capella-Gutiérrez et al., 2009), removed sequences with lower alignment coverage than 60%, built a Fasttree 2.1.10 (Price et al., 2010), and selected the branches of interest to build different datasets (FtsZ, tubulin-like proteins or to exclude FtsZ-like sequences). To avoid the over-representation of some taxa and build taxonomically balanced datasets, we removed redundant sequences (using CD-HIT; Fu et al., 2012) by taxonomic classes (following the GTDB taxonomy) and applied different cut-offs depending on the number of sequences (from 20 sequences, 95% identity, up to >250, 55%).

From these reduced datasets, we built different datasets for specific purposes: that is, focused on prokaryotic FtsZ (**Figure 2**), all tubulin/FtsZ and homologous proteins with a special focus on archaea (**Figures 4**, **5**), and tubulin/CetZ-like proteins (**Figure 6**). All trees were visualized and annotated using iTOL (Letunic and Bork, 2019). Raw data for the MSAs and respective phylogenetic trees are provided in **Supplementary Data**.

For the phylogenies of bacterial FtsZs (**Figure 2**), we randomly selected one sequence from each taxonomic order from the FtsZ prokaryotic dataset (note that some taxa such as Verrucomicrobiota, CPR were manually increased by selecting representative sequences). Then, we removed spurious sequences based on the visualization of the MSA (trimAl -gt 20) and the respective phylogenetic tree (Fasttree). The final set of sequences was aligned with MAFFT-linsi, and three filtering methods were applied: gap-trimmed alignment (–gt 0.2), BMGE-1.12 (–h 0.55 -m BLOSUM30), using trimAl (–gt 0.7). Due to some miss-aligned regions at the C-terminus in FtsZ from Verrucomicrobia, the manually trimmed alignment was re-aligned as follows: We ran MAFFT-Linsi excluding CPR-Verrucomicrobia divergent sequences, and these later ones were aligned independently; then, both MSAs were treated with trimAl -gt 0.2 and merged with MAFFT (–merge option). Each trimming method had two sub-datasets: one including bacterial and archaeal FtsZs (BA), and the other one containing exclusively bacterial FtsZs (B). From the manually trimmed alignments, B MSA was obtained by removing archaeal sequences from BA MSA, while in BMGE- and trimAl-gt70-trimming MSAs, B, and BA MSAs were aligned independently through MAFFT-Linsi.

For the phylogeny including FtsZ, tubulins, and others (**Figures 4**, **5**), the dataset was constructed by a selection of representative bacterial FtsZs and eukaryotic tubulins, plus an archaeal selection based on redundancy filtering and recovering the discarded paralogs from the remaining taxa. Then, this set of the sequence was aligned with MAFFT-Linsi and trimmed with trimAl (–gt 0.2) and BMGE (the same options as mentioned earlier).

For the phylogeny of tubulins and CetZ(–like) sequences (**Figure 6**), we manually selected a set of representative sequences branching basally to eukaryotic tubulins extracted from the tree of **Figure 3**.

Phylogenies were inferred using maximum-likelihood methods with IQ-TREE (Nguyen et al., 2015). We obtained branch support with ultrafast bootstrap (UFBoot2, option -B 1000; Hoang et al., 2018), and the evolutionary models of each set of the sequences were automatically selected using ModelFinder (Kalyaanamoorthy et al., 2017), following the BIC criterion (–m MFP). In addition, we also run UFBoot2 + NNI bootstraps (-B 1000 -bnni options), testing complex evolutionary models such as LG4X (–m LG4X) or C10-30 (–mset LG -madd LG + C10, LG + C20, LG + C30, LG + C10 + R + F, LG + C20 + R + F, LG + C30 + R + F –score-diff all).

Genomic context

To analyze the genome contexts of the tubulin/FtsZ protein family, the DNA sequence corresponding to 10 Kb surrounding (5 kb up/downstream) each of the respective coding genes was extracted. In that fragment, genes were predicted using Prodigal (Hyatt et al., 2010), with default parameters, and the predicted proteins annotated with HMMSCAN (Potter et al., 2018) on the Pfam database (Finn et al., 2011). The abundance of Pfam domains in these proteins was then quantified, counting only the presence and not the repetitions inside the same protein or genome context, that is, multiple copies of the same domain in one gene or multiple gene copies in one genome context was only counted once.

Definition of C-termini protein models and searches

To build the protein models (hidden Markov model, HMM) of the C-termini of the different proteins of interest (CetZ, TubZ, artubulins, Halo.Tub, and Asg.Tub), the sequences were aligned individually, and soft gap trimming was applied with

trimAl (–gt 0.1). Then, the C-termini fragments were extracted starting from the end of the Tubulin Pfam domain (N-terminus domain). Protein models were made using HMMBUILD (Potter et al., 2018) and were then used to perform hmmsearches against our dataset of the FtsZ/tubulin protein family without an e-value threshold. Raw data for the construction of protein models and respective searches are provided in **Supplementary Data**.

Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding authors.

Author contributions

DD: funding acquisition and supervision. CS-M and DS-N: investigation and formal analysis. All authors contributed to the article and approved the submitted version.

Funding

This publication was funded in part by the Gordon and Betty Moore Foundation through Grant GBMF9733 and by the Spanish Ministry of Economy and Competitiveness (Grant No. BFU2016-78326-P) to DD.

Acknowledgments

The authors thank Anja Spang and Berend Snel for the financial support of CS-M through a grant funded through Utrecht University and financed by the Netherlands Organisation for Scientific Research (NWO).

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.

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.2022.1100249/full#supplementary-material

References

Acebo, P., García de Lacoba, M., Rivas, G., Andreu, J. M., Espinosa, M., and del Solar, G. (1998). Structural features of the plasmid pMV158-encoded transcriptional repressor CopG, a protein sharing similarities with both helix-turn-helix and beta-sheet DNA binding proteins. *Proteins* 32, 248–261. doi: 10.1002/(sici)1097-0134(19980801)32:2<248::aid-prot11>3.0.co;2-d

Akıl, C., Ali, S., Tran, L. T., Gaillard, J., Li, W., Hayashida, K., et al. (2022). Structure and dynamics of Odinarchaeota tubulin and the implications for eukaryotic microtubule evolution. *Sci. Adv.* 8:eabm2225.

Ayala, J. A., Garrido, T., De Pedro, M. A., and Vicente, M. (1994). "Chapter 5 Molecular biology of bacterial septation," in *New comprehensive biochemistry Bacterial Cell Wall*, eds J.-M. Ghuysen and R. Hakenbeck (Amsterdam: Elsevier), 73–101. doi: 10.1016/S0167-7306(08)60408-1

Bi, E., and Lutkenhaus, J. (1991). FtsZ ring structure associated with division in *Escherichia coli*. *Nature* 354, 161–164. doi: 10.1038/354161a0

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Caspi, Y., and Dekker, C. (2018). Dividing the Archaeal Way: The ancient Cdv cell-division machinery. *Front. Microbiol.* 9:174. doi: 10.3389/fmicb.2018.00174

Cavalier-Smith, T., and Chao, E. E.-Y. (2020). Multidomain ribosomal protein trees and the planctobacterial origin of neomura (eukaryotes, archaebacteria). *Protoplasma* 257, 621–753. doi: 10.1007/s00709-019-01442-7

Chaaban, S., and Brouhard, G. J. (2017). A microtubule bestiary: Structural diversity in tubulin polymers. *Mol. Biol. Cell* 28, 2924–2931. doi: 10.1091/mbc.E16-05-0271

Coleman, G. A., Davín, A. A., Mahendrarajah, T. A., Szánthó, L. L., Spang, A., Hugenholtz, P., et al. (2021). A rooted phylogeny resolves early bacterial evolution. *Science* 372:eabe0511. doi: 10.1126/science.abe0511

Davis, B. K. (2002). Molecular evolution before the origin of species. *Prog. Biophys. Mol. Biol.* 79, 77–133. doi: 10.1016/S0079-6107(02)00012-3

Deng, X., Fink, G., Bharat, T. A. M., He, S., Kureisaite-Ciziene, D., and Löwe, J. (2017). Four-stranded mini microtubules formed by Prosthecobacter BtubAB show dynamic instability. *Proc. Natl. Acad. Sci. U.S.A.* 114, E5950–E5958. doi: 10.1073/pnas.1705062114

Duggin, I. G., Aylett, C. H. S., Walsh, J. C., Michie, K. A., Wang, Q., Turnbull, L., et al. (2015). CetZ tubulin-like proteins control archaeal cell shape. *Nature* 519, 362–365. doi: 10.1038/nature13983

Findeisen, P., Mühlhausen, S., Dempewolf, S., Hertzog, J., Zietlow, A., Carlomagno, T., et al. (2014). Six subgroups and extensive recent duplications characterize the evolution of the eukaryotic tubulin protein family. *Genome Biol. Evol.* 6, 2274–2288. doi: 10.1093/gbe/evu187

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281

Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Ithurbide, S., Gribaldo, S., Albers, S.-V., and Pende, N. (2022). Spotlight on FtsZ-based cell division in Archaea. *Trends Microbiol.* 30, 665–678. doi: 10.1016/j.tim.2022.01.005

Janke, C., and Magiera, M. M. (2020). The tubulin code and its role in controlling microtubule properties and functions. *Nat. Rev. Mol. Cell Biol.* 21, 307–326. doi: 10.1038/s41580-020-0214-3

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Larsen, R. A., Cusumano, C., Fujioka, A., Lim-Fong, G., Patterson, P., and Pogliano, J. (2007). Treadmilling of a prokaryotic tubulin-like protein, TubZ, required for plasmid stability in Bacillus thuringiensis. *Genes Dev.* 21, 1340–1352. doi: 10.1101/gad.1546107

Leger, M. M., Petrů, M., Žárský, V., Eme, L., Vlček, Č., Harding, T., et al. (2015). An ancestral bacterial division system is widespread in eukaryotic mitochondria. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10239–10246. doi: 10.1073/pnas.1421392112

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Liao, Y., Ithurbide, S., Evenhuis, C., Löwe, J., and Duggin, I. G. (2021). Cell division in the archaeon Haloferax volcanii relies on two FtsZ proteins with distinct functions in division ring assembly and constriction. *Nat. Microbiol.* 6, 594–605. doi: 10.1038/s41564-021-00894-z

Liu, Y., Makarova, K. S., Huang, W.-C., Wolf, Y. I., Nikolskaya, A. N., Zhang, X., et al. (2021). Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature* 593, 553–557. doi: 10.1038/s41586-021-03494-3

Lluch-Senar, M., Querol, E., and Piñol, J. (2010). Cell division in a minimal bacterium in the absence of ftsZ. *Mol. Microbiol.* 78, 278–289. doi: 10.1111/j.1365-2958.2010.07306.x

Makarova, K. S., and Koonin, E. V. (2010). Two new families of the FtsZ-tubulin protein superfamily implicated in membrane remodeling in diverse bacteria and archaea. *Biol. Dir.* 5:33. doi: 10.1186/1745-6150-5-33

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic Era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015

Moody, E. R., Mahendrarajah, T. A., Dombrowski, N., Clark, J. W., Petitjean, C., Offre, P., et al. (2022). An estimate of the deepest branches of the tree of life from ancient vertically evolving genes. *eLife* 11:e66695. doi: 10.7554/eLife.66695

Mottier-Pavie, V., Cenci, G., Vernì, F., Gatti, M., and Bonaccorsi, S. (2011). Phenotypic analysis of misato function reveals roles of noncentrosomal microtubules in *Drosophila* spindle formation. *J. Cell Sci.* 124, 706–717. doi: 10.1242/jcs.072348

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Palumbo, V., Pellacani, C., Heesom, K. J., Rogala, K. B., Deane, C. M., Mottier-Pavie, V., et al. (2015). Misato controls mitotic microtubule generation by stabilizing the TCP-1 tubulin chaperone complex. *Curr. Biol.* 25, 1777–1783. doi: 10.1016/j.cub.2015.05.033

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996–1004. doi: 10.1038/nbt.4229

Pende, N., Sogues, A., Megrian, D., Sartori-Rupp, A., England, P., Palabikyan, H., et al. (2021). SepF is the FtsZ anchor in archaea, with features of an ancestral cell division system. *Nat. Commun.* 12:3214. doi: 10.1038/s41467-021-23099-8

Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., and Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Res.* 46, W200–W204. doi: 10.1093/nar/gky448

Prakash, A., Jeffryes, M., Bateman, A., and Finn, R. D. (2017). The HMMER web server for protein sequence similarity search. *Curr. Protoc. Bioinformatics* 60, 3.15.1–3.15.23. doi: 10.1002/cpbi.40

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490

Pulschen, A. A., Mutavchiev, D. R., Culley, S., Sebastian, K. N., Roubinet, J., Roubinet, M., et al. (2020). Live Imaging of a Hyperthermophilic Archaeon reveals distinct roles for Two ESCRT-III homologs in ensuring a robust and symmetric division. *Curr. Biol.* 30, 2852–2859.e4. doi: 10.1016/j.cub.2020.05.021

Rivas-Marín, E., Canosa, I., and Devos, D. P. (2016). Evolutionary cell biology of division mode in the bacterial Planctomycetes-Verrucomicrobia- Chlamydiae Superphylum. *Front. Microbiol.* 7:1964. doi: 10.3389/fmicb.2016.01964

Schlieper, D., Oliva, M. A., Andreu, J. M., and Löwe, J. (2005). Structure of bacterial tubulin BtubA/B: Evidence for horizontal gene transfer. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9170–9175. doi: 10.1073/pnas.0502859102

Söding, J., Biegert, A., and Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 33, W244–W248. doi: 10.1093/nar/gki408

Subedi, B. P., Martin, W. F., Carbone, V., Duin, E. C., Cronin, B., Sauter, J., et al. (2021). Archaeal pseudomurein and bacterial murein cell wall biosynthesis share a common evolutionary ancestry. *FEMS Microbes* 2:xtab012. doi: 10.1093/femsmc/xtab012

TerBush, A. D., Yoshida, Y., and Osteryoung, K. W. (2013). FtsZ in chloroplast division: Structure, function and evolution. *Curr. Opin. Cell Biol.* 25, 461–470. doi: 10.1016/j.ceb.2013.04.006

van Wolferen, M., Pulschen, A. A., Baum, B., Gribaldo, S., and Albers, S.-V. (2022). The cell biology of archaea. *Nat. Microbiol.* 7, 1744–1755. doi: 10.1038/s41564-022-01215-8

Wagstaff, J., and Löwe, J. (2018). Prokaryotic cytoskeletons: Protein filaments organizing small cells. *Nat. Rev. Microbiol.* 16, 187–201. doi: 10.1038/nrmicro.2017.153

Yutin, N., and Koonin, E. V. (2012). Archaeal origin of tubulin. *Biol. Dir.* 7:10. doi: 10.1186/1745-6150-7-10

Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Bäckström, D., Juzokaite, L., Vancaester, E., et al. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541, 353–358. doi: 10.1038/nature21031

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 22 March 2023
DOI 10.3389/fmicb.2023.1145673

OPEN ACCESS

EDITED BY
Feng Gao,
Tianjin University,
China

REVIEWED BY
Claude Eliane Saint-Ruf,
Université Paris Cité,
France
Rohan Balakrishnan,
University of California,
San Diego,
United States

*CORRESPONDENCE
Bei-Wen Ying
✉ ying.beiwen.gf@u.tsukuba.ac.jp

SPECIALTY SECTION
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 16 January 2023
ACCEPTED 02 March 2023
PUBLISHED 22 March 2023

CITATION
Matsui Y, Nagai M and Ying B-W (2023) Growth rate-associated transcriptome reorganization in response to genomic, environmental, and evolutionary interruptions.
Front. Microbiol. 14:1145673.
doi: 10.3389/fmicb.2023.1145673

Growth rate-associated transcriptome reorganization in response to genomic, environmental, and evolutionary interruptions

Yuichiro Matsui, Motoki Nagai and Bei-Wen Ying*

School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan

The genomic, environmental, and evolutionary interruptions caused the changes in bacterial growth, which were stringently associated with changes in gene expression. The growth and gene expression changes remained unclear in response to these interruptions that occurred combinative. As a pilot study, whether and how bacterial growth was affected by the individual and dual interruptions of genome reduction, environmental stress, and adaptive evolution were investigated. Growth assay showed that the presence of the environmental stressors, i.e., threonine and chloramphenicol, significantly decreased the growth rate of the wild-type *Escherichia coli*, whereas not that of the reduced genome. It indicated a canceling effect in bacterial growth due to the dual interruption of the genomic and environmental changes. Experimental evolution of the reduced genome released the canceling effect by improving growth fitness. Intriguingly, the transcriptome architecture maintained a homeostatic chromosomal periodicity regardless of the genomic, environmental, and evolutionary interruptions. Negative epistasis in transcriptome reorganization was commonly observed in response to the dual interruptions, which might contribute to the canceling effect. It was supported by the changes in the numbers of differentially expressed genes (DEGs) and the enriched regulons and functions. Gene network analysis newly constructed 11 gene modules, one out of which was correlated to the growth rate. Enrichment of DEGs in these modules successfully categorized them into three types, i.e., conserved, responsive, and epistatic. Taken together, homeostasis in transcriptome architecture was essential to being alive, and it might be attributed to the negative epistasis in transcriptome reorganization and the functional differentiation in gene modules. The present study directly connected bacterial growth fitness with transcriptome reorganization and provided a global view of how microorganisms responded to genomic, environmental, and evolutionary interruptions for survival from wild nature.

KEYWORDS

negative epistasis, transcriptome reorganization, gene network, chromosomal periodicity, genome reduction, environmental stress, experimental evolution

Introduction

Bacterial growth was primarily disturbed by genomic, environmental, and evolutionary interruptions. How the individual interruption contributed to growth fitness was intensively studied in laboratories (Baba et al., 2006; Basan et al., 2020; Kinsler et al., 2020), although these interruptions commonly happened combinative in wild nature. As a representative case of genomic interruption, the genome reduction could cause habitat specialization (Nicks and Rahn-Lee, 2017; Simonsen, 2022), such as increased temperature for thermophiles (Sabath et al., 2013) and decreased pH for acidophiles (Cortez et al., 2022), which was supposed to be the consequence of adaptative evolution in nature (Martinez-Cano et al., 2014; Chu et al., 2021). How the growth fitness of these microbes was influenced by genome reduction and the environmental or evolutionary changes was challenging to analyze in wild nature (Lewis et al., 2021). Instead, genome reduction was conducted in the laboratory by deleting nonessential genomic sequences for bacterial growth (Feher et al., 2007; Kato and Hashimoto, 2007). High-throughput growth assay demonstrated that the genome reduction commonly led to the fitness decrease (Karcagi et al., 2016; Kurokawa et al., 2016), although it might benefit metabolic engineering (Mizoguchi et al., 2007; Morimoto et al., 2008; Vernyik et al., 2020). The decreased fitness could be rapidly improved by experimental evolution (Kawecki et al., 2012; Kurokawa and Ying, 2019), probably owing to the increased mutability of reduced genomes (Nishimura et al., 2017; Lao et al., 2022). The evolutionary improved growth fitness was correlated to the environmental variation (Kurokawa et al., 2022). These findings suggested that the combinations of genomic, environmental, and evolutionary interruptions would cause varied impacts on bacterial growth. The quantitative understanding of the contribution of these interruptions to growth fitness remained largely insufficient.

The gene expression pattern was commonly used to interpret bacterial growth, as it was stringently associated with the exponential growth rate (Scott et al., 2010; Klumpp and Hwa, 2014). Significant correlations of gene expression to growth fitness were reported to a large extent (Klumpp et al., 2009; Wytock and Motter, 2019; Liu et al., 2020), such as the growth rate-correlated gene clusters (Matsumoto et al., 2013) and the environmental adaptation-correlated transcriptomes (Ying et al., 2013; Murakami et al., 2015; Feugeas et al., 2016). Computational analyses observed the theoretical features of the transcriptome commonly, such as the power law (Zipf's rule) (Ying and Yama, 2018), chromosomal periodicity (Mathelier and Carbone, 2010; Nagai et al., 2020), and epistasis (Ying et al., 2013). Accordingly, there was somehow universality in transcriptome reorganization in response to the genomic, environmental, and evolutionary interruptions. Whether the dual interruption additively accumulated the changes in growth and transcriptome caused by individual interruption was an intriguing question.

The present study employed the *Escherichia coli* (*E. coli*) strains carrying the wild-type, reduced, and evolved genomes. The growth fitness and transcriptome were quantitatively evaluated under regular and environmentally stressed conditions. The growth change and the associated transcriptome reorganization in response to the genomic, environmental, and evolutionary interruptions were analyzed. We tried to illustrate an overview of the growth fitness-associated gene network and discover the working principle participating in transcriptome reorganization responsible for being alive.

Results

Canceling effect in bacterial growth

The contribution of genome reduction to growth fitness under environmental stress was examined. The wild-type W3110 (N0) and its genome-reduced (N28) strains and their experimentally evolved strains (N0e and N28e) were tested (Figure 1A). The genome size of N28 was approximately 21% smaller than that of N0, and the growth of N28 was significantly slower than that of N0 (Kurokawa et al., 2016). N0e and N28e were experimentally evolved for approximately 1,000 generations in the M63 medium, and N28e showed an improved growth rate compared to N28 (Kurokawa et al., 2022). These strains were the only collection (candidates) well-satisfied with the present study. As the environmental stressors, a number of chemical additives were tested preliminary, i.e., threonine (Thr), chloramphenicol (Cm), NaCl, ampicillin (Amp), and serine (Ser), because they were reported to inhibit the bacterial growth (Bearer and Neet, 1978; Malik et al., 2005; Gutierrez et al., 2013; Matsumoto et al., 2013; Rau et al., 2016). Firstly, the contribution of these additives to the growth of N0 was evaluated in a concentration gradient (Figure 1B; Supplementary Figure S1A). Their concentrations causing the growth rate of N0 to decrease by ~30% compared to that in the regular condition were decided. Secondly, the additives were tested for the growth of N28 at their decided concentrations (Figure 1C; Supplementary Figure S1B). As N28 failed to grow in the presence of NaCl, Amp, and Ser, Thr and Cm were used for the following analyses. Note that the biological function and molecular mechanism of the additives were nonessential for the study; whether they could cause the decreased growth rates of N0 and N28 was crucial.

Repeated growth assays showed that the significant repression in growth was caused by the stressors (Thr and Cm) for the wild-type genome (N0), regardless of the experimental evolution (N0e) (Figure 1C, left). However, the stressor-mediated decrease in growth fitness was undetected in the reduced genome (N28), indicating the reduced responsivity of the reduced genome to the two stressors. Experimental evolution improved the growth fitness of the reduced genome under the regular condition, as previously reported (Choe et al., 2019; Kurokawa et al., 2022), and in the presence of Thr or Cm (Figure 1C, right). It seemed that the evolved N28 (N28e) regained the growth fitness and the responsivity to the stressors. The changes in growth rate indicated a negative epistatic effect in the dual inhibition of growth (Figure 1D). The experimental evolution of the reduced genome released mutual inhibition of growth, i.e., either genome reduction or environmental stress inhibited bacterial growth. Transcriptome analysis was subsequently performed to interpret the canceling effect in growth.

Homeostatic chromosomal periodicity of transcriptome

Gene expression patterns were initially analyzed with hierarchical clustering (HC) and principal component analysis (PCA). HC resulted in the primary branch between N0 and N28 (Figure 2A), and PCA showed the most significant differentiation between N0 and N28 at PC1 (Figure 2B). It demonstrated the most significant differentiation in the gene expression pattern was caused by genome reduction. Additionally, the secondary branches were divergent in response to the stressors and experimental evolution in the N0 and N28 clusters,

10.3389/fmicb.2023.1145673

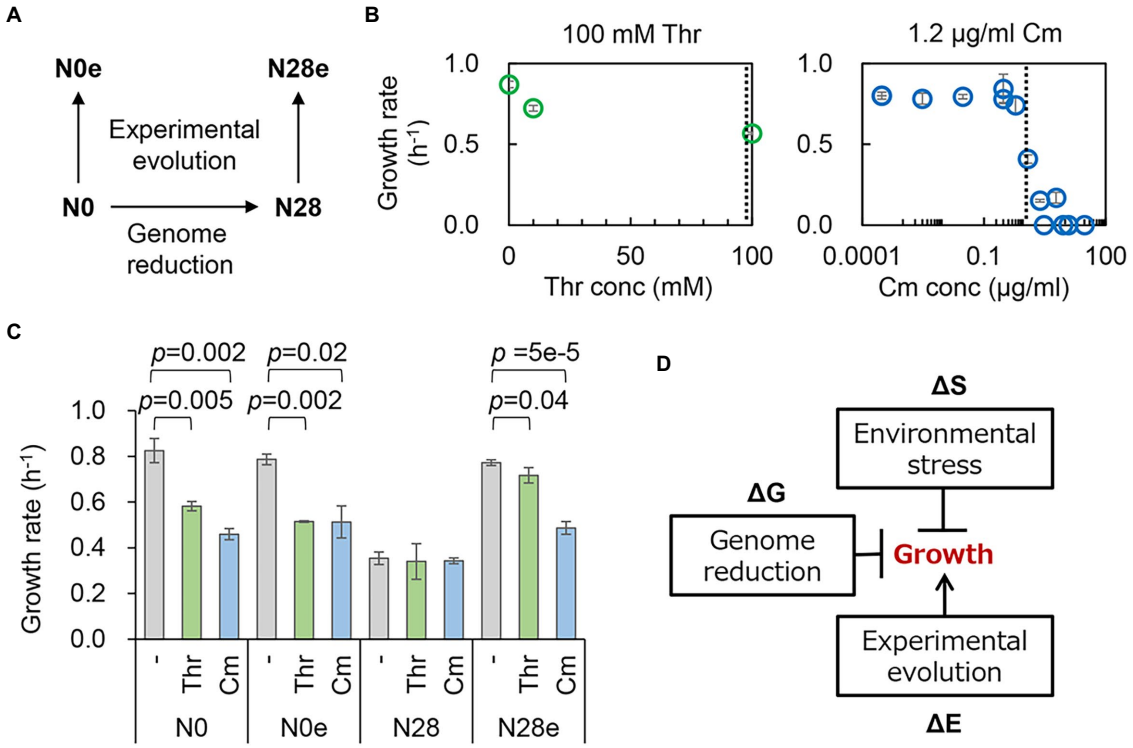


FIGURE 1
Changes in growth fitness in response to genomic, environmental, and evolutionary changes. **(A)** An overview of the *E. coli* strains. The wild-type and reduced genomes are indicated as N0 and N28, respectively. The evolved N0 and N28 are indicated as N0e and N28e, respectively. **(B)** Contribution of the concentration gradient of the additives to growth fitness. The selected concentrations of the additives are indicated by the broken lines. **(C)** *E. coli* growth rates. N0, N28, N0e, and N28e indicate the wild-type, reduced, evolved, and evolved reduced genomes, respectively. Thr and Cm stand for threonine and chloramphenicol, respectively. The statistical significance (*p* values) of Welch's *t*-tests are indicated. **(D)** Schematic drawing of the contributions of genomic, environmental, and evolutionary changes to bacterial growth. The changes mediated by genome reduction, environmental stressors, and experimental evolution are represented as ΔG, ΔS, and ΔE, respectively. The arrow and T cross from the boxes stand for the improvement and inhibition of growth, respectively.

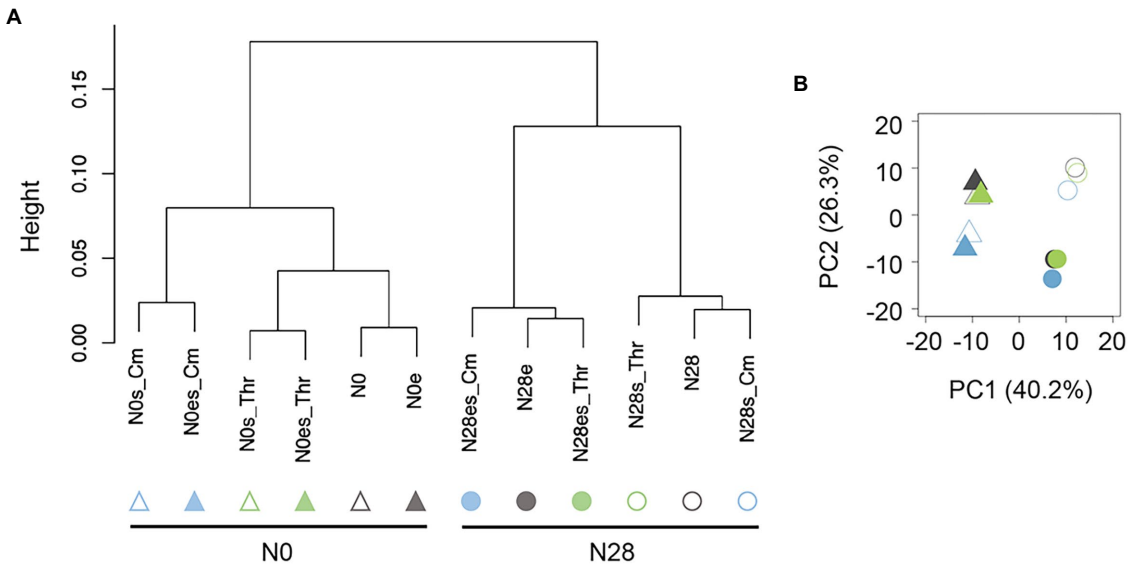


FIGURE 2
Global gene expression patterns. **(A)** Hierarchical clustering of gene expression. **(B)** Principal component analysis. Triangles and circles represent the wild-type (N0) and reduced (N28) genomes, respectively. Green and blue indicate the addition of Thr and Cm in the culture, respectively. The open and closed signals represent the ancestor and evolved genomes, respectively.

respectively (Figure 2A). It strongly suggested that the wild-type and reduced genomes were highly sensitive to environmental stress and experimental evolution, respectively. These results supported the differentiated changes in the growth fitness of N0 and N28 (Figure 1B).

On the other hand, the global transcriptome reorganization remained homeostatic, despite the differentiation in gene expression patterns due to genome reduction, environmental stress, and experimental evolution. The chromosomal periodicity of the transcriptome was analyzed as previously reported (Nagai et al., 2020). The wavelengths of the maximal power spectra acquired by the Fourier transform of the gene expression were different between N0 and N28 (Figure 3A). The transcriptomes all presented a common six-period of statistical significance (Figure 3B and Supplementary Table S1), consistent with our previous finding (Nagai et al., 2020). The finding first demonstrated that the chromosomal periodicity of the transcriptome was highly conserved, regardless of the genomic, environmental, and evolutionary changes.

Negative epistasis in transcriptome reorganization

The genetic concept of epistasis was introduced as described previously (Ying et al., 2013). The epistatic change was compared with the additive change, which was evaluated by the sum of individual changes that were different or equivalent to the simultaneous change. In the present study, the changes in transcriptome caused by genome reduction, environmental stress, and experimental evolution were designated as ΔG, ΔS, and ΔE, respectively (Figure 4A), as similar as mentioned in describing the changes in the growth (Figure 1D). The changes in the transcriptome were analyzed in varied viewpoints, i.e., changes in transcriptional levels, differentially expressed genes (DEGs), and enriched regulations and functions (Figure 4A, shadow). Firstly, the transcriptional changes caused by any single genomic (ΔG), environmental (ΔS), and evolutionary (ΔE) interruptions were calculated (Figure 4B, vertical and horizontal arrows). The simultaneous changes caused by any pair of the three interruptions (ΔGE, ΔGS, ΔES) were evaluated as the changes between the two transcriptomes with dual interruptions (Figure 4B, diagonal arrows). The additive changes caused by any pair of the three interruptions (ΔG + ΔE, ΔG + ΔS, ΔE + ΔS) were calculated as the sum of any single changes and were compared to the simultaneous changes by regression (Figure 4C, red lines). The regression slopes varied from 0.36 to 0.77, which were all smaller than 1 (Figure 4C). It demonstrated a negative epistatic effect in transcriptome reorganization triggered by dual interruptions. That is, the changes in gene expression were suppressed by each other of any pair of the three interruptions. For example, the genome reduction (ΔG) somehow canceled out the environmental

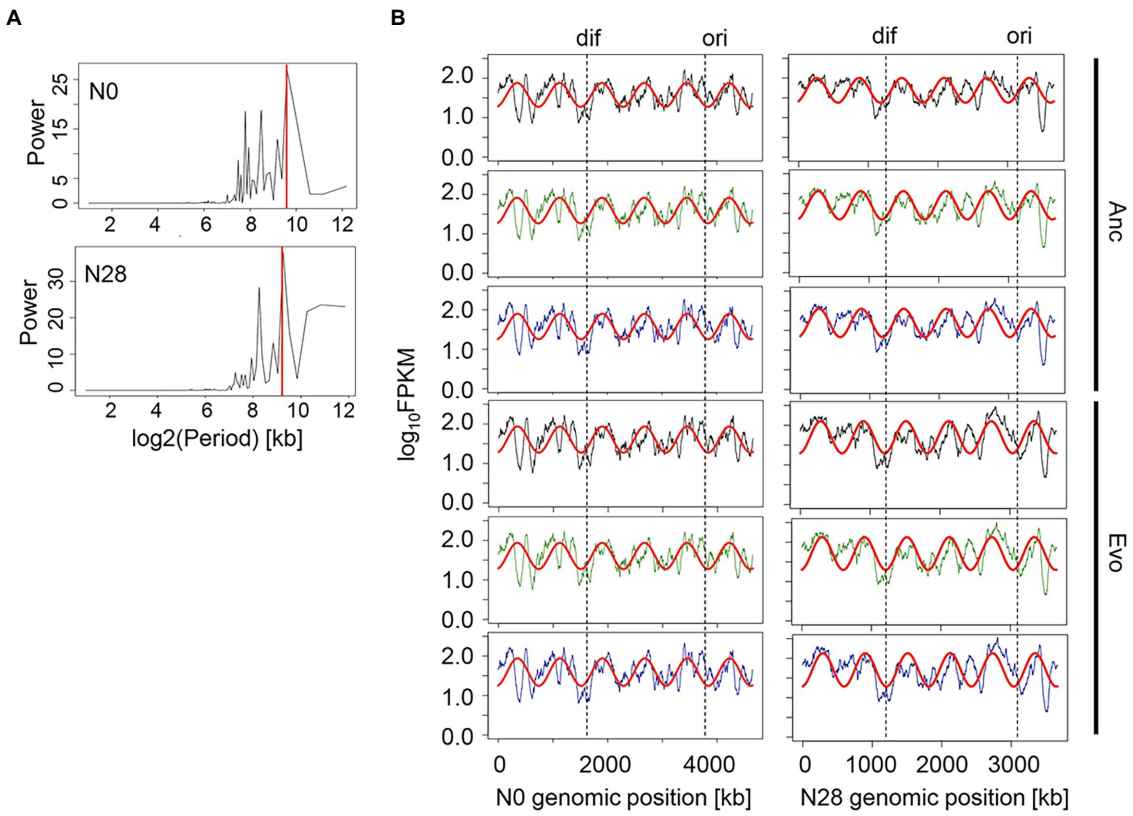


FIGURE 3
Chromosomal periodicity of transcriptomes. **(A)** Fourier transform. The examples of N0 and N28 are shown. The red lines indicate the highest power spectra (max-peak) estimated by the Fourier transform. **(B)** Chromosomal periodicity. The red curves represent the fitted periods of the transcriptomes according to the max-peaks estimated by the Fourier transform. The transcriptional levels for every 1-kb sliding window and 100-kb smoothing are shown. *Ori* and *dif* are indicated. The left and right panels show the wild-type and reduced genomes, respectively. Anc and Evo represent the ancestral and evolved populations, respectively. Green and blue indicate the addition of Thr and Cm in the culture, respectively.

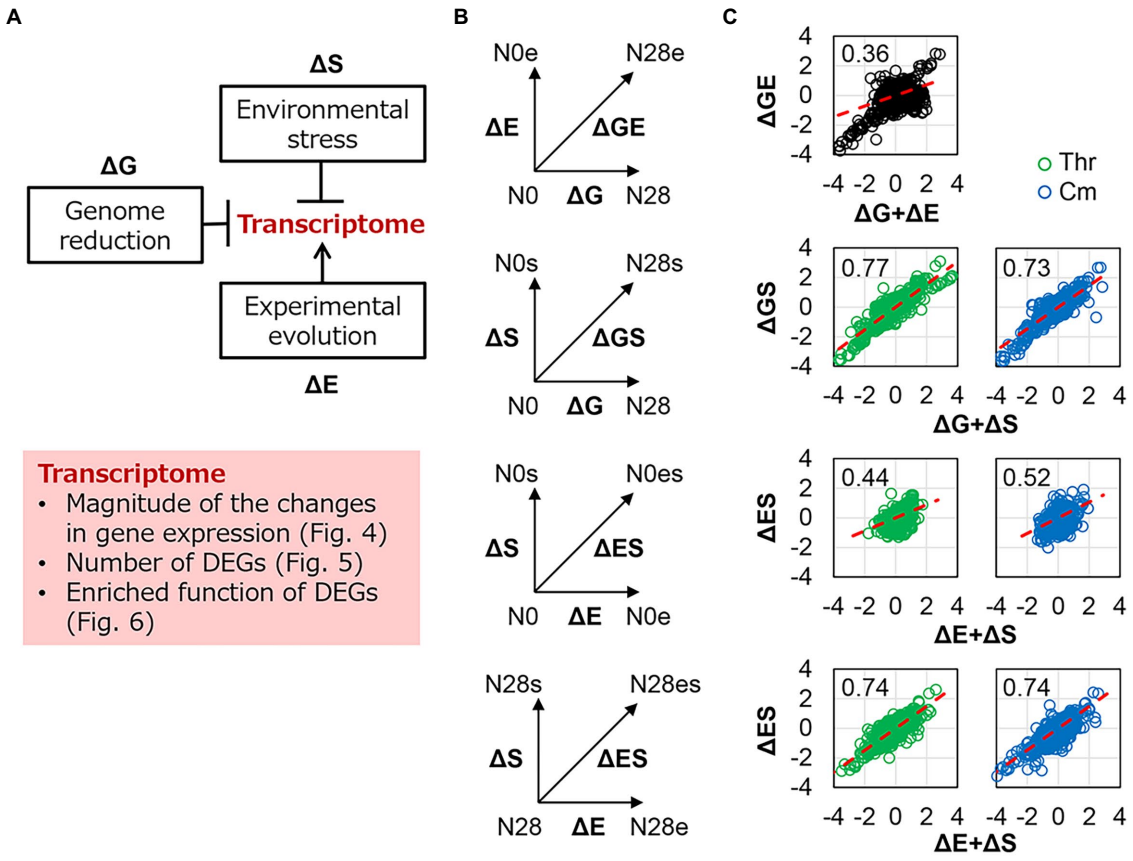


FIGURE 4
Epistasis in transcriptome reorganization. **(A)** Schematic drawing of the contributions of genomic, environmental, and evolutionary changes to transcriptomes. The box shadowed in pink highlighted the analytical points. **(B)** Schemes of evaluating the epistasis in transcriptome reorganization. ΔG, ΔE, and ΔS represent the transcriptional changes caused by the single interruption of genome reduction, environmental stress, and experimental evolution, respectively. ΔGE, ΔGS, and ΔES represent the simultaneous transcriptional changes caused by the dual interruption of any pairs. **(C)** Epistasis in transcriptome reorganization. The horizontal and vertical axes show the additive changes (the sum of the changes caused by any single interruption) and the simultaneous changes (the changes caused by dual interruptions) in gene expression, respectively. The linear regression slopes are indicated, which were theoretically interpreted in the Materials and Methods. The red dashed lines represent the slope of 1. Green and blue indicate the addition of Thr and Cm in the culture, respectively.

stress (ΔS) in transcriptome reorganization (Figure 4C, ΔGS). The findings agreed with the insignificant changes in growth fitness of the reduced genome under environmental stress (Figure 1B) and were consistent with our previous finding (Ying et al., 2013). It suggested that the magnitude of the global changes in gene expression was restricted within a certain level, which might benefit the transcriptome homeostasis, such as the conserved chromosomal periodicity (Figure 3).

Differentially expressed genes (DEGs) induced by any single or dual interruptions of the genomic, environmental, and evolutionary changes (Figures 4A,B) were determined (Figure 5A and Supplementary Figure S2). For example, 1,237, 809, and 846 DEGs were significantly identified in the ΔG-, ΔS-, and ΔGS-mediated changes in gene expression (Figure 5A), and their overlaps were further analyzed (Figure 5B, box, broken line). Comparing all pairs of DEGs (Supplementary Figure S2) observed that the DEGs were largely overlapped (Figure 5B; Supplementary Figure S3), which strongly supported the epistasis in transcriptome reorganization (Figure 4). ΔS-mediated DEGs largely overlapped with ΔG-mediated DEGs (Figure 5B, ΔGS), with up to as many as approximately 70% overlaps (Supplementary Figure S3). More ΔS-mediated DEGs were detected in N0 and more ΔE-mediated DEGs in N28 (Figure 5B, ΔES). It was consistent with the divergence in growth fitness (Figure 1C) and gene expression pattern (Figure 2) between the wild-type and reduced genomes. There was no overlap between ΔE and ΔS in N0 because there were only three DEGs in ΔE (Figure 5A, ΔGE). Note that the DEGs identified by Rank Product showed similar results (Supplementary Figure S4).

Gene functions related to the negative epistasis

Enrichment analysis was performed to determine the transcriptional regulation and genetic function participating in the transcriptome reorganization. Regulon enrichment of the DEGs showed that a single regulon σ38 and an assortment of regulons, e.g., GadW and GadX, were significant in ΔG and ΔS, respectively (Figure 6A). These regulons were known to be responsible for stress response (Loewen et al., 1998; Tucker et al., 2003).

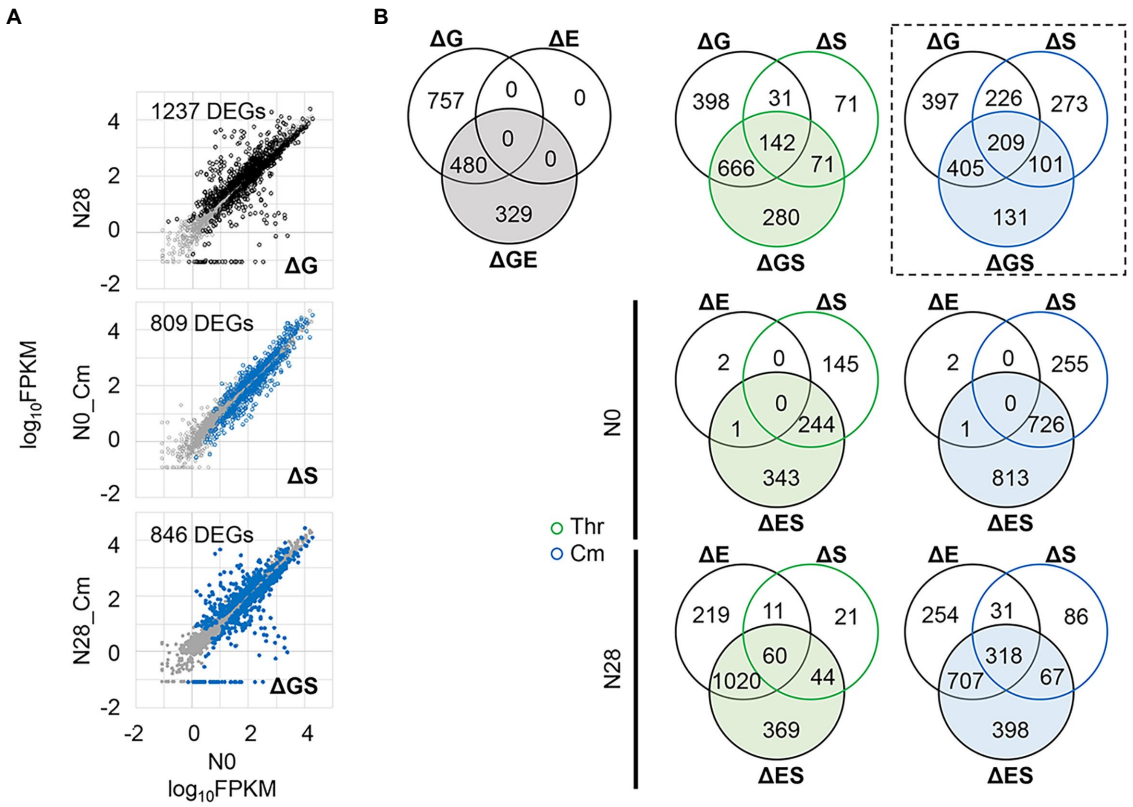


FIGURE 5
Differentially expressed genes (DEGs). **(A)** An example of identified DEGs. The comparisons of the ΔG-, ΔS-, and ΔGS-mediated changes are shown, in which the number of DEGs is indicated. The DEGs and the other genes are highlighted (in black or blue) and shown in gray, respectively. The gene expression level is shown on a logarithmic scale. **(B)** Venn diagram of DEGs. The numbers of the DEGs mediated by single and dual interruptions are shown. Green and blue indicate the addition of Thr and Cm in the culture, respectively. The box in broken lines indicates the example shown in **A**.

Nevertheless, they were insignificant in ΔGS (Figure 6A, ΔGS). The reduced number of enriched regulons was due to the dual interruption of genome reduction and environmental stress. Such offset tendency was also detected in ΔES (Figure 6A, ΔES). Gene orthology (GO) enrichment of the DEGs showed a reduced number of enriched GO terms due to the dual interruption (Supplementary Figure S5). It strongly supported the negative epistasis in transcriptome reorganization regarding biological processes. Note that the enrichment analysis failed to identify any biological function in the genes of equivalent changes in the additive and simultaneous manners (Figure 4C, upper), as these genes mostly encoded the conserved or predicted proteins (Supplementary Table S5).

Furthermore, the gene categories (GCs) of the structural component and enzyme were significantly enriched, regardless of the genomic, environmental, and evolutionary changes (Figure 6B). The offset tendency of dual interruption was also observed. The enzyme seemed to be a core GC participating in the dual interruptions. Note that the reduced number of DEGs did not result in the reduced number of enriched functions. The results indicated that the negative epistasis occurred in the gene expression and biological function. Additionally, KEGG pathway enrichment of the shared DEGs assigned in the enriched regulons and GCs showed that the amino acid metabolisms played a role in the transcriptome reorganization to a great extent (Supplementary Figure S6).

Gene modules responsible for the epistatic changes in transcriptome and fitness

To figure out the negative epistasis in the transcriptome, novel gene network construction was performed with the weighted gene co-expression network analysis (WGCNA), as described previously (Langfelder and Horvath, 2008; Bhatia et al., 2022). A total of 3,290 genes were classified into 11 modules, which comprised 30 to 1,118 genes per module (Supplementary Figure S7). Functional correlation analysis showed that only one of these modules (M4), which comprised 271 genes (Supplementary Table S3), showed a significant negative correlation to the growth rate (Figure 7A; Supplementary Table S2). Functional enrichment analysis of the genes clustered in M4 showed that the biological processes assigned in the gene orthologs (GO) related to homeostasis were considerably enriched (Figure 7C). It suggested that bacterial growth was stringently associated with functional homeostasis, consistent with the conserved chromosomal periodicity of the transcriptome (Figure 3).

Furthermore, whether the DEGs were significantly enriched in these modules was first evaluated. The modules M1, M3, M5, and M8 (Figure 7A, blue) failed to enrich the DEGs, regardless of the genomic, environmental, and evolutionary changes (Figure 7B). It suggested that these modules were conserved or irresponsible for the genomic, environmental, and evolutionary changes. The DEGs were mainly enriched in the modules M2, M4, M10, and M11 (Figure 7A, green),

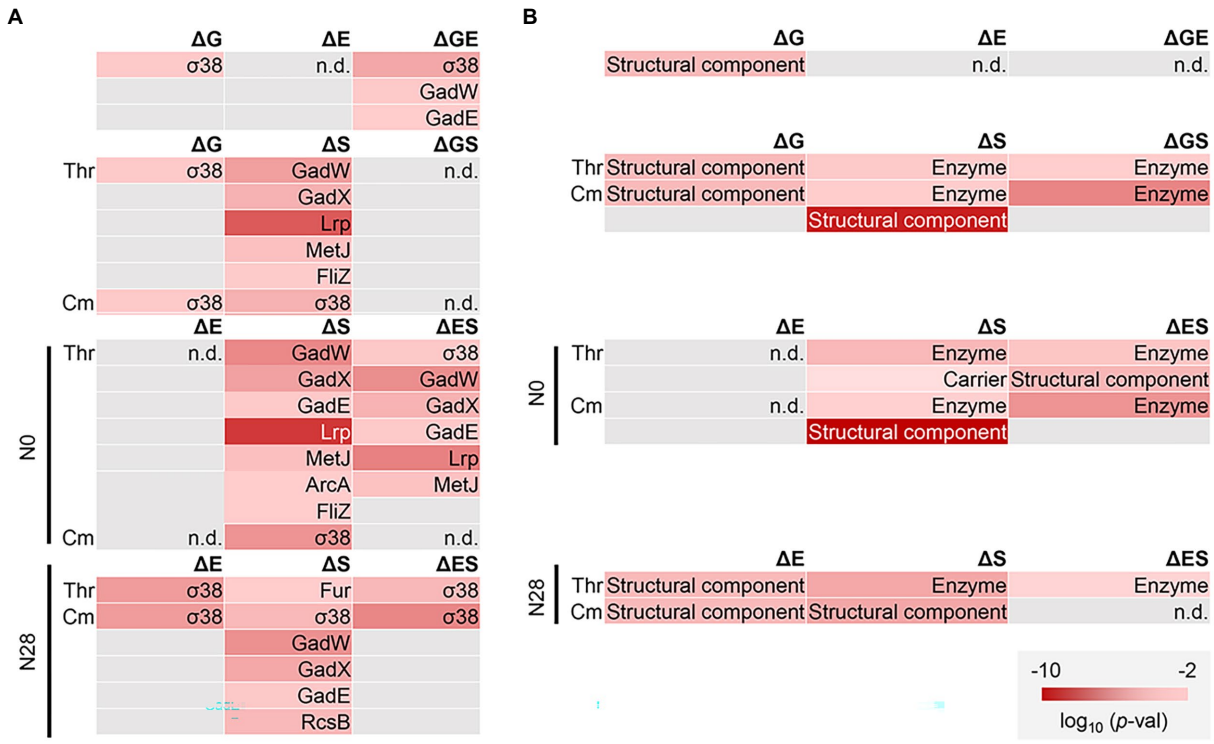


FIGURE 6
Functional enrichment of DEGs. **(A)** Regulon enrichment. Statistically significant regulons are shown. **(B)** Gene category enrichment. Statistically significant gene categories are shown. Gradation in red indicates the normalized *p* values in the logarithmic scale.

which revealed that these modules were highly responsive to individual and dual interruptions (Figure 7B). Additionally, the modules M6, M7, and M9 (Figure 7, yellow) specifically enriched the environmental stress-mediated DEGs (ΔS), which turned out to be insignificant once either genome reduction or experimental evolution happened simultaneously (ΔGS, ΔES) (Figure 7B). These three modules were somehow stringently related to the negative epistasis in growth changes and transcriptome reorganization. In summary, the newly constructed gene networks were roughly categorized into three types, i.e., conserved, responsive, and epistatic modules (Figure 7A). Functional enrichment of the epistatic modules found the metabolic process and translation mechanism were significant in M6 and M9, respectively (Supplementary Figure S8). No function was significantly enriched in M7. The varied enrichment results suggested that the negative epistasis in transcriptome reorganization was not based on any specific function but owing to the global coordination. The discriminated functions of gene modules might play a crucial role in maintaining the homeostatic transcriptome architecture for being alive in response to genomic, environmental, and evolutionary changes.

Discussion

The growth fitness, defined as the growth rate during the exponential phase, was employed to quantitatively evaluate the contribution of the genomic, environmental, and evolutionary changes. The research target in the present study was the steady growth but not the stress response of the *E. coli* strains in the presence of the stressor. To search for the proper stressors, five different additives, i.e., Thr, Cm, NaCl, ampicillin (Amp), and serine (Ser), were initially introduced to the bacterial culture because these additives were reported to reduce the bacterial growth rate (Gutierrez et al., 2013; Matsumoto et al., 2013; Rau et al., 2016). The stressors of Thr and Cm were finally chosen out of the five additives, as they successfully reduced the growth rate but maintained steady (exponential) growth of both strains (Figure 1). As known, Cm was a ribosome-binding inhibitor (Levin et al., 2017), and Thr could lead to isoleucine depletion, consequently, inhibit bacterial growth with a single carbon source (Bearer and Neet, 1978; Hama et al., 1991). They both disturbed the amino acid and protein biosynthesis, which was reported to be correlated to the growth rate of *E. coli* (Basan et al., 2015; Peebo et al., 2015; Aida et al., 2022). The related molecular mechanism might remain functional in the reduced genome N28. On the other hand, adding NaCl, Amp, and Ser failed to cause the reduced growth rate but was somehow lethal to N28 (Supplementary Figure S1). NaCl, Amp, and Ser were known to change the permeability, disturb the cell wall, and cause starvation, respectively (Munita and Arias, 2016; Rau et al., 2016; Yu and Liao, 2018). These stressors might be too severe for the reduced genome, in which the functional genes were deleted or the related molecular mechanisms were inactivated. It might cause a considerably prolonged lag time, which made the exponential growth undetectable within 48 h, the time limit of the temporal growth assay. Thus, these stressors were out of the scope of the present study.

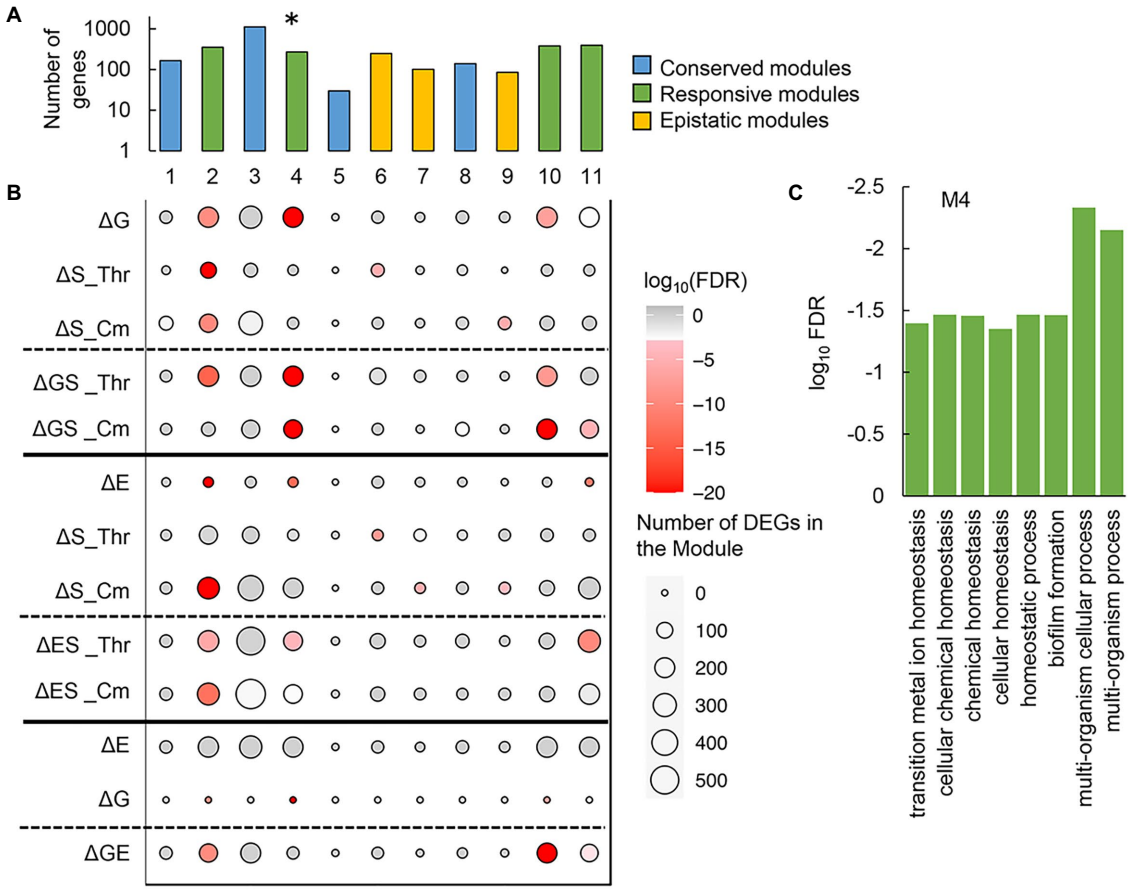


FIGURE 7
Gene modules clustered by weighted gene co-expression network analysis (WGCNA). **(A)** Gene modules newly constructed by WGCNA. A total of eleven modules newly constructed are shown. Color variation indicates the assigned types of gene modules. The asterisk indicates a significant correlation to the growth rate (p<0.01). **(B)** Enrichment of the DEGs in the gene modules. The numbers of the DEGs mediated by any of the individual and dual interruptions are represented as circles. Gradation in red indicates the statistical significance of enrichment. **(C)** Functional enrichment of M4. The enriched GO terms of high significance (FDR<0.05) are indicated. Statistical significance is shown in a logarithmic scale.

The present study applied the genetic concept of epistasis, initially used to describe the canceling effect caused by multiple mutations in adaptative evolution (Khan et al., 2011; Hartl, 2014), to interpret the canceling effect in the fitness changes responding to the dual interruption of genomic and environmental changes. The canceling effect in response to environmental changes has been found between the heat shock and alcohol stresses (Piper, 1995), and the hydrogen peroxide and ethanol stresses (Semchyshyn, 2014). These findings were known as the cross-stress resistance (Święciło, 2016), describing how the microorganisms acquired strong or multiple resistance. The so-called regulatory cross-talk (Święciło, 2016) was proposed to be the mechanism of cross-stress resistance; that is, a single transcriptional factor responded to various stresses. It might contribute to the negative epistasis observed in the present study. For example, the genes regulated by the transcription factor σ38 were commonly enriched in the ΔG- and ΔS-mediated DEGs (Figure 6A). σ38 was probably responsible for both genome reduction and environmental stress; thus, the growth change was insignificant when the dual interruption occurred (Figure 1B). Taking account of the findings of the enriched functions of translation and ribosome assembly (Figure 6 and Supplementary Figure S8) and the reduced protein synthesis by promoting the formation of 100S ribosome complexes (Trösch and Willmund, 2019), the canceling effect in growth and the negative epistasis might be attributed to the activated σ38 regulon. Note that the canceling effect was based on the fact that N28 growing slower than N0 was independent of the stressors. Although the presence of Cm of Thr could inhibit the ribosomes for translation or cause the depletion of isoleucine, either the abundance of the remaining active ribosomes or the lowered concentration of isoleucine might be sufficient for the slowly growing N28. N28 might benefit from its reduced growth rate. Despite these suspected molecular mechanisms, the negative epistasis in transcriptome reorganization could well explain the canceling effect. It demonstrated that mutual inhibition occurred among the genetic changes (i.e., mutations) and between the genomic and environmental changes.

In addition, to discover the direction of transcriptional changes, the DEGs could be further distinguished between up- and down-regulated by RankProd. Roughly, there seemed to be more downregulated DEGs than upregulated ones in N28 (Supplementary Table S6), indicating the repressed expression of the reduced genome in response to environmental and evolutionary interruptions. Either the genomic (ΔG) or the evolutionary (ΔE)

interruption caused more upregulated DEGs, whereas the dual interruptions (ΔGS, ΔGE, ΔES) led to more downregulated DEGs (Supplementary Table S6). Such reversed changes in the number of DEGs supported the epistasis in transcriptome reorganization (Figure 4). The environmental interruption triggered somehow distinguished changes depending on the stressors and genomes. More downregulated DEGs responding to Cm (ΔS_Cm) were commonly found in N0 and N28, whereas more upregulated DEGs responding to Thr (ΔS_Thr) were found in N0 and reversed in N28 (Supplementary Table S6). For instance, dadA, D-amino acid dehydrogenase, was upregulated in N0 (N0, ΔS_Thr); in contrast, no DEGs were related to amino acid metabolism in N28 (N28, ΔS_Thr), indicating that genome reduction reduced the isoleucine suppression by the additional Thr (Supplementary Table S7). The 50S ribosomal subunit proteins, e.g., *rpmD*, *rplO*, and *rplE*, were the upregulated DEGs in N0 (N0, ΔS_Cm); in contrast, no ribosome- or translation-related DEGs in N28 (N28, ΔS_Cm), indicating that genome reduction reduced the inhibition of Cm to 50S ribosome (Supplementary Table S7). The suppression of cellular function by additional Thr or Cm was somehow alleviated by genome reduction. Taken together, considerable changes must have occurred in the amino acid metabolism-, ribosome-, and translation-related functions to achieve the epistatic changes. Further studies on molecular mechanisms are required to clarify the epistatic effect observed in the genomic, environmental, and evolutionary interruptions.

In wild nature, microorganisms constantly face various environmental changes and frequently experience genomic changes, e.g., genome reduction or gene loss, to adapt to their habit (Batut et al., 2014; Giovannoni et al., 2014; Simonsen, 2022). The present study directly connected the genomic, environmental, and evolutionary changes with bacterial growth and transcriptome. Multilevel analyses successfully discovered the negative epistasis in transcriptome reorganization, which was assumed to benefit the homeostasis of transcriptome architecture. Gene network analysis revealed functional differentiation in the gene modules, illustrating an overview of the global optimization of the transcriptome for maintaining growth fitness. These novel findings provided a better understanding of how living organisms responded to the individual and dual interruptions of genome reduction, environmental stress, and adaptive evolution, compensating for known molecular mechanisms and gene functions.

Materials and methods

Bacterial strains and culture

The wild-type *E. coli* strain W3110 and the genome-reduced strain (N28) were from the National BioResource Project, National Institute of Genetics, Japan. The reduced genomes were constructed in an accumulated manner, as previously described (Mizoguchi et al., 2008). The evolved *E. coli* strains were adopted from our previous study, in which the strains have undergone experimental evolution in the minimal medium M63 for approximately 1,000 generations (Nishimura et al., 2017). The M63 medium was used for bacterial culture, of which the composition and preparation were described in detail elsewhere (Ishizawa et al., 2015). The chemical compounds, i.e., threonine, serine, NaCl, ampicillin, and chloramphenicol, were commercially available (Wako). Their concentrations in the culture were 100 mM, 75 mM, 0.45 M, 4 μg/mL, and 1.2 μg/mL, respectively, which were determined as the condition where the growth rate of the wild type (N0) decreased by ~30% from the regular culture. The bacterial culture was performed in a 96-well microplate for growth assay and a test tube for RNA sequencing, as described following.

Growth assay

Every 200 μL of bacterial culture was dispensed to each well of the 96-well microplate (Coaster), as previously described (Kurokawa et al., 2016; Kurokawa and Ying, 2017). In brief, the microplate was incubated at 37°C in a plate reader (EPOCH2, BioTek) shaking at 567 cpm (cycles per minute) for 48 h. The temporal changes in OD_{600} were measured at 30-min intervals. The growth rate was calculated using the following equation between any two consecutive points (Eq. 1)

$$\mu_i = \frac{\ln\left(\frac{C_{i+1}}{C_i}\right)}{t_{i+1} - t_i} \quad (1)$$

Here, t_i and t_{i+1} are the culture times at the two consecutive measurement points, and C_i and C_{i+1} are the OD_{600} at time points t_i and t_{i+1}. The average of three consecutive μ_i showing the largest mean and the smallest variance was determined as the growth rate. The mean of the biological triplicates was defined as the growth fitness and used in the analyses.

RNA sequencing

Every 5 mL bacterial culture in a test tube was incubated in a bioshaker (BR-23FP, Taitec) at 200 rpm, 37°C. A precision particle size analyzer (Multisizer 4, Beckman Coulter) with an aperture of a pore size of 20 μL was used to evaluate the cell concentration. 20 μL of culture was suspended in a dedicated 25 mL cuvette (Beckman Coulter) containing 10 mL of diluent (Isoton II, Beckman Coulter). The *E. coli* cells were collected during the exponential growth phase (i.e., $5 \times 10^7 \sim 2 \times 10^8$ cells/mL). The test tubes were quickly transferred into ice and mixed with Falcon containing 5 mL of a 10% phenol ethanol solution. The Falcon was subsequently centrifuged at 7,000 rpm, 4°C, for 3 min, the supernatant was removed, and the pellet was frozen at −80°C for future use. The pellets were thawed for at least 10 min, and the total RNA was purified using the RNeasy Mini Kit (QIAGEN) and RNase-Free DNase Set (QIAGEN) according to product instructions. Purified total RNA was dissolved in RNase-free water and frozen at −80°C. The rRNAs were removed using the Ribo-Zero Plus rRNA Depletion Kit (Illumina), and the mRNA libraries were prepared using the Ultra Directional RNA Library Prep Kit for Illumina (NEBNext). The paired-end sequencing (150 bp × 2) was performed using the Novaseq6000 next-generation sequencer (Illumina). Biological replications were performed for all conditions ($N = 2 \sim 6$; N0 and N28 under regular conditions were repeated six and four times, respectively; and the others were all duplicates). The raw data sets were deposited in the DDBJ Sequence Read Archive under the accession numbers DRA013430, DRA013683, and DRA015318.

Data processing and normalization

The reference genome W3110, obtained from GenBank under the accession number AP009048.1, was mapped for the paired-end FASTQ obtained by RNAseq. The mismatch parameter (mp) was set to 2 using the mapping software Bowtie 2 (Langmead and Salzberg, 2012). The obtained read counts were converted to FPKM values according to the gene length and total read count values. Global normalization of the FPKM values was performed to reach an identical mean value in the logarithmic scale in all datasets. The gene expression level was determined as the logarithmic value of FPKM, and the biological replicates were all summarized in Supplementary Table S4. The dataset was used for the following computational analyses.

Computational analyses

Computational analyses were performed with R. Hierarchical clustering (HC) and principal component analysis (PCA) were performed using the functions of "clusterSample" and "prcomp," respectively. In HC, the "dist.method" and "hclust.method" were set as "spearman" and "ward.D2," respectively. In PCA, the "scale" was set as "F." A total of 3,290 genes common in the wild type and the reduced genomes were used in HC and PCA. Differentially expressed genes (DEGs) were identified using the R package DESeq2 (Love et al., 2014) and the Bioconductor software package RankProd (Breitling et al., 2004; Hong et al., 2006). The read counts and logarithmic FPKM values were used as the input data for DESeq2 and RankProd analyses. The DEGs were determined according to the false discovery rate (FDR) (Storey, 2002). 4,427 and 3,290 genes were used to identify the DEGs for the wild type and the reduced genomes, respectively. In the comparison within the same genome of either N0 or N28 (ΔE, ΔS, and ΔES), either 4,227 or 3,290 genes were used in the analysis (Supplementary Table S6). Once comparing the two genomes of N0 and N28 (ΔG, ΔGS, and ΔGE), the common 3,290 genes in both genomes were used in the analysis.

Functional enrichment

Functional enrichments of the DEGs were performed according to the features of transcriptional regulation, gene category, gene ontology, and metabolic pathway. A total of 53 transcriptional factors (TFs) (Salgado et al., 2013) comprising more than 10 regulatees and 19 gene categories (GCs) (Riley et al., 2006) of more than 30 genes were subjected to enrichment analysis, as previously described (Ying et al., 2015; Ying and Yama, 2018). A total of 20 GCs were used in the analysis. The statistical significance of the DEGs enriched in TFs and GCs was evaluated by the binomial test with Bonferroni correction. The functional enrichment of gene ontology (GO) (Ashburner et al., 2000; Carbon et al., 2009) and Kyoto encyclopedia of genes and genomes (KEGG) pathway (Kanehisa and Goto, 2000; Kanehisa et al., 2016) analysis was performed using the Database for Annotation Visualization and Integrated Discovery version 6.7 (DAVID) (Huang da et al., 2009). The statistical significance was according to FDR.

Chromosomal periodicity analysis

A standard Fourier transform was used to evaluate the chromosomal periodicity of the transcriptome, as previously described (Nagai et al., 2020). The periodicity analysis was performed using the function of "periodogram" in the R package. The CDS information was obtained from the DDBJ databank under the accession number AP009048. The periodicity was evaluated with a sliding distance of 1 kb and shown in 100-kb bins of genomic length. The approximate curves of the periodicity were calculated using the max peaks of the periodograms. They were fitted by minimizing the square error of the approximate curves and the series of expression values. The genomic position of *ori* was according to the previous reports (Mizoguchi et al., 2008; Kurokawa et al., 2016). The statistical significance of the periodicity was assessed with Fisher's g test (Wichert et al., 2004), which was performed using the GeneCycle package in R.

Epistasis analysis

The epistasis in transcriptome reorganization was performed as described previously (Ying et al., 2013). The transcriptional changes caused by genome reduction or environmental stressors for a given gene were denoted as ΔG or ΔS, respectively. The changes caused by the genome reduction and the environmental stressor simultaneously or separately were defined as ΔGS or ΔG + ΔS, respectively. The linear regression of ΔGS and ΔG + ΔS was performed using the least squares method (Eq. 2). The epistasis in transcriptome reorganization was determined according to the slope of α, as follows.

$$\Delta GS = \alpha\left(\Delta G + \Delta S\right) \tag{2}$$

α = 1, additivity
α > 1, positive epistasis
α < 1, negative epistasis.

Gene network analysis

The weighted gene co-expression network analysis (WGCNA) was performed with the R package of WGCNA (Langfelder and Horvath, 2008) under the developer's instruction. A total of 3,290 genes commonly encoded in the N0 and N28 genomes were subjected to the gene network analysis, where the logarithmic FPKM values were used as the input data. The "step-by-step" method was used to determine the parameters for network construction. The soft threshold of the co-expression network clustering was decided by Scale Free Topology Model Fit, where the R^2 was approximately 0.9. The FPKM values were converted into the topological overlap matrix (TOM). Hierarchical clustering was performed to divide the genes into various co-expression modules with the "dynamic tree cut" method. The gene modules of similar expression profiles were merged according to the module eigengenes using the "mergeCloseModules" function with a height cut of 0.25. The parameters "softPower," "minModuleSize," and "MEDissThres" were set at 12, 50, and 0.25, respectively. The other parameters

were set as default. A total of 3,290 genes common in the wild type and reduced genomes were used to construct the network. The statistical significance of the DEGs enriched in the merged modules was evaluated by the binomial test with Bonferroni correction as described above.

Data availability statement

The data presented in the study are deposited in the DDBJ Sequence Read Archive, accession numbers DRA013430, DRA013683, and DRA015318.

Author contributions

YM performed the experiments and draft the manuscript. YM, MN, and B-WY analyzed the data. B-WY conceived the research and rewrote the manuscript. All authors approved the final manuscript.

Funding

This work was supported by the JSPS KAKENHI Grant-in-Aid for Scientific Research (B) (grant number 19H03215) and partially by Grant-in-Aid for Challenging Exploratory Research (grant number 21 K19815).

Acknowledgments

We thank NBRP for providing the *E. coli* strains carrying the wild-type and reduced genomes.

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.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1145673/full#supplementary-material

References

Aida, H., Hashizume, T., Ashino, K., and Ying, B. W. (2022). Machine learning-assisted discovery of growth decision elements by relating bacterial population dynamics to environmental diversity. *Elife* 11:e76846. doi: 10.7554/eLife.76846

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: Tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., et al. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol. Syst. Biol.* 2:2006.0008. doi: 10.1038/msb4100050

Basan, M., Honda, T., Christodoulou, D., Hörl, M., Chang, Y. F., Leoncini, E., et al. (2020). A universal trade-off between growth and lag in fluctuating environments. *Nature* 584, 470–474. doi: 10.1038/s41586-020-2505-4

Basan, M., Zhu, M., Dai, X., Warren, M., Sévin, D., Wang, Y. P., et al. (2015). Inflating bacterial cells by increased protein synthesis. *Mol. Syst. Biol.* 11:836. doi: 10.15252/msb.20156178

Batut, B., Knibbe, C., Marais, G., and Daubin, V. (2014). Reductive genome evolution at both ends of the bacterial population size spectrum. *Nat. Rev. Microbiol.* 12, 841–850. doi: 10.1038/nrmicro3331

Bearer, C. F., and Neet, K. E. (1978). Threonine inhibition of the aspartokinase--homoserine dehydrogenase I of *Escherichia coli*. Threonine binding studies. *Biochemistry* 17, 3512–3516. doi: 10.1021/bi00610a014

Bhatia, R. P., Kirit, H. A., Predeus, A. V., and Bollback, J. P. (2022). Transcriptomic profiling of *Escherichia coli* K-12 in response to a compendium of stressors. *Sci. Rep.* 12:8788. doi: 10.1038/s41598-022-12463-3

Breitling, R., Armengaud, P., Amtmann, A., and Herzyk, P. (2004). Rank products: A simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett.* 573, 83–92. doi: 10.1016/j.febslet.2004.07.055

Carbon, S., Ireland, A., Mungall, C. J., Shu, S. Q., Marshall, B., Lewis, S., et al. (2009). AmiGO: Online access to ontology and annotation data. *Bioinformatics* 25, 288–289. doi: 10.1093/bioinformatics/btn615

Choe, D., Lee, J. H., Yoo, M., Hwang, S., Sung, B. H., Cho, S., et al. (2019). Adaptive laboratory evolution of a genome-reduced *Escherichia coli*. *Nat. Commun.* 10:935. doi: 10.1038/s41467-019-08888-6

Chu, X., Li, S., Wang, S., Luo, D., and Luo, H. (2021). Gene loss through pseudogenization contributes to the ecological diversification of a generalist Roseobacter lineage. *ISME J.* 15, 489–502. doi: 10.1038/s41396-020-00790-0

Cortez, D., Neira, G., González, C., Vergara, E., and Holmes, D. S. (2022). A large-scale genome-based survey of acidophilic bacteria suggests that genome streamlining is an adaption for life at low pH. *Front. Microbiol.* 13:803241. doi: 10.3389/fmicb.2022.803241

Feher, T., Papp, B., Pal, C., and Posfai, G. (2007). Systematic genome reductions: Theoretical and experimental approaches. *Chem. Rev.* 107, 3498–3513. doi: 10.1021/cr0683111

Feugeas, J. P., Tourret, J., Launay, A., Bouvet, O., Hoede, C., Denamur, E., et al. (2016). Links between transcription, environmental adaptation and gene variability in *Escherichia coli*: Correlations between gene expression and gene variability reflect growth efficiencies. *Mol. Biol. Evol.* 33, 2515–2529. doi: 10.1093/molbev/msw105

Giovannoni, S. J., Cameron Thrash, J., and Temperton, B. (2014). Implications of streamlining theory for microbial ecology. *ISME J.* 8, 1553–1565. doi: 10.1038/ismej.2014.60

Gutierrez, A., Laureti, L., Crussard, S., Abida, H., Rodríguez-Rojas, A., Blázquez, J., et al. (2013). β-Lactam antibiotics promote bacterial mutagenesis via an RpoS-mediated reduction in replication fidelity. *Nat. Commun.* 4:1610. doi: 10.1038/ncomms2607

Hama, H., Kayahara, T., Tsuda, M., and Tsuchiya, T. (1991). Inhibition of homoserine dehydrogenase I by L-serine in *Escherichia coli*. *J. Biochem.* 109, 604–608. doi: 10.1093/oxfordjournals.jbchem.a123427

Hartl, D. L. (2014). What can we learn from fitness landscapes? *Curr. Opin. Microbiol.* 21, 51–57. doi: 10.1016/j.mib.2014.08.001

Hong, F., Breitling, R., McEntee, C. W., Wittner, B. S., Nemhauser, J. L., and Chory, J. (2006). RankProd: A bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics* 22, 2825–2827. doi: 10.1093/bioinformatics/btl476

Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211

Ishizawa, Y., Ying, B. W., Tsuru, S., and Yomo, T. (2015). Nutrient-dependent growth defects and mutability of mutators in *Escherichia coli*. *Genes Cells* 20, 68–76. doi: 10.1111/gtc.12199

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070

Karcagi, I., Draskovits, G., Umenhoffer, K., Fekete, G., Kovács, K., Méhi, O., et al. (2016). Indispensability of horizontally transferred genes and its impact on bacterial genome streamlining. *Mol. Biol. Evol.* 33, 1257–1269. doi: 10.1093/molbev/msw009

Kato, J., and Hashimoto, M. (2007). Construction of consecutive deletions of the *Escherichia coli* chromosome. *Mol. Syst. Biol.* 3:132. doi: 10.1038/msb4100174

Kawecki, T. J., Lenski, R. E., Ebert, D., Hollis, B., Olivieri, I., and Whitlock, M. C. (2012). Experimental evolution. *Trends Ecol. Evol.* 27, 547–560. doi: 10.1016/j.tree.2012.06.001

Khan, A. I., Dinh, D. M., Schneider, D., Lenski, R. E., and Cooper, T. F. (2011). Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* 332, 1193–1196. doi: 10.1126/science.1203801

Kinsler, G., Geiler-Samerotte, K., and Petrov, D. A. (2020). Fitness variation across subtle environmental perturbations reveals local modularity and global pleiotropy of adaptation. *Elife* 9:e61271. doi: 10.7554/eLife.61271

Klumpp, S., and Hwa, T. (2014). Bacterial growth: Global effects on gene expression, growth feedback and proteome partition. *Curr. Opin. Biotechnol.* 28, 96–102. doi: 10.1016/j.copbio.2014.01.001

Klumpp, S., Zhang, Z., and Hwa, T. (2009). Growth rate-dependent global effects on gene expression in bacteria. *Cells* 139, 1366–1375. doi: 10.1016/j.cell.2009.12.001

Kurokawa, M., Nishimura, I., and Ying, B. W. (2022). Experimental evolution expands the breadth of adaptation to an environmental gradient correlated with genome reduction. *Front. Microbiol.* 13:826894. doi: 10.3389/fmicb.2022.826894

Kurokawa, M., Seno, S., Matsuda, H., and Ying, B. W. (2016). Correlation between genome reduction and bacterial growth. *DNA Res.* 23, 517–525. doi: 10.1093/dnares/dsw035

Kurokawa, M., and Ying, B. W. (2017). Precise, high-throughput analysis of bacterial growth. *J. Vis. Exp.* 127:56197. doi: 10.3791/56197

Kurokawa, M., and Ying, B. W. (2019). Experimental challenges for reduced genomes: The cell model *Escherichia coli*. *Microorganisms* 8:3. doi: 10.3390/microorganisms8010003

Langfelder, P., and Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* 9:559. doi: 10.1186/1471-2105-9-559

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lao, Z., Matsui, Y., Ijichi, S., and Ying, B. W. (2022). Global coordination of the mutation and growth rates across the genetic and nutritional variety in *Escherichia coli*. *Front. Microbiol.* 13:990969. doi: 10.3389/fmicb.2022.990969

Levin, B. R., McCall, I. C., Perrot, V., Weiss, H., Ovesepian, A., and Baquero, F. (2017). A numbers game: Ribosome densities, bacterial growth, and antibiotic-mediated stasis and death. *mBio* 8:e02253-16. doi: 10.1128/mBio.02253-16

Lewis, W. H., Tahon, G., Geesink, P., Sousa, D. Z., and Ettema, T. J. G. (2021). Innovations to culturing the uncultured microbial majority. *Nat. Rev. Microbiol.* 19, 225–240. doi: 10.1038/s41579-020-00458-8

Liu, L., Kurokawa, M., Nagai, M., Seno, S., and Ying, B. W. (2020). Correlated chromosomal periodicities according to the growth rate and gene expression. *Sci. Rep.* 10:15531. doi: 10.1038/s41598-020-72389-6

Loewen, P. C., Hu, B., Strutinsky, J., and Sparling, R. (1998). Regulation in the rpoS regulon of *Escherichia coli*. *Can. J. Microbiol.* 44, 707–717. doi: 10.1139/w98-069

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Malik, M., Lu, T., Zhao, X., Singh, A., Hattan, C. M., Domagala, J., et al. (2005). Lethality of quinolones against mycobacterium smegmatis in the presence or absence of chloramphenicol. *Antimicrob. Agents Chemother.* 49, 2008–2014. doi: 10.1128/AAC.49.5.2008-2014.2005

Martinez-Cano, D. J., Reyes-Prieto, M., Martinez-Romero, E., Partida-Martínez, L. P., Latorre, A., Moya, A., et al. (2014). Evolution of small prokaryotic genomes. *Front. Microbiol.* 5:742. doi: 10.3389/fmicb.2014.00742

Mathelier, A., and Carbone, A. (2010). Chromosomal periodicity and positional networks of genes in *Escherichia coli*. *Mol. Syst. Biol.* 6:366. doi: 10.1038/msb.2010.21

Matsumoto, Y., Murakami, Y., Tsuru, S., Ying, B. W., and Yomo, T. (2013). Growth rate-coordinated transcriptome reorganization in bacteria. *BMC Genomics* 14:808. doi: 10.1186/1471-2164-14-808

Mizoguchi, H., Mori, H., and Fujio, T. (2007). *Escherichia coli* minimum genome factory. *Biotechnol. Appl. Biochem.* 46, 157–167. doi: 10.1042/BA20060107

Mizoguchi, H., Sawano, Y., Kato, J., and Mori, H. (2008). Superpositioning of deletions promotes growth of *Escherichia coli* with a reduced genome. *DNA Res.* 15, 277–284. doi: 10.1093/dnares/dsn019

Morimoto, T., Kadoya, R., Endo, K., Tohata, M., Sawada, K., Liu, S., et al. (2008). Enhanced recombinant protein productivity by genome reduction in Bacillus subtilis. *DNA Res.* 15, 73–81. doi: 10.1093/dnares/dsn002

Munita, J. M., and Arias, C. A. (2016). Mechanisms of antibiotic resistance. *Microbiol. Spectr.* 4, 1–24. doi: 10.1128/microbiolspec.VMBF-0016-2015

Murakami, Y., Matsumoto, Y., Tsuru, S., Ying, B. W., and Yomo, T. (2015). Global coordination in adaptation to gene rewiring. *Nucleic Acids Res.* 43, 1304–1316. doi: 10.1093/nar/gku1366

Nagai, M., Kurokawa, M., and Ying, B. W. (2020). The highly conserved chromosomal periodicity of transcriptomes and the correlation of its amplitude with the growth rate in *Escherichia coli*. *DNA Res.* 27:dsaa018. doi: 10.1093/dnares/dsaa018

Nicks, T., and Rahn-Lee, L. (2017). Inside out: Archaeal Ectosymbionts suggest a second model of reduced-genome evolution. *Front. Microbiol.* 8:384. doi: 10.3389/fmicb.2017.00384

Nishimura, I., Kurokawa, M., Liu, L., and Ying, B. W. (2017). Coordinated changes in mutation and growth rates induced by genome reduction. *mBio* 8:e00676-17. doi: 10.1128/mBio.00676-17

Peebo, K., Valgepea, K., Maser, A., Nahku, R., Adamberg, K., and Vilu, R. (2015). Proteome reallocation in *Escherichia coli* with increasing specific growth rate. *Mol. BioSyst.* 11, 1184–1193. doi: 10.1039/C4MB00721B

Piper, P. W. (1995). The heat shock and ethanol stress responses of yeast exhibit extensive similarity and functional overlap. *FEMS Microbiol. Lett.* 134, 121–127. doi: 10.1111/j.1574-6968.1995.tb07925.x

Rau, M. H., Calero, P., Lennen, R. M., Long, K. S., and Nielsen, A. T. (2016). Genome-wide *Escherichia coli* stress response and improved tolerance towards industrially relevant chemicals. *Microb. Cell Factories* 15:176. doi: 10.1186/s12934-016-0577-5

Riley, M., Abe, T., Arnaud, M. B., Berlyn, M. K., Blattner, F. R., Chaudhuri, R. R., et al. (2006). *Escherichia coli* K-12: A cooperatively developed annotation snapshot--2005. *Nucleic Acids Res.* 34, 1–9. doi: 10.1093/nar/gkj405

Sabath, N., Ferrada, E., Barve, A., and Wagner, A. (2013). Growth temperature and genome size in bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation. *Genome Biol. Evol.* 5, 966–977. doi: 10.1093/gbe/evt050

Salgado, H., Peralta-Gil, M., Gama-Castro, S., Santos-Zavaleta, A., Muñiz-Rascado, L., García-Sotelo, J. S., et al. (2013). RegulonDB v8.0: Omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res.* 41, D203–D213. doi: 10.1093/nar/gks1201

Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z., and Hwa, T. (2010). Interdependence of cell growth and gene expression: Origins and consequences. *Science* 330, 1099–1102. doi: 10.1126/science.1192588

Semchyshyn, H. M. (2014). Hormetic concentrations of hydrogen peroxide but not ethanol induce cross-adaptation to different stresses in budding yeast. *Int. J. Microbiol.* 2014:485792. doi: 10.1155/2014/485792

Simonsen, A. K. (2022). Environmental stress leads to genome streamlining in a widely distributed species of soil bacteria. *ISME J.* 16, 423–434. doi: 10.1038/s41396-021-01082-x

Storey, J. D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B.* 64, 479–498. doi: 10.1111/1467-9868.00346

Święciło, A. (2016). Cross-stress resistance in *Saccharomyces cerevisiae* yeast--new insight into an old phenomenon. *Cell Stress Chaperones* 21, 187–200. doi: 10.1007/s12192-016-0667-7

Trösch, R., and Willmund, F. (2019). The conserved theme of ribosome hibernation: From bacteria to chloroplasts of plants. *Biol. Chem.* 400, 879–893. doi: 10.1515/hsz-2018-0436

Tucker, D. L., Tucker, N., Ma, Z., Foster, J. W., Miranda, R. L., Cohen, P. S., et al. (2003). Genes of the GadX-GadW regulon in *Escherichia coli*. *J. Bacteriol.* 185, 3190–3201. doi: 10.1128/JB.185.10.3190-3201.2003

Vernyik, V., Karcagi, I., Timar, E., Nagy, I., Gyorkei, A., Papp, B., et al. (2020). Exploring the fitness benefits of genome reduction in *Escherichia coli* by a selection-driven approach. *Sci. Rep.* 10:7345. doi: 10.1038/s41598-020-64074-5

Wichert, S., Fokianos, K., and Strimmer, K. (2004). Identifying periodically expressed transcripts in microarray time series data. *Bioinformatics* 20, 5–20. doi: 10.1093/bioinformatics/btg364

Wytock, T. P., and Motter, A. E. (2019). Predicting growth rate from gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 116, 367–372. doi: 10.1073/pnas.1808080116

Ying, B. W., Matsumoto, Y., Kitahara, K., Suzuki, S., Ono, N., Furusawa, C., et al. (2015). Bacterial transcriptome reorganization in thermal adaptive evolution. *BMC Genomics* 16:802. doi: 10.1186/s12864-015-1999-x

Ying, B. W., Seno, S., Kaneko, F., Matsuda, H., and Yomo, T. (2013). Multilevel comparative analysis of the contributions of genome reduction and heat shock to the *Escherichia coli* transcriptome. *BMC Genomics* 14:25. doi: 10.1186/1471-2164-14-25

Ying, B. W., and Yama, K. (2018). Gene expression order attributed to genome reduction and the steady cellular state in *Escherichia coli*. *Front. Microbiol.* 9:2255. doi: 10.3389/fmicb.2018.02255

Yu, H., and Liao, J. C. (2018). A modified serine cycle in *Escherichia coli* coverts methanol and CO_2 to two-carbon compounds. *Nat. Commun.* 9:3992. doi: 10.1038/s41467-018-06496-4

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 29 March 2023
DOI 10.3389/fmicb.2023.1102250

OPEN ACCESS

EDITED BY
Feng Gao,
Tianjin University, China

REVIEWED BY
Francisco Barona-Gomez,
Leiden University, Netherlands
Michele Cloete,
Agricultural Research Council-Vegetable and Ornamental Plants (ARC-VOPI), South Africa

*CORRESPONDENCE
Xiao-Yang Zhi
✉ xyzhi@ynu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 18 November 2022
ACCEPTED 14 March 2023
PUBLISHED 29 March 2023

CITATION
Wang M, Li C-J, Zhang Z, Li P-P, Yang L-L and Zhi X-Y (2023) The evolution of morphological development is congruent with the species phylogeny in the genus *Streptomyces*. *Front. Microbiol.* 14:1102250. doi: 10.3389/fmicb.2023.1102250

The evolution of morphological development is congruent with the species phylogeny in the genus *Streptomyces*

Min Wang[1,2†], Cong-Jian Li[1†], Zhen Zhang[1], Pan-Pan Li[1], Ling-Ling Yang[1] and Xiao-Yang Zhi[1*]

[1]Key Laboratory of Microbial Diversity in Southwest China of Ministry of Education, School of Life Sciences, Yunnan Institute of Microbiology, Yunnan University, Kunming, China, [2]Zhaotong Health Vocational College, Zhaotong, China

As the canonical model organism to dissect bacterial morphological development, *Streptomyces* species has attracted much attention from the microbiological society. However, the evolution of development-related genes in *Streptomyces* remains elusive. Here, we evaluated the distribution of development-related genes, thus indicating that the majority of these genes were ubiquitous in *Streptomyces* genomes. Furthermore, the phylogenetic topologies of related strict orthologous genes were compared to the species tree of *Streptomyces* from both concatenation and single-gene tree analyses. Meanwhile, the reconciled gene tree and normalization based on the number of parsimony-informative sites were also employed to reduce the impact of phylogenetic conflicts, which was induced by uncertainty in single-gene tree inference based merely on the sequence and the bias in the amount of phylogenetic information caused by variable numbers of parsimony-informative sites. We found that the development-related genes had higher congruence to the species tree than other strict orthologous genes. Considering that the development-related genes could also be tracked back to the common ancestor of *Streptomyces*, these results suggest that morphological development follows the same pattern as species divergence.

KEYWORDS

***Streptomyces*, morphological development, comparative phylogenetics, phylogenetic conflict, Robinson-Foulds distance**

Introduction

Streptomyces are Gram-positive and filamentous bacteria belonging to the phylum *Actinobacteria*. They have complex multicellular life cycles involving the transformation of a vegetative mycelium into the reproductive spores that are essential for their propagation (Kämpfer, 2015). Moreover, it is also the largest genus of prokaryotes, comprising 685 species with a validly published name (827 species including synonyms, data from LPSN)[1] (Parte et al., 2020). The flourishing of species diversity is undoubtedly the result of its

1 https://lpsn.dsmz.de/genus/streptomyces

broad ecological adaptability (Andam et al., 2016; Li et al., 2019) and is also linked to its irreplaceable role in producing commercially and medically important bioactive substances (Anderson and Wellington, 2001). As we know, morphological differentiation and secondary metabolism are two physiologically coupled biological processes that strongly underpin the environmental adaptability of *Streptomyces* species (Sun et al., 2017; Xu et al., 2019). They have attracted the most interest and scrutiny of microbiological society.

The life cycle of *Streptomyces* species begins with the germination of a spore, followed by the formation of vegetative mycelia, including tip extension and the initiation of new branches (Flärdh and Buttner, 2009; van Dissel et al., 2014). Then, in response to nutrient depletion and other signals, the morphological differentiation and the production of secondary metabolites is initiated (Chater, 2006). Most of the genes involved in this intricate developmental process, e.g., *bld* genes related to the deficiency of aerial mycelium (Willey et al., 1991), *whi* genes related to spore formation (Soliveri et al., 2000), *ram* genes related to the hydrophobic protein synthesis (Capstick et al., 2007), *ssg* genes associated with the sporulation-specific cell division (Kormanec and Sevcikova, 2002), were considerably widespread in *Streptomyces*. Intriguingly, these two coupled biological processes were governed by complex regulatory networks (Bibb, 2005; Ohnishi et al., 2005). However, distinct from morphological differentiation, the secondary metabolism exhibits a high level of species specificity (Liu et al., 2013). Only some genes involved in regulation and signal transduction might be conserved among *Streptomyces* species (Bibb, 2005; Rodríguez et al., 2013).

Comparative genomics indicated that morphological development and its regulation in *Streptomyces* was probably achieved through the progressive acquisition of the laterally transferred DNA (Chater and Chandra, 2006). But most development-related genes may have been assembled in ancestors of *Streptomyces* species, at least before *Streptomyces* species branched off from the actinobacterial phylogenetic tree (Chandra and Chater, 2014). In this respect, *Streptomyces* differs from cyanobacteria, which also have the capacity for multicellular differentiation and a more complex early evolutionary history on the filamentous morphology (Sanchez-Baracaldo et al., 2005; Schirrmeister et al., 2011; Hammerschmidt et al., 2021). After that, the pan-genome analysis reaffirmed that many genes involved in stress response and morphological development were commonly expressed in *Streptomyces* species (Kim et al., 2015). Since 2013, fueled by rapid advances in high-throughput sequencing, the expansion of genome data of *Streptomyces*, especially the non-model streptomycetes (Lee et al., 2020), will aid in better comprehending the evolution of *Streptomyces* development. The relationship between gene phylogenies of specific core biosynthetic gene clusters and species phylogeny has been uncovered separately in a limited number of species (Joynt and Seipke, 2018; Waglechner et al., 2019; Navarro-Muñoz et al., 2020; Creamer et al., 2021). Nevertheless, the phylogenies of morphological development-related genes and their relationships with the phylogeny of species in *Streptomyces* remain elusive.

In this study, to address the above question, we performed an in-depth comparative phylogenetic investigation based on a large genome dataset of *Streptomyces*. The distributions of 65 development-related genes supported the previous conclusion that these genes were progressively assembled during the early evolution of *Streptomyces,* and some genes had been equipped in the genome of the common ancestor. Remarkably, the phylogenetic comparisons revealed that development-related genes had a higher degree of topological congruences to the species tree for both concatenated and single-gene trees. Whether by reconciling single-gene trees with the species tree using amalgamated likelihood estimation or by normalizing the topological distance of single-gene trees with a linear regression model, the development-related genes are always the gene sets that are considerably concordant with the species tree. These results demonstrated that the evolution of morphological development of *Streptomyces* kept in step with the speciation of this largest taxonomic taxon over the past 380 million years (McDonald and Currie, 2017).

Materials and methods

Genome and gene datasets preparation

The genome data (≥5 Mb) of type strains of 403 validly published *Streptomyces* species were collected from NCBI Assembly (344 genomes), GCM Type (43 genomes), and JGI IMG (16 genomes) databases (**Supplementary Table 1**). To reduce the redundancy of sequence data, the genome-pairwise average nucleotide identities (ANIs) were calculated by FastANI (Jain et al., 2018). And, if the genomes of two species shared an ANI value greater than 95%, which is the threshold for demarcating prokaryotic species (Chun et al., 2018), the genome of the species whose taxonomic name that had the priority of publication was retained (Li et al., 2022). Therefore, 388 genomes constituted a basic genome dataset (BGD) and were annotated using Prokka (version 1.12) with default parameters (Seemann, 2014). It needs to be stated that all subsequent phylogenetic analyses are based on the protein sequence. Firstly, 40 universal marker genes were extracted from BGD by SpecI (Mende et al., 2013), and the resultant gene datasets were denoted as UGs. Genomes with less than 38 extracted universal markers were further excluded from the BGD (**Supplementary Table 2**). Secondly, to collect the translated products of development-related genes in BGD, the reference protein sequences encoded by 65 development-related genes were retrieved from UniProt[2] (**Table 1**) and used as queries to search against BGD by blastp (Altschul et al., 1990) with an *e*-value cut-off of 1×10^{-5}. In each genome, all hits with identity ≥40% and coverage ≥50% were kept to constitute the original datasets for 65 development-related genes (denoted as DGs).

Random genome datasets generation

Because of the large number of genomes in the basic genome dataset, the current orthology inference method cannot obtain sufficient strict orthogroups [theoretically refer to a gene family including genes that are orthologous to each other and do not involve pairs of inparalogs (Boeckmann et al., 2011); technically

2 https://www.uniprot.org

TABLE 1 The detailed information of development-related proteins in the genus *Streptomyces*.

Function category	Name	Length (aa)	Entry (UniProt)	Description	Organism	References
A-factor	AdpA	405	Q9S166	AdpA, A-factor-responsive transcriptional activator	*S. griseus*	Ohnishi et al., 1999
	ArpA	276	Q9ZN78	A-factor receptor protein	*S. griseus*	Onaka et al., 1995
	AfsA	301	P18394	2-oxo-3-(phosphooxy)propyl 3-oxoalkanoate synthase	*S. griseus*	Horinouchi et al., 1989
	BprA	282	B1VN94	(4-alkanoyl-5-oxo-2,5-dihydrofuran-3-yl)methyl phosphate reductase	*S. griseus*	Ohnishi et al., 2008
Aerial hypha formation	SigJ	198	Q9K2W6	ECF subfamily RNA polymerase sigma factor	*S. coelicolor*	Mazurakova et al., 2006
	SigN	278	Q9ADM4	Putative RNA polymerase sigma factor	*S. coelicolor*	Dalton et al., 2007
	BldA	402	Q93JF9	BldA-regulated nucleotide binding protein	*S. coelicolor*	Bentley et al., 2002
	BldB	98	Q7AKF6	Regulator, BldB	*S. coelicolor*	Bentley et al., 2002; Eccleston et al., 2006
	BldC	68	Q9RKJ5	Developmental transcriptional regulator BldC (MerR family)	*S. coelicolor*	Hunt et al., 2005
	BldD	167	Q7AKQ8	DNA-binding protein	*S. coelicolor*	Elliot et al., 1998
	BldG	113	Q9WVX8	Anti-sigma factor antagonist (Anti-sigma-B factor antagonist)/Anti-sigma-B factor antagonist	*S. coelicolor*	Bignell et al., 2003
	BldKA	343	Q93IU3	ABC transporter integral membrane protein BldKA	*S. coelicolor*	Bentley et al., 2002
	BldKB	600	Q93IU2	ABC transporter lipoprotein BldKB	*S. coelicolor*	Chávez et al., 2011
	BldKC	325	Q93IU1	ABC transporter integral membrane protein BldKC	*S. coelicolor*	Bentley et al., 2002
	BldKD	353	Q93IU0	ABC transporter intracellular ATPase subUT BldKD	*S. coelicolor*	Bentley et al., 2002
	BldKE	381	Q8CJS2	Peptide transport system ATP-binding subUT	*S. coelicolor*	Bentley et al., 2002
	BldM	203	Q7AKI8	Putative two-component regulator (isolation response regulatory proteins)	*S. coelicolor*	Bentley et al., 2002
	BldN	177	Q9WX11	RNA polymerase sigma factor	*S. coelicolor*	Bentley et al., 2002
	ClpP1	219	Q9F315	ATP-dependent Clp protease proteolytic subUT 1	*S. coelicolor*	de Crecy-Lagard et al., 1999
	ClpX	428	Q9F316	ATP-dependent Clp protease ATP-binding subUT ClpX	*S. coelicolor*	Bentley et al., 2002
	SigH	361	Q9RIT0	RNA polymerase sigma factor	*S. coelicolor*	Potúèková et al., 1995
	Sti1	144	P61152	Subtilase-type protease inhibitor	*S. coelicolor*	Kim et al., 2005
Cell division	CrgA	84	Q9XA10	Cell division protein CrgA	*S. coelicolor*	Bentley et al., 2002; Del Sol et al., 2003
	DpsA	187	Q9R408	DNA-binding protein from starved cells	*S. coelicolor*	Facey et al., 2009
	DpsB	425	O86816	DNA-binding protein from starved cells	*S. coelicolor*	Facey et al., 2009
	DpsC	353	Q9K3L0	DNA-binding protein from starved cells	*S. coelicolor*	Facey et al., 2009
	FtsI	651	Q9Z5V7	FtsI	*S. coelicolor*	Bennett et al., 2009
	FtsK	917	O86810	DNA translocase FtsK	*S. coelicolor*	Bentley et al., 2002
	FtsQ	264	P45518	Cell division protein FtsQ	*S. coelicolor*	McCormick and Losick, 1996

(Continued)

10.3389/fmicb.2023.1102250

TABLE 1 (Continued)

Function category	Name	Length (aa)	Entry (UniProt)	Description	Organism	References
	FtsW	456	Q9ZBA6	Peptidoglycan glycosyltransferase	*S. coelicolor*	Mistry et al., 2008
	MreB	361	Q9L1G6	Cell shape-determining protein MreB	*S. coelicolor*	Heichlinger et al., 2011
	ParB1	381	Q9S6U1	Putative plasmid partitioning protein, ParB2	*S. coelicolor*	Kim et al., 2000; Bentley et al., 2002
	SffA	596	Q9RKX5	Putative FtsK/SpoIIIE family protein	*S.coelicolor*	Ausmees et al., 2007
	Smc	1,186	Q9ZBQ2	Chromosome associated protein	*S. coelicolor*	Bentley et al., 2002; Dedrick et al., 2009
	SmeA	64	Q9RKX4	Putative small membrane protein	*S.coelicolor*	Ausmees et al., 2007
	DivIVA	398	Q9S2X4	Cell division protein DivIVA	*S. coelicolor*	Xu et al., 2008
	FtsZ	399	P45500	Cell division protein FtsZ	*S. coelicolor*	McCormick et al., 1994
	ParA	310	Q8VWE4	Putative partitioning protein ParA	*S. coelicolor*	Bentley et al., 2002; Ditkowski et al., 2010
	ParJ	333	Q9ADA5	Filament polymerization regulator ParJ	*S. coelicolor*	Xu et al., 2020
Hydrophobin	ChpA	252	Q8CJY7	Chaplin-A	*S. coelicolor*	Bentley et al., 2002
	ChpB	237	Q9X7U2	Chaplin-B	*S. coelicolor*	Bentley et al., 2002
	ChpC	259	Q9AD93	Chaplin-C	*S. coelicolor*	Bentley et al., 2002
	ChpD	75	Q9L1J9	Chaplin-D	*S. coelicolor*	Bentley et al., 2002
	ChpE	82	Q9X9Z2	Chaplin-E	*S. coelicolor*	Bentley et al., 2002
	ChpF	88	Q9KYG7	Chaplin-F	*S. coelicolor*	Bentley et al., 2002
	ChpG	90	Q9KYH3	Chaplin-G	*S. coelicolor*	Bentley et al., 2002
	ChpH	77	Q9AD92	Chaplin-H	*S. coelicolor*	Bentley et al., 2002
	RamA	635	O88039	ABC transporter ATP-binding protein	*S. coelicolor*	Ma and Kendall, 1994; Bentley et al., 2002
	RamB	608	Q7AKE5	ABC transporter ATP-binding protein	*S. coelicolor*	Ma and Kendall, 1994; Bentley et al., 2002
	RamC	930	O88037	Probable SapB synthase	*S. coelicolor*	Bentley et al., 2002
	RamR	202	Q7AKE4	Two-component system response regulator/rapid aerial mycelium regulator	*S. coelicolor*	Ma and Kendall, 1994; Bentley et al., 2002
	RamS	42	O88038	Lanthionine-containing peptide SapB precursor RamS	*S. coelicolor*	Bentley et al., 2002
Sporulation	SigF	287	P37971	RNA polymerase sigma-F factor	*S. coelicolor*	Kim et al., 2008; Ditkowski et al., 2010
	SsfR	326	Q9L688	Putative transcriptional regulator	*S. griseus*	Jiang and Kendrick, 2000
	SsgA	136	Q9X9U2	Regulator (Sporulation/cell division regulator)	*S. coelicolor*	Bentley et al., 2002

(Continued)

10.3389/fmicb.2023.1102250

TABLE 1 (Continued)

Function category	Name	Length (aa)	Entry (UniProt)	Description	Organism	References
	SsgB	137	Q9L268	Sporulation-specific cell division protein SsgB	*S. coelicolor*	Xu et al., 2009
	SsgR	241	Q9X9U3	Transcriptional regulator	*S. coelicolor*	Bentley et al., 2002
	WhiA	328	Q9Z515	Probable cell division protein WhiA (hypothetical protein)	*S. coelicolor*	Aínsa et al., 2000
	WhiB	87	Q7AKN0	Transcriptional regulator WhiB	*S. coelicolor*	Davis and Chater, 1992
	WhiD	112	Q7AKI9	Transcriptional regulator WhiD	*S. coelicolor*	Soliveri et al., 2000
	WhiE	627	P42534	Putative polyketide hydroxylase	*S. coelicolor*	Komaki et al., 2015
	WhiG	280	P17211	RNA polymerase sigma factor WhiG	*S. coelicolor*	Chater et al., 1989
	WhiH	295	Q7AKF5	Sporulation transcription factor, WhiH	*S. coelicolor*	Genay et al., 2007
	WhiI	220	O69859	Two-component regulator	*S. coelicolor*	Aínsa et al., 1999; Bentley et al., 2002
	WhiJ	283	Q9F3F8	Transcriptional regulator WhiJ	*S. coelicolor*	Aínsa et al., 2010

refer to a gene family including genes that are not only homologous to each other but also single copy in the source genomes]. Therefore, we used a random resampling strategy similar to previous works (Canback et al., 2002; Zhi et al., 2014) to decrease the number of genomes and to cover as much species diversity presented in BGD as possible. Briefly, a reference species tree (denoted as RST) was first reconstructed based on a concatenated alignment of UGs, using IQ-tree (version 2.1.2) (Nguyen et al., 2015). Due to the asymmetry of RST in the number of species on the deep branches, it was re-rooted to enable the numbers of species contained in the two deepest branches to be close. And then, based on re-rooted RST, all species were divided into groups based on different clades (**Supplementary Table 3**). The corresponding genomes were randomly selected in terms of the species constitution of various clades in RST. Finally, random sampling was repeated to generate 100 random genome datasets (denoted as RGDs, **Supplementary Table 4**).

Phylogenetic tree inference

All RGDs were individually subjected to orthology inference using OrthoFinder (version 2.5.2) (Emms and Kelly, 2015, 2019). The orthogroups with precisely one gene from each genome were regarded as the strict orthogroups (sOGs) and were aligned separately by ClustalO (version 1.2.3) (Sievers et al., 2011). Then, all alignments of sOGs were concatenated, and the ambiguous sites were trimmed by trimAl (version 1.4) (Capella-Gutierrez et al., 2009) with the "automated" option. A species tree based on supermatrix (denoted as ST1) was inferred using IQ-tree, based on the best model with the following options: "–m TEST –B 1000 –T AUTO." To alleviate the bias induced by a single method, an additional species tree based on supertree (denoted as ST2) under the multispecies coalescent model was also built using ASTRAL (version 5.7.4) (Mirarab et al., 2014). The single-gene maximum likelihood (ML) trees used in the supertree approach were inferred using IQ-tree with options: "-m TEST –score-diff ALL –B 1000 –wbtl –T AUTO."

Besides ST1 and ST2, other two concatenated ML trees were reconstructed based on UGs and DGs, respectively. Because the universal marker gene had at most one sequence in each genome, UGs can be treated like sOGs, and their concatenated tree (denoted as UT) could be built using the same method. However, unlike UG, the development-related gene sequences belonging to one RGD were first extracted from the original DG (based on BGD) and then subjected to the orthology assessment by UPhO (Ballesteros and Hormiga, 2016). To a DG family, its subfamilies that formed the largest monophyletic subtrees in the tree of the whole family were recognized as strict orthogroups and used to build the concatenated ML tree (denoted as DT). In summary, for each RGD, two species trees (ST1 and ST2), two concatenated ML trees (UT and DT), and a batch of single-gene trees (genes in sOGs and DGs) were reconstructed as well.

Phylogenetic topology comparison

The Robinson-Foulds (RF) distance (Briand et al., 2020; Hayati and Chindelevitch, 2020) was employed to quantify the difference

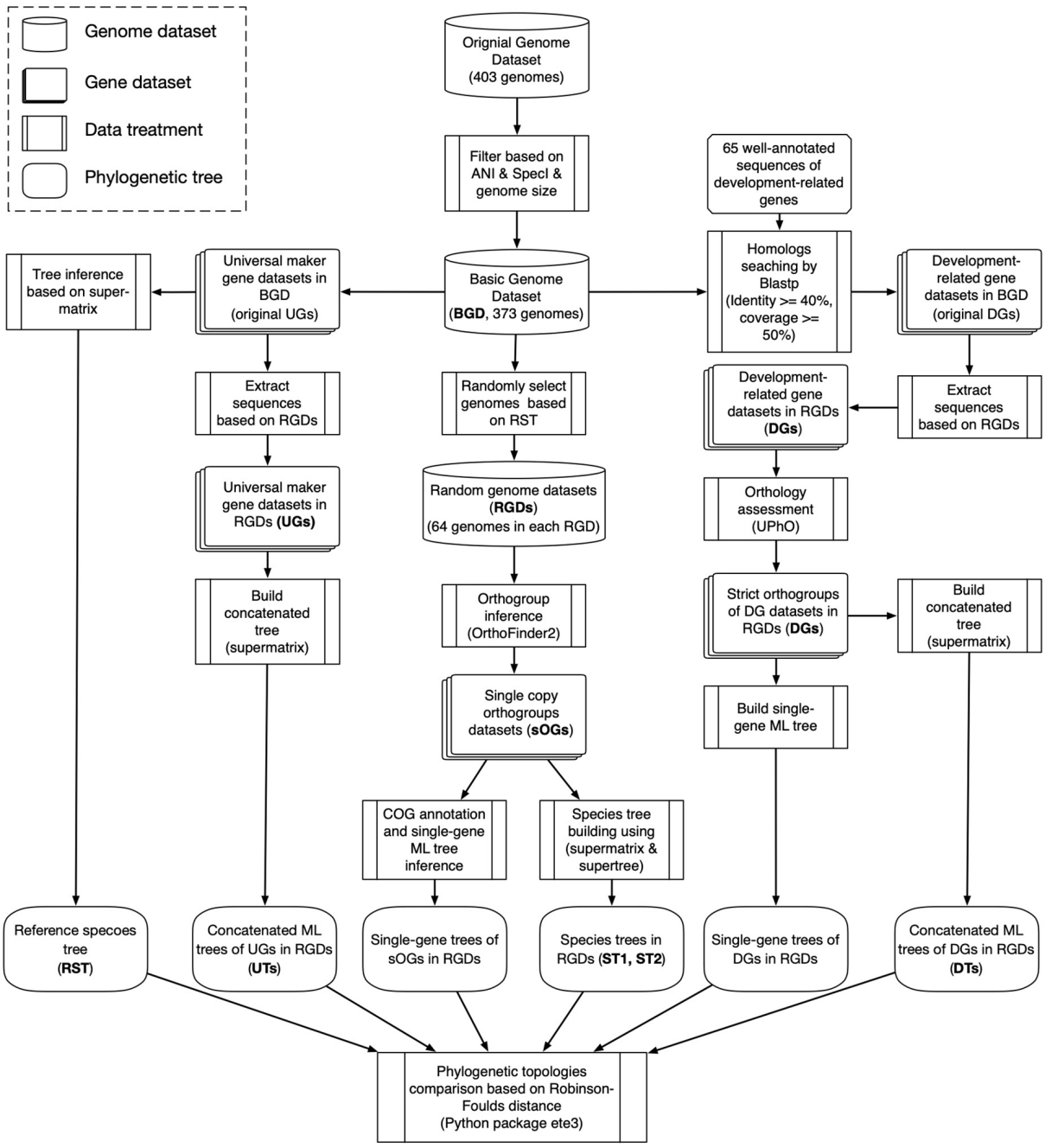


FIGURE 1
Flowchart of data analysis in this study.

in phylogenetic topologies by using the Python package ete3 (version 3.1.2). The species trees were compared with concatenated ML trees and single-gene trees. However, because multichotomous branching profoundly affects the RF distance calculation, the single-gene trees with multichotomous branches involving ≥50% of species were excluded from the topological comparison. To examine the difference in phylogenetic topologies among various functional groups, the Clusters of Orthologous Genes (COG) annotation was performed on all sOGs using the online version of eggNOG (version 5.0) (Huerta-Cepas et al., 2019). The single-gene trees of sOGs could be divided into subgroups according to the COG categories (Galperin et al., 2021), and their RF distances to species trees (ST1 and ST2) were analyzed separately. The whole data analysis procedure in this work is illustrated schematically in **Figure 1**.

The underpinning evolutionary processes causing phylogenetic conflict are diverse (Schrempf and Szöllősi, 2020). Besides biological factors like gene duplication and loss, and horizontal gene transfer, a wide variety of analytical factors, e.g., the tree reconstruction based on sequence alone (Szöllősi et al., 2013) and the limited amount of phylogenetic information (Steel, 2005), can lead to overestimating the phylogenetic conflict. Here, to reduce the

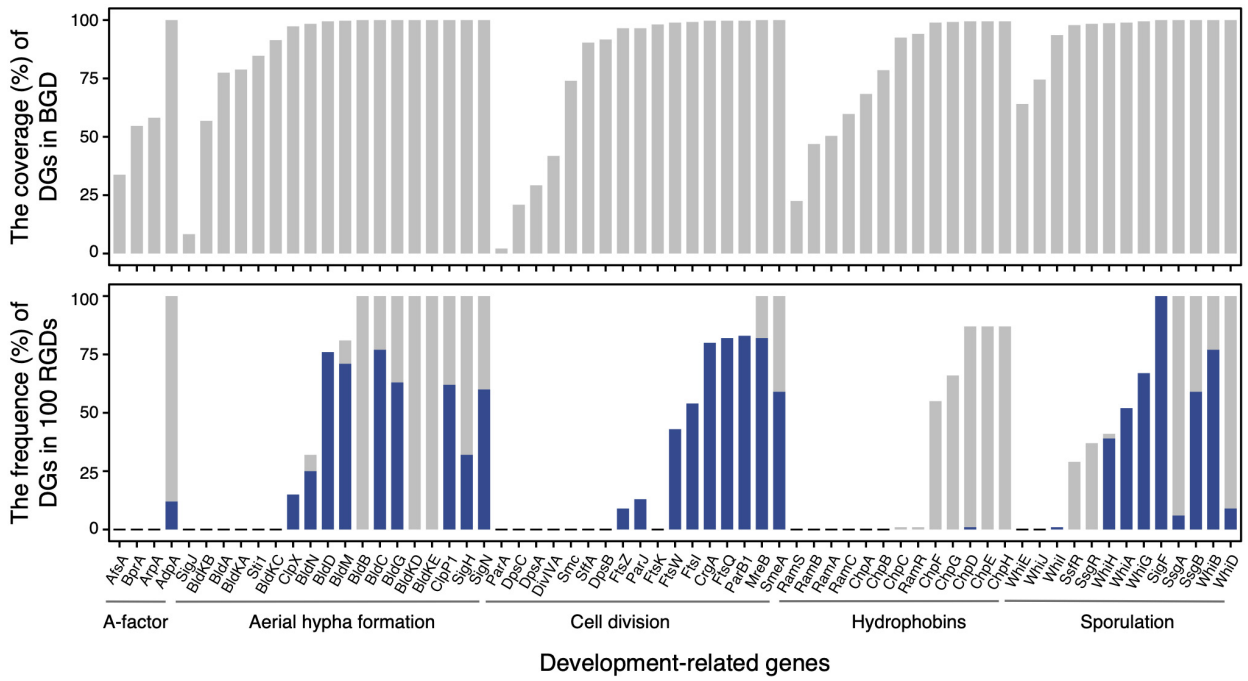


FIGURE 2
The distributions of development-related genes in the basic genome dataset **(upper panel)** and 100 random genome datasets **(lower panel)**. In the lower panel, the frequencies of genome dataset (DG) as core gene family (having homolog in each genome) and strict orthogroup (subfamily assessed by UPhO) are visualized as gray and blue bars, respectively.

discordance between gene trees inferred based on sequence alone, single-gene trees were reconciled using amalgamated likelihood estimation (ALE version 0.4) (Szöllősi et al., 2013, 2015). Briefly, the ultrafast bootstrap tree distributions generated by IQ-tree were converted into ALE objects using ALEobserve. And then, each ALE object was reconciled using ALEml_undated with the species tree.

Additionally, to reduce the gene tree incongruence due to differences in the number of parsimony-informative sites (PIS), the RF distance between the species tree and single-gene trees were normalized using a linear model between the PIS numbers and the actual RF distances. All RF distance values were sorted by corresponding PIS numbers ascendingly and divided into equal-sized bins (each bin contained 1,000 RF values). In each bin, the top 5% of RF values (the 5th percentile of RF values) representing "good" single-gene trees were selected for regression analysis with the relevant PIS numbers. A linear model was used to fit a line ($RF = a\log_{10}N_{PIS}+b$, N_{PIS}: the number of PIS) using least squares fitting. Then, all original RF values were normalized using the formula $RF' = \frac{RF}{a\log_{10}N_{PIS}+b}$.

Results

The morphological development predated the species divergence of *Streptomyces*

Based on 40 universal markers, a reference species tree (RST) of 373 *Streptomyces* genomes (basic genome datasets, BGD) was reconstructed. According to the evolutionary relationships reflected in RST, 373 *Streptomyces* species were segregated into eight major clades labeled as C1–C8 (**Supplementary Figure 1**). The species numbers of these clades were 1 (C1), 2 (C2), 18 (C3), 159 (C4), 36 (C5), 56 (C6), 15 (C7), and 86 (C8), respectively (**Supplementary Table 3**). Namely, the two deepest internal branches of the re-rooted RST possessed four descendant clades: C1–C4 (180 species) and C5–C8 (193 species). Parts of *Streptomyces* genomes were randomly selected according to the proportion of species number in eight clades, and their genomes constituted a small random genome dataset (RGD) containing 64 species (**Supplementary Table 4**). It's worth noting that the number of species in each major clade is varied; hence the probabilities of 373 species being selected are also different. The highest proportion of overlapped species between RGDs was less than 37.2%. And each species was re-sampled an average of 17.3 times (standard deviation, sd: 6.2).

In terms of the distribution of homologous sequences, as expected, 44 of 65 development-related genes were identified over 90% of genomes in BGD (**Figure 2**, upper panel). Those genes ought to play crucial roles in the development of *Streptomyces*. For instance, AdpA, an essential early regulator of the development (Horinouchi, 2007), and MreB, involved in the spore cell wall assembly (Heichlinger et al., 2011) could be identified in all *Streptomyces* genomes. Nevertheless, nearly a third of the development-related genes showed varying degrees of loss. Most of them were either unnecessary for morphological development or could be replaced by alternative genes. For example, the narrow phylogenetic distribution of the SapB biosynthesis gene cluster, including the *ramCSAB*

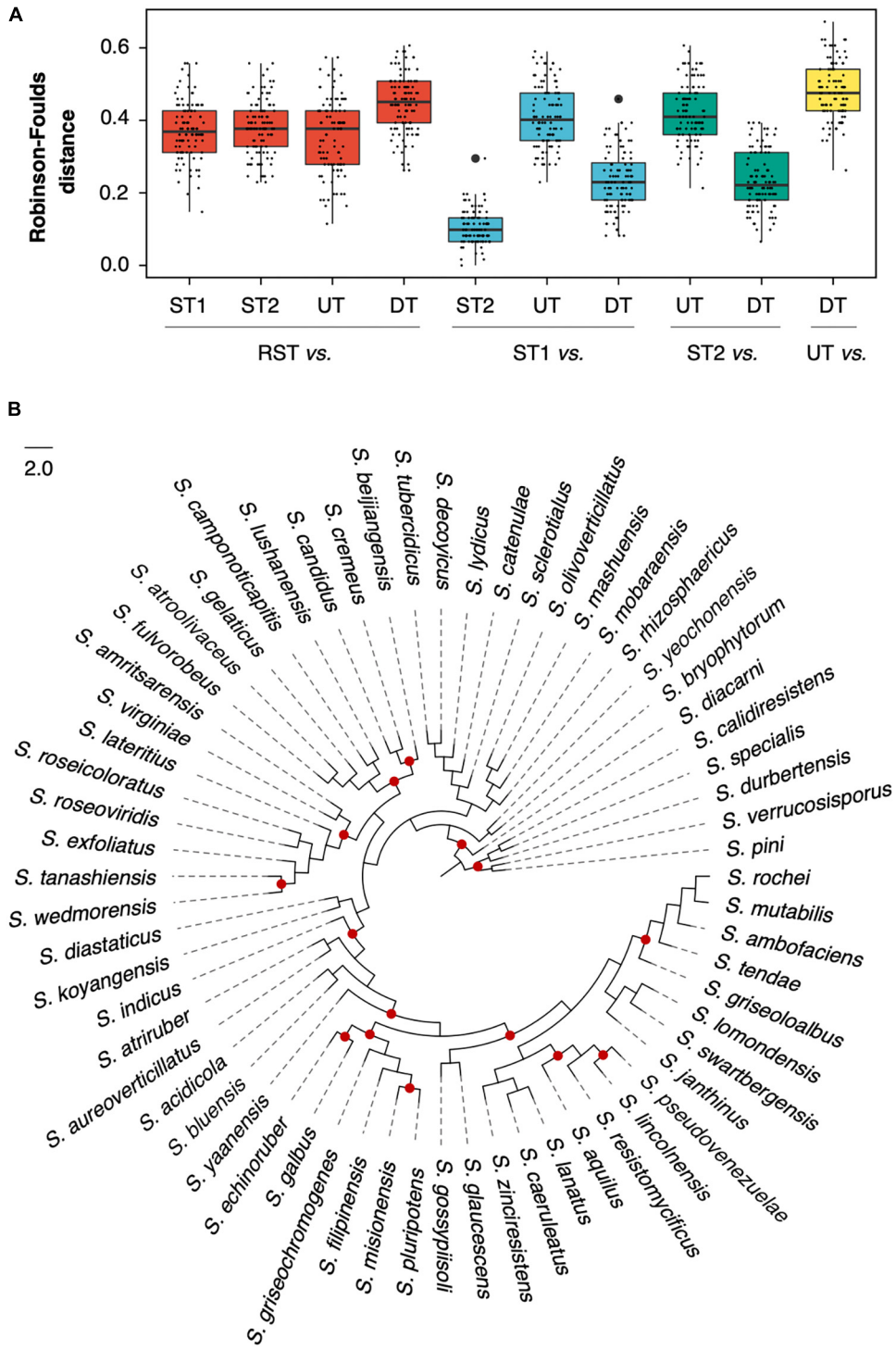


FIGURE 3
Topological comparison among species trees and concatenated maximum likelihood (ML) trees. **(A)** Robinson-Foulds distance between two compared trees in one of 100 random genome datasets (RGDs); **(B)** the species tree of RGD008 based on the supermatrix approach. The red dots represent where the phylogenetic conflict between DT and ST1 occurred.

operon and *ramR* (a response regulator gene that controls the activation of the convergently transcribed *ramCSAB* operon), might relate to a SapB-independent pathway that is mediated by the chaplins (Nguyen et al., 2002; O'Connor et al., 2002). These results are consistent with the previous report (Chandra and Chater, 2014; Kim et al., 2015). However, the distribution of homologous sequences cannot reflect evolutionary events such as horizontal gene transfer and gene duplication. In other words, a gene of late origin can exhibit a wider phylogenetic distribution through HGT. Therefore, we evaluated the orthologies of these DGs and further investigated the distribution of orthologous DGs.

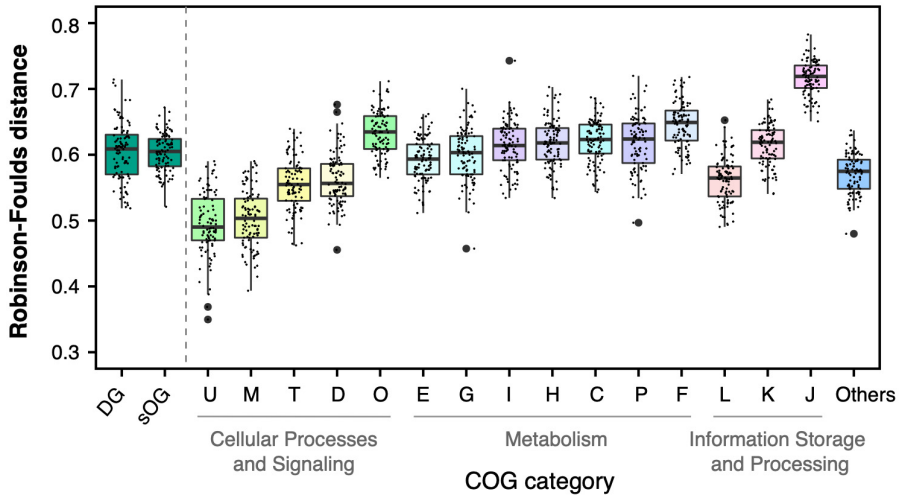


FIGURE 4
Topological comparison between single-gene trees and species tree based on supermatrix (ST1). Each single data point represented a mean value of Robison-Foulds distances between single-gene trees in the corresponding dataset [e.g., genome dataset (DG)] to ST1 in one of 100 random genome datasets (RGDs).

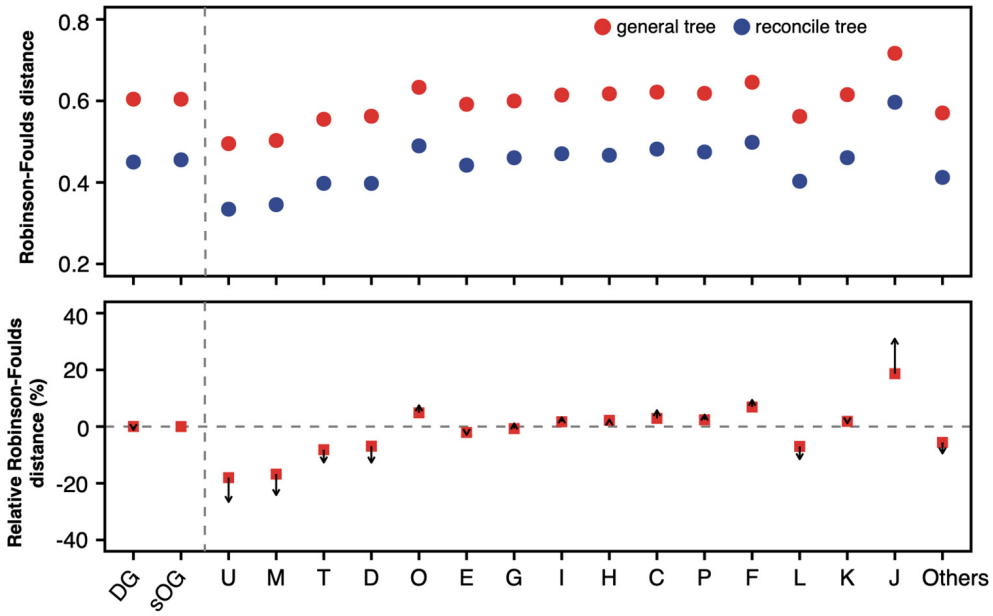


FIGURE 5
Robinson-Foulds distance of general gene tree and reconciled gene tree to supermatrix (ST1). For a gene dataset, 100 mean values of Robinson-Foulds (RF) distances from 100 random genome datasets (RGDs) were averaged and shown in the **(upper panel)**. Red solid circles represent the average RF distance of general gene trees to ST1, and solid blue circles represent the average RF distance of reconciled gene trees to ST1. The relative RF distance was calculated as formula, $\frac{\overline{RF}_{t_J \to ST1} - \overline{RF}_{t_{sOG} \to ST1}}{\overline{RF}_{sOG \to ST1}}$, where t_J is general gene tree of the gene in Clusters of Orthologous Genes (COG) J. In the **(lower panel)**, arrows start from the relative RF distance of general gene trees and point to the relative RF distance of reconciled gene trees.

However, because it is difficult to obtain a reliable and stable phylogenetic result for the large sequence dataset, evaluating the orthology of DG directly based on BGD is not feasible. Therefore, the assessment of orthology was carried out based on RGDs in this work. With the reduction of the BGD to an RGD comprising 64 genomes, the development-related protein sequences only associated with focal RGD were extracted from the original DGs. In focal RGD, the development-related protein family distributed in all 64 genomes (number of sequences ≥64) was regarded as a core protein family and subjected to the assessment of orthology. The subfamilies of the core protein family corresponding to strict orthogroups were finally retained. Due to the high similarity between some query sequences (e.g., the similarities among SigH, SigF, and SigN were more than 70%), the identical strict orthogroup

might be extracted from different DGs. In this case, only the strict orthogroup shared the highest similarity to the corresponding reference was extracted for further investigation. In addition, five protein families (ClpP1, SigN, SigF, WhiD, and SmeA) could be decomposed into two subfamilies as strict orthogroups and involved in 33 RGDs. Particularly, this situation occurs mainly in two protein families: ClpP1 (19 RGDs) and SigN (10 RGDs). These 33 RGDs included one and only one development-related gene that has this situation except for RGD064 and RGD090. All these strict orthogroups were retained for subsequent analysis.

The distributions of orthologous DGs in BGD were exhibited through the frequencies of DGs as core protein families (gray bars) and as strict orthogroups (blue bars) in 100 RGDs (**Figure 2**, lower panel). A total of 29 DGs involved in aerial hyphae and mature spore formation, cell division, and hydrophobin synthesis were recognized as core protein families in more than 50 % of RGDs. And 17 of 29 could extract strict orthogroups. These protein families from which strict orthogroups can be extracted tend to be more conserved than others. Notably, three protein families (BldB, BldKE, and BldKD) failed to extract the strict orthogroups, due to the high identity between sequences (resulting in a multifurcating tree). In addition, even though the strict orthogroups could be retrieved, the trimmed alignments were so short and conservative that the comparable trees of BldC, BldM, SsgB, WhiB, CrgA, and SmeA were unavailable. Besides the genome incompleteness and the species specificity of certain proteins, the progressive acquisition of these genes in the early ancestors of *Streptomyces* might also account for the differences in protein family distribution. Based on the larger genome datasets and relatively comprehensive protein collection, these results demonstrated that development-related genes had already been equipped in the genome of the common ancestor of *Streptomyces* and underwent a vertical evolutionary history, confirming that the origin of development should occur before the flourishing of its species diversity.

The concatenated DGs phylogeny has a higher degree of congruence with the species tree

In each RGD, two species trees based on supermatrix (ST1) and supertree (ST2) approaches and two concatenated ML trees, DT (based on DGs) and UT (based on UGs), were reconstructed. And a batch of single-gene trees was also inferred for strict orthogroups (sOGs) (496 60.8 trees) and DGs (10 2.6 trees). The Robinson-Foulds (RF) distance was implemented to evaluate the topological differences. Noticeably, when RST was compared to ST1, ST2, DT, and UT; it was required to be pruned to fit the different comparison objects. As shown in **Figure 3A**, the RF distances from UT to RST (mean sd: 0.36 0.1) were not significantly different from the distances from ST1 to RST (0.37 0.08) (Student's *t*-test, $p = 0.167$). Compared to RST, ST1 differed in both the underlying gene dataset and the number of species involved; but UT only differed in the number of species involved. Therefore, it indicated that the random resampling strategy could achieve the goal of maintaining topological stability when reducing the number of species involved. Nevertheless, in addition to the underlying gene dataset and the number of species involved, the species tree-building approach of ST2 was also different. That might cause a significant difference between the distances from ST2 to RST and that from UT to RST (*t*-test, $p = 0.048$). The RF distances from DT to RST were significantly higher than that from ST1, ST2, and UT to RST (*t*-test, all *p*-values $< 10^{-5}$).

When ST1 was used as a reference, the RF distances from DT to ST1 decreased dramatically (from 0.45 0.08 to 0.23 0.08) and were even significantly lower than that from UT to ST1 (*t*-test, $p < 10^{-5}$). For example, **Figure 3B** illustrates the degree of congruence between DT and ST1, which were generated from dataset RGD008 (RF = 0.23). Given the negligible difference between ST1 and ST2 (their average RF distance was only 0.10 0.05), the distances of DT decreased similarly when ST2 was used as a reference. For further understanding of the distances from DT to ST1, sOGs were divided according to COG functional categories and then used to infer concatenated ML trees for each COG category. The average RF distances from COG categories to ST1 ranged from 0.11 (COG Others, including orthogroups that could not be assigned to COGs or whose COG category appeared once in 100 RGDs) to 0.33 (COG P) (**Supplementary Table 5**). Interestingly, DT had a lower RF distance to ST1 than half of the COG categories. Concerning concatenation analysis, the phylogenies of concatenated DGs had a higher degree of congruence with the specie trees of the genus *Streptomyces*.

The single-gene phylogenies of DGs have no significant difference from that of sOGs on phylogenetic conflict

The concatenation approach integrating phylogenetic information from multiple gene loci suppresses the phylogenetic heterogeneity between different genes (Pease et al., 2016; Wu et al., 2018; Jiang et al., 2020). To investigate the phylogenetic conflicts of single genes with the species tree, all single-gene trees in different gene datasets were individually compared with the species tree (**Figure 4** and **Supplementary Figure 2**, detailed data see **Supplementary Table 5**). For each gene dataset, the mean value of RF distances from single-gene trees to the species tree was calculated and used to measure the topological difference between the whole gene dataset and the species tree. The distribution of mean values of RF distances from sOGs to ST1 demonstrated that the single-gene tree had a higher degree of incongruence to the species tree than the concatenated tree, as expected.

The mean RF distances from DGs to ST1 (0.60 0.04) were not significantly different from the mean RF distances from sOGs to ST1 (0.60 0.03) (*t*-test, $p = 0.9$). And it is consistent with the results of concatenation analysis, DGs had lower RF distance to ST1 than parts of COG categories (e.g., COGs F and J). sOGs belonging to different COG categories showed different degrees of topological incongruence with species trees. For instance, COG M (cell wall/membrane/envelope biogenesis) and COG U (intracellular trafficking, secretion, and vesicular transport) had relatively lower mean RF distances to the species tree. And, at the higher level of functional classification, the single-gene trees of COGs related to cellular processes and signaling were closer to the species tree than those of other COGs involved in metabolism

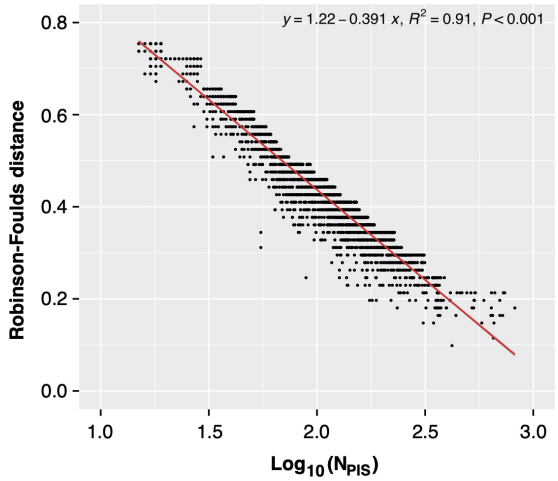


FIGURE 6
Regression analysis between Robinson-Foulds (RF) distance [between the single-gene tree and supermatrix (ST1)] and the number of parsimony-informative sites.

and information storage and processing. Astonishingly, COG J (translation, ribosomal structure, and biogenesis), including all ribosomal proteins usually used in phylogenomic analysis exhibited the highest degree of incongruence with the species tree. However, the concatenated tree of COG J had only 0.26 RF distance on average to ST1 (**Supplementary Table 5**). This suggested that the concatenation would exert a complex effect on species tree reconstruction.

The phylogenetic inference of a single gene based on sequence alone often lacks enough information to confidently support one gene tree topology. This may be one of the major reasons why the single gene has a higher degree of topological incongruence than the concatenated alignment. By reconciling the gene tree with a putative species tree based on the joint likelihood, its incongruence with the species tree was substantially reduced (**Figure 5** and **Supplementary Figure 3**). However, the mean RF distances of reconciled gene trees are still higher than those of the corresponding concatenation trees. Interestingly, although the RF distances were decreased after the gene trees were reconciled, the magnitudes of the RF distance reduction of various gene datasets relative to sOGs were discrepant. Generally, gene datasets whose single-gene trees were closer to the species tree, such as COGs M and U, had more magnitudes of RF distance reduction. In contrast, gene datasets whose single-gene trees were more distant from the species tree, such as COG J, had fewer magnitudes of RF distance reduction.

DGs exhibited the highest topological concordance with the species tree based on the normalized RF distance

As we know, the number of information sites used for phylogenetics profoundly impacted the topology of the resultant tree. Therefore, the limited number of parsimony-informative sites might contribute to the topological incongruence between the single gene tree and the species tree. Meanwhile, we noted that the numbers of parsimony-informative sites of proteins seem to be negatively correlated to their RF distances (Pearson's r = -0.51, $p < 0.01$, **Supplementary Table 5**). Therefore, to determine the correlation between the number of parsimony-informative sites and the topological incongruence, all single genes in different gene datasets were divided into bins according to the number of parsimony-informative sites. The RF distance values in each bin were sorted ascendingly, and the top 5% of RF values representing ones with "good" single-gene phylogeny (similar to species tree) were used for the regression analysis (**Figure 6** and **Supplementary Figure 4**).

The actual RF distances were normalized based on the linear regression model (**Figure 7** and **Supplementary Figure 5**). However, the normalized RF distance scale changed, which made the direct comparison to the original RF distance inapplicable. So, just as reconciled gene tree comparison, relative RF distance was used to present the variation of RF distance caused by the normalization based on the number of parsimony-informative sites. The relative mean RF distance from single-gene trees of COG J to ST1 decreased from 18.6 to 1.2%; on the contrary, the relative mean RF distance from single-gene trees of COG L to ST1 increased from –6.9 to 6.0%. Generally, the regression model decreased the RF distances of genes with more parsimony-informative sites (such as COGs U, M, E, G, C, and L). In contrast, the RF distances of genes with a lower number of parsimony-informative sites were increased (such as COGs J, K, and DGs). Remarkably, the RF distances of DGs had dramatically changed relative to sOGs same as COG J (decreased from <0.1 to –13.2%), and became the lowest ones. Taken together, these results indicated that development-related genes exhibited more congruent single-gene phylogenies to the species tree versus strict orthogroups, especially when technical factors like the limitation of parsimony-informative sites are excluded.

Discussion

The development of *Streptomyces*, including both morphological and physiological differentiation, was controlled by the same complex regulatory system and endowed streptomycetes with the ability to cope with complex living environments. Based on a large genome dataset, we found that development-related genes present differential distribution patterns in *Streptomyces* genomes. Parts of development-related genes had been equipped in the genome of the common ancestor of *Streptomyces* bacteria. Although it remains unknown whether these are sufficient to support the morphological development of the common ancestor of *Streptomyces*, the origin of these genes can be traced back to the early stages of *Streptomyces* speciation or even earlier. The phylogenetic distribution of homologous genes is an essential tool for studying the origin of genes and relevant biological functions, relying on the assumption that ancient genes often have a broader phylogenetic distribution. However, the natural evolutionary process is not entirely immune to confounding factors, such as horizontal gene transfer, which inevitably complicate the evolution of the gene family. In this study, the orthologies of development-related genes were also considered besides their phylogenetic distribution.

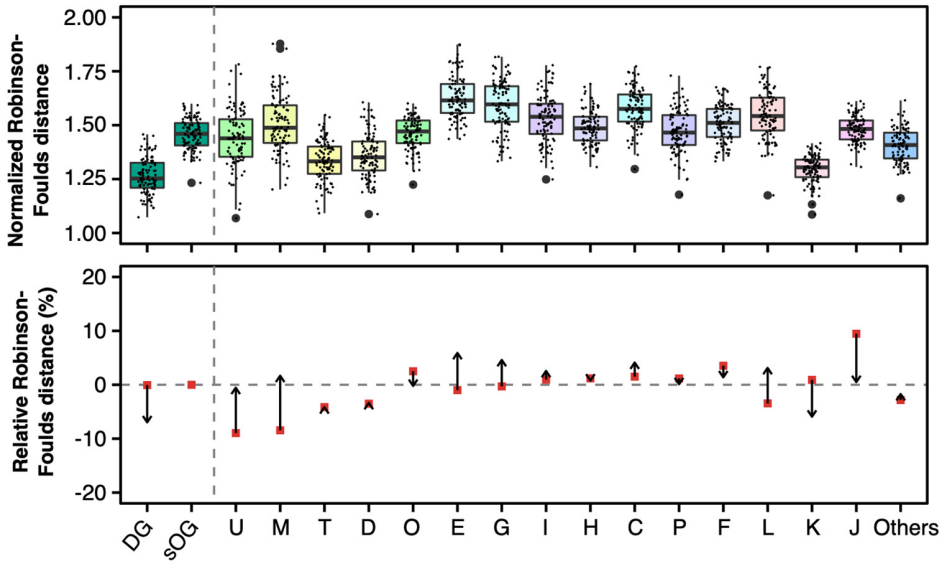


FIGURE 7
Normalized Robison-Foulds distances **(upper panel)** and relative normalized Robison-Foulds distances (**lower panel**) between single-gene trees and supermatrix (ST1). Like the lower panel of **Figure 4**, the arrows start from the relative Robison-Foulds (RF) distances of general gene trees and point to the relative RF distances after normalization based on the number of parsimony-informative sites.

Furthermore, we compared the phylogenetic topologies of related genes to the species tree on both concatenation and single-gene levels. We found that the evolution of development-related genes had higher congruence to the species tree. Especially when the analytical factors, such as tree reconstruction based on sequence alone and the limited amount of phylogenetic information, were excluded, the development-related genes exhibited the lowest degree of topological incongruence with the species tree. This is consistent with the above conclusion that parts of development-related genes have been assembled in the genome of the comment ancestor of *Streptomyces*. Consequently, the subsequent evolution of these genes should be synchronized with the history of species differentiation. Or in other words, the development-related genes documented the evolutionary history of *Streptomyces*. Meanwhile, it must be noted that the phylogenies of some development-related genes originating earlier in *Streptomyces* were relatively complex. These genes were either not conservative enough, or their genuine orthologies were obscured by other evolutionary events like lateral gene transfer, gene duplication, and loss.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: doi: 10.5061/dryad.00000005d.

Author contributions

MW and C-JL: initial drafting, data analysis, figures generation, and discussion. ZZ and P-PL: raw data processing and discussion. L-LY: discussion and draft correction. X-YZ: project coordination, discussion, and draft correction. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the National Natural Science Foundation of China (Grant Number: 32060003); Program for Excellent Young Talents, Yunnan University; and grants from the Major Science and Technology Projects of Yunnan Province (Digitalization, development and application of biotic resource, 202002AA100007).

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.

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.2023.1102250/full#supplementary-material

References

Aínsa, J. A., Bird, N., Ryding, N. J., Findlay, K. C., and Chater, K. F. (2010). The complex whiJ locus mediates environmentally sensitive repression of development of *Streptomyces coelicolor* A3(2). *Antonie Van Leeuwenhoek* 98, 225–236. doi: 10.1007/s10482-010-9443-3

Ainsa, J. A., Parry, H. D., and Chater, K. F. (1999). A response regulator-like protein that functions at an intermediate stage of sporulation in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* 34, 607–619. doi: 10.1046/j.1365-2958.1999.01630.x

Aínsa, J. A., Ryding, N. J., Hartley, N., Findlay, K. C., Bruton, C. J., and Chater, K. F. (2000). WhiA, a protein of unknown function conserved among gram-positive bacteria, is essential for sporulation in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* 182, 5470–5478. doi: 10.1128/JB.182.19.5470-5478.2000

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Andam, C. P., Doroghazi, J. R., Campbell, A. N., Kelly, P. J., Choudoir, M. J., and Buckley, D. H. (2016). A latitudinal diversity gradient in terrestrial bacteria of the genus *Streptomyces*. *mBio* 7, e2200–e2215. doi: 10.1128/mBio.02200-15

Anderson, A. S., and Wellington, E. M. (2001). The taxonomy of Streptomyces and related genera. *Int J. Syst. Evol. Microbiol.* 51, 797–814. doi: 10.1099/00207713-51-3-797

Ausmees, N., Wahlstedt, H., Bagchi, S., Elliot, M. A., Buttner, M. J., and Flärdh, K. (2007). SmeA, a small membrane protein with multiple functions in *Streptomyces* sporulation including targeting of a SpoIIIE/FtsK-like protein to cell division septa. *Mol. Microbiol.* 65, 1458–1473. doi: 10.1111/j.1365-2958.2007.05877.x

Ballesteros, J. A., and Hormiga, G. (2016). A New orthology assessment method for phylogenomic data: Unrooted phylogenetic orthology. *Mol. Biol. Evol.* 33, 2117–2134. doi: 10.1093/molbev/msw069

Bennett, J. A., Yarnall, J., Cadwallader, A. B., Kuennen, R., Bidey, P., Stadelmaier, B., et al. (2009). Medium-dependent phenotypes of *Streptomyces coelicolor* with mutations in ftsI or ftsW. *J. Bacteriol.* 191, 661–664. doi: 10.1128/JB.01048-08

Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A.-M., Challis, G. L., Thomson, N. R., James, K. D., et al. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417, 141–147. doi: 10.1038/417141a

Bibb, M. J. (2005). Regulation of secondary metabolism in streptomycetes. *Curr. Opin. Microbiol.* 8, 208–215. doi: 10.1016/j.mib.2005.02.016

Bignell, D. R. D., Lau, L. H., Colvin, K. R., and Leskiw, B. K. (2003). The putative anti-anti-sigma factor BldG is post-translationally modified by phosphorylation in *Streptomyces coelicolor*. *FEMS Microbiol. Lett.* 225, 93–99. doi: 10.1016/S0378-1097(03)00504-4

Boeckmann, B., Robinson-Rechavi, M., Xenarios, I., and Dessimoz, C. (2011). Conceptual framework and pilot study to benchmark phylogenomic databases based on reference gene trees. *Brief Bioinform.* 12, 423–435. doi: 10.1093/bib/bbr034

Briand, S., Dessimoz, C., El-Mabrouk, N., Lafond, M., and Lobinska, G. (2020). A generalized robinson-foulds distance for labeled trees. *BMC Genomics* 21:779. doi: 10.1186/s12864-020-07011-0

Canback, B., Andersson, S. G. E., and Kurland, C. G. (2002). The global phylogeny of glycolytic enzymes. *Proc. Natl. Acad. Sci. U.S.A.* 99, 6097–6102. doi: 10.1073/pnas.082112499

Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Capstick, D. S., Willey, J. M., Buttner, M. J., and Elliot, M. A. (2007). SapB and the chaplins: Connections between morphogenetic proteins in *Streptomyces coelicolor*. *Mol. Microbiol.* 64, 602–613. doi: 10.1111/j.1365-2958.2007.05674.x

Chandra, G., and Chater, K. F. (2014). Developmental biology of *Streptomyces* from the perspective of 100 actinobacterial genome sequences. *FEMS Microbiol. Rev.* 38, 345–379. doi: 10.1111/1574-6976.12047

Chater, K. F. (2006). Streptomyces inside-out: A new perspective on the bacteria that provide us with antibiotics. *Phil. Trans. R. Soc. B* 361, 761–768. doi: 10.1098/rstb.2005.1758

Chater, K. F., and Chandra, G. (2006). The evolution of development in *Streptomyces* analysed by genome comparisons. *FEMS Microbiol. Rev.* 30, 651–672. doi: 10.1111/j.1574-6976.2006.00033.x

Chater, K. F., Bruton, C. J., Plaskitt, K. A., Buttner, M. J., Méndez, C., and Helmann, J. D. (1989). The developmental fate of S. *coelicolor* hyphae depends upon a gene product homologous with the motility σ factor of B. *subtilis*. *Cell* 59, 133–143. doi: 10.1016/0092-8674(89)90876-3

Chávez, A., Forero, A., Sánchez, M., Rodríguez-Sanoja, R., Mendoza-Hernández, G., Servín-Gonzalez, L., et al. (2011). Interaction of SCO2127 with BldKB and its possible connection to carbon catabolite regulation of morphological differentiation in *Streptomyces coelicolor*. *Appl. Microbiol. Biotechnol.* 89, 799–806. doi: 10.1007/s00253-010-2905-8

Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D. R., da Costa, M. S., et al. (2018). Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 68, 461–466. doi: 10.1099/ijsem.0.002516

Creamer, K. E., Kudo, Y., Moore, B. S., and Jensen, P. R. (2021). Phylogenetic analysis of the salinipostin γ-butyrolactone gene cluster uncovers new potential for bacterial signalling-molecule diversity. *Microbial. Genomics* 7:000568. doi: 10.1099/mgen.0.000568

Dalton, K. A., Thibessard, A., Hunter, J. I. B., and Kelemen, G. H. (2007). A novel compartment, the ?subapical stem? of the aerial hyphae, is the location of a sigN-dependent, developmentally distinct transcription in *Streptomyces coelicolor*. *Mol. Microbiol.* 64, 719–737. doi: 10.1111/j.1365-2958.2007.05684.x

Davis, N. K., and Chater, K. F. (1992). The Streptomyces coelicolor whiB gene encodes a small transcription factor-like protein dispensable for growth but essential for sporulation. *Mol. Gen. Genet.* 232, 351–358. doi: 10.1007/BF00266237

de Crecy-Lagard, V., Servant-Moisson, P., Viala, J., Grandvalet, C., and Mazodier, P. (1999). Alteration of the synthesis of the Clp ATP-dependent protease affects morphological and physiological differentiation in *Streptomyces*. *Mol. Microbiol.* 32, 505–517. doi: 10.1046/j.1365-2958.1999.01364.x

Dedrick, R. M., Wildschutte, H., and McCormick, J. R. (2009). Genetic interactions of smc, ftsK, and parB genes in *Streptomyces coelicolor* and their developmental genome segregation phenotypes. *J. Bacteriol.* 191, 320–332. doi: 10.1128/JB.00858-08

Del Sol, R., Pitman, A., Herron, P., and Dyson, P. (2003). The product of a developmental gene, crgA, that coordinates reproductive growth in *Streptomyces* belongs to a novel family of small actinomycete-specific proteins. *J. Bacteriol.* 185, 6678–6685. doi: 10.1128/JB.185.22.6678-6685.2003

Ditkowski, B., Troæ, P., Ginda, K., Donczew, M., Chater, K. F., Zakrzewska-Czerwińska, J., et al. (2010). The actinobacterial signature protein ParJ (SCO1662) regulates ParA polymerization and affects chromosome segregation and cell division during *Streptomyces* sporulation: ParJ – regulator of ParA polymerization. *Mol. Microbiol.* 78, 1403–1415. doi: 10.1111/j.1365-2958.2010.07409.x

Eccleston, M., Willems, A., Beveridge, A., and Nodwell, J. R. (2006). Critical residues and novel effects of overexpression of the *Streptomyces coelicolor* developmental protein BldB: Evidence for a critical interacting partner. *J. Bacteriol.* 188, 8189–8195. doi: 10.1128/JB.01119-06

Elliot, M., Damji, F., Passantino, R., Chater, K., and Leskiw, B. (1998). The bldD gene of *Streptomyces coelicolor* A3(2): A regulatory gene involved in morphogenesis and antibiotic production. *J. Bacteriol.* 180, 1549–1555. doi: 10.1128/JB.180.6.1549-1555.1998

Emms, D. M., and Kelly, S. (2015). OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:157. doi: 10.1186/s13059-015-0721-2

Emms, D. M., and Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238. doi: 10.1186/s13059-019-1832-y

Facey, P. D., Hitchings, M. D., Saavedra-Garcia, P., Fernandez-Martinez, L., Dyson, P. J., and Del Sol, R. (2009). Streptomyces coelicolor Dps-like proteins: Differential dual roles in response to stress during vegetative growth and in nucleoid condensation during reproductive cell division. *Mol. Microbiol.* 73, 1186–1202. doi: 10.1111/j.1365-2958.2009.06848.x

Flärdh, K., and Buttner, M. J. (2009). *Streptomyces* morphogenetics: Dissecting differentiation in a filamentous bacterium. *Nat. Rev. Microbiol.* 7, 36–49. doi: 10.1038/nrmicro1968

Galperin, M. Y., Wolf, Y. I., Makarova, K. S., Vera Alvarez, R., Landsman, D., and Koonin, E. V. (2021). COG database update: Focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* 49, D274–D281. doi: 10.1093/nar/gkaa1018

Genay, M., Decaris, B., and Dary, A. (2007). Implication of stringent response in the increase of mutability of the whiG and whiH genes during *Streptomyces coelicolor* development. *Mutat. Res.* 624, 49–60. doi: 10.1016/j.mrfmmm.2007.03.016

Hammerschmidt, K., Landan, G., Domingues Kümmel Tria, F., Alcorta, J., and Dagan, T. (2021). The order of trait emergence in the evolution of cyanobacterial multicellularity. *Genome Biol. Evol.* 13:evaa249. doi: 10.1093/gbe/evaa249

Hayati, M., and Chindelevitch, L. (2020). Computing the distribution of the Robinson-Foulds distance. *Comput. Biol. Chem.* 87:107284. doi: 10.1016/j.compbiolchem.2020.107284

Heichlinger, A., Ammelburg, M., Kleinschnitz, E.-M., Latus, A., Maldener, I., Flärdh, K., et al. (2011). The MreB-Like Protein Mbl of *Streptomyces coelicolor* A3(2) depends on MreB for proper localization and contributes to spore wall synthesis. *J. Bacteriol.* 193, 1533–1542. doi: 10.1128/JB.01100-10

Horinouchi, S. (2007). Mining and polishing of the treasure trove in the bacterial genus streptomyces. *Biosci. Biotechnol. Biochem.* 71, 283–299. doi: 10.1271/bbb.60627

Horinouchi, S., Suzuki, H., Nishiyama, M., and Beppu, T. (1989). Nucleotide sequence and transcriptional analysis of the *Streptomyces* griseus gene (afsA) responsible for A-factor biosynthesis. *J. Bacteriol.* 171, 1206–1210. doi: 10.1128/jb.171.2.1206-1210.1989

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085

Hunt, A. C., Servín-González, L., Kelemen, G. H., and Buttner, M. J. (2005). The bldC developmental locus of *Streptomyces coelicolor* encodes a member of a family of small DNA-binding proteins related to the DNA-binding domains of the MerR family. *J. Bacteriol.* 187, 716–728. doi: 10.1128/JB.187.2.716-728.2005

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114. doi: 10.1038/s41467-018-07641-9

Jiang, H., and Kendrick, K. E. (2000). Characterization of ssfR and ssgA, Two genes involved in sporulation of *Streptomyces griseus*. *J. Bacteriol.* 182, 5521–5529. doi: 10.1128/JB.182.19.5521-5529.2000

Jiang, X., Edwards, S. V., and Liu, L. (2020). the multispecies coalescent model outperforms concatenation across diverse phylogenomic data sets. *Syst. Biol.* 69, 795–812. doi: 10.1093/sysbio/syaa008

Joynt, R., and Seipke, R. F. (2018). A phylogenetic and evolutionary analysis of antimycin biosynthesis. *Microbiology* 164, 28–39. doi: 10.1099/mic.0.000572

Kämpfer, P. (2015). "Streptomyces," in *Bergey's manual of systematics of archaea and bacteria*, eds W. Whitman, F. Rainey, P. Kämpfer, M. Trujillo, J. Chun, P. DeVos, et al. (New York, NY: John Wiley & Sons). doi: 10.1002/9781118960608.gbm00191

Kim, D.-W., Chater, K., Lee, K.-J., and Hesketh, A. (2005). Changes in the extracellular proteome caused by the absence of the bldA gene product, a developmentally significant tRNA, reveal a new target for the pleiotropic regulator AdpA in *Streptomyces coelicolor*. *J. Bacteriol.* 187, 2957–2966. doi: 10.1128/JB.187.9.2957-2966.2005

Kim, E. S., Song, J. Y., Kim, D. W., Chater, K. F., and Lee, K. J. (2008). A possible extended family of regulators of sigma factor activity in *Streptomyces coelicolor*. *J. Bacteriol.* 190, 7559–7566. doi: 10.1128/JB.00470-08

Kim, H.-J., Calcutt, M. J., Schmidt, F. J., and Chater, K. F. (2000). Partitioning of the linear chromosome during sporulation of *Streptomyces coelicolor* A3(2) involves an oriC -Linked parAB locus. *J. Bacteriol.* 182, 1313–1320. doi: 10.1128/JB.182.5.1313-1320.2000

Kim, J.-N., Kim, Y., Jeong, Y., Roe, J.-H., Kim, B.-G., and Cho, B.-K. (2015). Comparative genomics reveals the core and accessory genomes of *Streptomyces* Species. *J. Microbiol. Biotechnol.* 25, 1599–1605. doi: 10.4014/jmb.1504.04008

Komaki, H., Ichikawa, N., Hosoyama, A., Fujita, N., and Igarashi, Y. (2015). Draft genome sequence of marine-derived *Streptomyces* sp. TP-A0598, a producer of anti-MRSA antibiotic lydicamycins. *Stand. Genomic Sci.* 10:58. doi: 10.1186/s40793-015-0046-5

Kormanec, J., and Sevcikova, B. (2002). The stress-response sigma factor σH controls the expression of ssgB, a homologue of the sporulation-specific cell division gene ssgA, in *Streptomyces coelicolor* A3(2). *Mol. Gen. Genomics* 267, 536–543. doi: 10.1007/s00438-002-0687-0

Lee, N., Hwang, S., Kim, J., Cho, S., Palsson, B., and Cho, B.-K. (2020). Mini review: Genome mining approaches for the identification of secondary metabolite biosynthetic gene clusters in *Streptomyces*. *Comput. Struct. Biotechnol. J.* 18, 1548–1556. doi: 10.1016/j.csbj.2020.06.024

Li, C.-J., Wang, M., Zhang, Z., Zhi, X.-Y., and Yang, L.-L. (2022). Genome-based analyses reveal heterotypic synonyms of *Streptomyces* species and associated subspecies. *Arch. Microbiol.* 204:581. doi: 10.1007/s00203-022-03204-1

Li, Y., Pinto-Tomás, A. A., Rong, X., Cheng, K., Liu, M., and Huang, Y. (2019). Population genomics insights into adaptive evolution and ecological differentiation in *Streptomycetes*. *Appl. Environ. Microbiol.* 85, e2555–e2518. doi: 10.1128/AEM.02555-18

Liu, G., Chater, K. F., Chandra, G., Niu, G., and Tan, H. (2013). Molecular regulation of antibiotic biosynthesis in *Streptomyces*. *Microbiol. Mol. Biol. Rev.* 77, 112–143. doi: 10.1128/MMBR.00054-12

Ma, H., and Kendall, K. (1994). Cloning and analysis of a gene cluster from *Streptomyces coelicolor* that causes accelerated aerial mycelium formation in Streptomyces lividans. *J. Bacteriol.* 176, 3800–3811. doi: 10.1128/jb.176.12.3800-3811.1994

Mazurakova, V., Sevcikova, B., Rezuchova, B., and Kormanec, J. (2006). Cascade of sigma factors in streptomycetes: Identification of a new extracytoplasmic function sigma factor σJ that is under the control of the stress-response sigma factor σH in *Streptomyces coelicolor* A3(2). *Arch. Microbiol.* 186, 435–446. doi: 10.1007/s00203-006-0158-9

McCormick, J. R., and Losick, R. (1996). Cell division gene ftsQ is required for efficient sporulation but not growth and viability in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* 178, 5295–5301. doi: 10.1128/jb.178.17.5295-5301.1996

McCormick, J. R., Su, E. P., Driks, A., and Losick, R. (1994). Growth and viability of *Streptomyces coelicolor* mutant for the cell division gene ftsZ. *Mol. Microbiol.* 14, 243–254. doi: 10.1111/j.1365-2958.1994.tb01285.x

McDonald, B. R., and Currie, C. R. (2017). Lateral Gene transfer dynamics in the ancient bacterial genus *Streptomyces*. *mBio* 8, e644–e617. doi: 10.1128/mBio.00644-17

Mende, D. R., Sunagawa, S., Zeller, G., and Bork, P. (2013). Accurate and universal delineation of prokaryotic species. *Nat. Methods* 10, 881–884. doi: 10.1038/nmeth.2575

Mirarab, S., Reaz, R., Bayzid, M., Zimmermann, T., Swenson, M. S., and Warnow, T. (2014). ASTRAL: Genome-scale coalescent-based species tree estimation. *Bioinformatics* 30, i541–i548. doi: 10.1093/bioinformatics/btu462

Mistry, B. V., Del Sol, R., Wright, C., Findlay, K., and Dyson, P. (2008). FtsW is a dispensable cell division protein required for Z-ring stabilization during sporulation septation in *Streptomyces coelicolor*. *J. Bacteriol.* 190, 5555–5566. doi: 10.1128/JB.00398-08

Navarro-Muñoz, J. C., Selem-Mojica, N., Mullowney, M. W., Kautsar, S. A., Tryon, J. H., Parkinson, E. I., et al. (2020). A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* 16, 60–68. doi: 10.1038/s41589-019-0400-9

Nguyen, K. T., Willey, J. M., Nguyen, L. D., Nguyen, L. T., Viollier, P. H., and Thompson, C. J. (2002). A central regulator of morphological differentiation in the multicellular bacterium *Streptomyces coelicolor*. *Mol. Microbiol.* 46, 1223–1238. doi: 10.1046/j.1365-2958.2002.03255.x

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

O'Connor, T. J., Kanellis, P., and Nodwell, J. R. (2002). The ramC gene is required for morphogenesis in *Streptomyces coelicolor* and expressed in a cell type-specific manner under the direct control of RamR. *Mol. Microbiol.* 45, 45–57. doi: 10.1046/j.1365-2958.2002.03004.x

Ohnishi, Y., Ishikawa, J., Hara, H., Suzuki, H., Ikenoya, M., Ikeda, H., et al. (2008). Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J. Bacteriol.* 190, 4050–4060. doi: 10.1128/JB.00204-08

Ohnishi, Y., Kameyama, S., Onaka, H., and Horinouchi, S. (1999). The A-factor regulatory cascade leading to streptomycin biosynthesis in *Streptomyces griseus*?: Identification of a target gene of the A-factor receptor. *Mol. Microbiol.* 34, 102–111. doi: 10.1046/j.1365-2958.1999.01579.x

Ohnishi, Y., Yamazaki, H., Kato, J., Tomono, A., and Horinouchi, S. (2005). AdpA, a central transcriptional regulator in the A-factor regulatory cascade that leads to morphological development and secondary metabolism in *Streptomyces griseus*. *Biosci. Biotechnol. Biochem.* 69, 431–439. doi: 10.1271/bbb.69.431

Onaka, H., Ando, N., Nihira, T., Yamada, Y., Beppu, T., and Horinouchi, S. (1995). Cloning and characterization of the A-factor receptor gene from *Streptomyces griseus*. *J. Bacteriol.* 177, 6083–6092. doi: 10.1128/jb.177.21.6083-6092.1995

Parte, A. C., Sardà Carbasse, J., Meier-Kolthoff, J. P., Reimer, L. C., and Göker, M. (2020). List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. *Int. J. Syst. Evol. Microbiol.* 70, 5607–5612. doi: 10.1099/ijsem.0.004332

Pease, J. B., Haak, D. C., Hahn, M. W., and Moyle, L. C. (2016). Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biol.* 14:e1002379. doi: 10.1371/journal.pbio.1002379

Potúèková, L., Kelemen, G. H., Findlay, K. C., Lonetto, M. A., Buttner, M. J., and Kormanec, J. (1995). A new RNA polymerase sigma factor,? ^{F} is required for the late stages of morphological differentiation in *Streptomyces* spp. *Mol. Microbiol.* 17, 37–48. doi: 10.1111/j.1365-2958.1995.mmi_17010037.x

Rodríguez, H., Rico, S., Díaz, M., and Santamaría, R. I. (2013). Two-component systems in Streptomyces: Key regulators of antibiotic complex pathways. *Microb. Cell Fact.* 12:127. doi: 10.1186/1475-2859-12-127

Sanchez-Baracaldo, P., Hayes, P. K., and Blank, C. E. (2005). Morphological and habitat evolution in the Cyanobacteria using a compartmentalization approach. *Geobiology* 3, 145–165. doi: 10.1111/j.1472-4669.2005.00050.x

Schirrmeister, B. E., Antonelli, A., and Bagheri, H. C. (2011). The origin of multicellularity in cyanobacteria. *BMC Evol. Biol.* 11:45. doi: 10.1186/1471-2148-11-45

Schrempf, D., and Szöllősi, G. (2020). "The sources of phylogenetic conflicts," in *phylogenetics in the genomic era*, eds C. Scornavacca, F. Delsuc, and N. Galtier, 3.1:1–3.1:23. Available online at: https://hal.inria.fr/PGE

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75

Soliveri, J. A., Gomez, J., Bishai, W. R., and Chater, K. F. (2000). Multiple paralogous genes related to the Streptomyces coelicolor developmental regulatory gene whiB are present in Streptomyces and other actinomycetes. *Microbiology* 146, 333–343. doi: 10.1099/00221287-146-2-333

Steel, M. (2005). Should phylogenetic models be trying to 'fit an elephant'? *Trends Genet.* 21, 307–309. doi: 10.1016/j.tig.2005.04.001

Sun, D., Liu, C., Zhu, J., and Liu, W. (2017). Connecting metabolic pathways: Sigma factors in *Streptomyces* spp. *Front. Microbiol.* 8:2546. doi: 10.3389/fmicb.2017.02546

Szöllősi, G. J., Davín, A. A., Tannier, E., Daubin, V., and Boussau, B. (2015). Genome-scale phylogenetic analysis finds extensive gene transfer among fungi. *Phil. Trans. R. Soc. B* 370:20140335. doi: 10.1098/rstb.2014.0335

Szöllősi, G. J., Rosikiewicz, W., Boussau, B., Tannier, E., and Daubin, V. (2013). Efficient exploration of the space of reconciled gene trees. *Syst. Biol.* 62, 901–912. doi: 10.1093/sysbio/syt054

van Dissel, D., Claessen, D., and van Wezel, G. P. (2014). Morphogenesis of *Streptomyces* in submerged cultures. *Adv. Appl. Microbiol.* 89, 1–45. doi: 10.1016/B978-0-12-800259-9.00001-9

Waglechner, N., McArthur, A. G., and Wright, G. D. (2019). Phylogenetic reconciliation reveals the natural history of glycopeptide antibiotic biosynthesis and resistance. *Nat. Microbiol.* 4, 1862–1871. doi: 10.1038/s41564-019-0531-5

Willey, J., Santamaria, R., Guijarro, J., Geistlich, M., and Losick, R. (1991). Extracellular complementation of a developmental mutation implicates a small sporulation protein in aerial mycelium formation by S. *coelicolor*. *Cell* 65, 641–650. doi: 10.1016/0092-8674(91)90096-H

Wu, M., Kostyun, J. L., Hahn, M. W., and Moyle, L. C. (2018). Dissecting the basis of novel trait evolution in a radiation with widespread phylogenetic discordance. *Mol. Ecol.* 27, 3301–3316. doi: 10.1111/mec.14780

Xu, H., Chater, K. F., Deng, Z., and Tao, M. (2008). A Cellulose synthase-like protein involved in hyphal tip growth and morphological differentiation in *Streptomyces*. *J. Bacteriol.* 190, 4971–4978. doi: 10.1128/JB.01849-07

Xu, L., Ye, K.-X., Dai, W.-H., Sun, C., Xu, L.-H., and Han, B.-N. (2019). Comparative genomic insights into secondary metabolism biosynthetic gene cluster distributions of marine *Streptomyces*. *Mar. Drugs* 17:498. doi: 10.3390/md17090498

Xu, M., Wang, W., Waglechner, N., Culp, E. J., Guitor, A. K., and Wright, G. D. (2020). GPAHex-A synthetic biology platform for Type IV–V glycopeptide antibiotic production and discovery. *Nat. Commun.* 11:5232. doi: 10.1038/s41467-020-19138-5

Xu, Q., Traag, B. A., Willemse, J., McMullan, D., Miller, M. D., Elsliger, M.-A., et al. (2009). Structural and functional characterizations of SsgB, a conserved activator of developmental cell division in morphologically complex actinomycetes. *J. Biol. Chem.* 284, 25268–25279. doi: 10.1074/jbc.M109.018564

Zhi, X.-Y., Yao, J.-C., Tang, S.-K., Huang, Y., Li, H.-W., and Li, W.-J. (2014). The futalosine pathway played an important role in menaquinone biosynthesis during early prokaryote evolution. *Genome Biol. Evol.* 6, 149–160. doi: 10.1093/gbe/evu007

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 14 April 2023
DOI 10.3389/fmicb.2023.1117017

OPEN ACCESS

EDITED BY
Feng Gao,
Tianjin University,
China

REVIEWED BY
Carlos Henrique Camargo,
Adolfo Lutz Institute,
Brazil
Ulisses Padua Pereira,
State University of Londrina,
Brazil
Norma Velazquez-Guadarrama,
Federico Gómez Children's Hospital,
Mexico

*CORRESPONDENCE
Lars Jelsbak
✉ lj@bio.dtu.dk

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 06 December 2022
ACCEPTED 27 March 2023
PUBLISHED 14 April 2023

CITATION
Zhao Y, Chen D, Ji B, Zhang X, Anbo M and Jelsbak L (2023) Whole-genome sequencing reveals high-risk clones of *Pseudomonas aeruginosa* in Guangdong, China. *Front. Microbiol.* 14:1117017. doi: 10.3389/fmicb.2023.1117017

Whole-genome sequencing reveals high-risk clones of *Pseudomonas aeruginosa* in Guangdong, China

Yonggang Zhao[1†], Dingqiang Chen[2†], Boyang Ji[3], Xingju Zhang[4], Mikkel Anbo[1] and Lars Jelsbak[1*]

[1]Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens Lyngby, Denmark, [2]Department of Laboratory Medicine, Microbiome Medicine Center, Zhujiang Hospital, Southern Medical University, Guangzhou, Guangdong, China, [3]Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden, [4]BGI-Shenzhen, Shenzhen, China

The ever-increasing prevalence of infections produced by multidrug-resistant or extensively drug-resistant *Pseudomonas aeruginosa* is commonly linked to a limited number of aptly-named epidemical 'high-risk clones' that are widespread among and within hospitals worldwide. The emergence of new potential high-risk clone strains in hospitals highlights the need to better and further understand the underlying genetic mechanisms for their emergence and success. *P. aeruginosa* related high-risk clones have been sporadically found in China, their genome sequences have rarely been described. Therefore, the large-scale sequencing of multidrug-resistance high-risk clone strains will help us to understand the emergence and transmission of antibiotic resistances in *P. aeruginosa* high-risk clones. In this study, 212 *P. aeruginosa* strains were isolated from 2 tertiary hospitals within 3years (2018–2020) in Guangdong Province, China. Whole-genome sequencing, multi-locus sequence typing (MLST) and antimicrobial susceptibility testing were applied to analyze the genomic epidemiology of *P. aeruginosa* in this region. We found that up to 130 (61.32%) of the isolates were shown to be multidrug resistant, and 196 (92.45%) isolates were Carbapenem-Resistant *Pseudomonas aeruginosa*. MLST analysis demonstrated high diversity of sequence types, and 18 reported international high-risk clones were identified. Furthermore, we discovered the co-presence of *exoU* and *exoS* genes in 5 collected strains. This study enhances insight into the regional research of molecular epidemiology and antimicrobial resistance of *P. aeruginosa* in China. The high diversity of clone types and regional genome characteristics can serve as a theoretical reference for public health policies and help guide measures for the prevention and control of *P. aeruginosa* resistance.

KEYWORDS

***Pseudomonas aeruginosa*, high-risk clone, multidrug resistance, genomic epidemiology, carbapenem resistant**

1. Introduction

Pseudomonas aeruginosa (*P. aeruginosa*) is an opportunistic pathogen and one of the major causes of nosocomial infections, particularly affecting intensive care units (ICU) and immunocompromised patients, with a very high morbidity and mortality rate (Botelho et al., 2019; Thornton and Parkins, 2023). In addition, *P. aeruginosa* is the most frequent chronic

respiratory infection contributing to the pathogenesis of cystic fibrosis (*CF*) and other chronic underlying diseases (Kawalek et al., 2020). The innate and acquired antibiotics resistance mechanism of *P. aeruginosa* includes reduced permeability, antibiotic efflux, expression change, antibiotic modification/degradation, target protection and target modification (including the acquisition of insensitive functional target) (Pang et al., 2019).

Infections caused by multidrug-resistant (MDR) or extensively drug-resistant (XDR) *P. aeruginosa* are extremely difficult to treat (Kawalek et al., 2020). In Asia, the MDR and XDR rates of *P. aeruginosa* nosocomial pneumonia were reported to 42.8 and 4.9%, respectively (Pang et al., 2019). Over the last few decades, the occurrence of epidemic outbreaks caused by antibiotic resistance strains within hospital environments has been given more attention (Talebi Bezmin Abadi et al., 2019). Recent works have provided further evidence that a small group of MDR/XDR global clones are disseminated in many hospitals worldwide (Telling et al., 2018; Hu et al., 2021; Gravningen et al., 2022). The most prevalent of these globally disseminated clones include ST235, ST111, ST233, ST244, ST357, ST308, ST175, ST277, ST654 and ST298, and they have been denominated high-risk clones (Oliver et al., 2015; del Barrio-Tofiño et al., 2020). Furthermore, some types of *P. aeruginosa* strains have developed resistance to carbapenems, called Carbapenem-Resistant *P. aeruginosa* (CRPA). The production of carbapenemases, especially Ambler class B Metallo-β-lactamases (MBLs), is an important mechanism of carbapenem resistance in *P. aeruginosa* (Yoon and Jeong, 2021). Moreover, in the absence of Carbapenemase-mediated resistance, the loss of outer membrane porin *OprD*, together with the overexpression of efflux pumps or *ampC* have been shown to be the main mechanisms leading to carbapenem resistance (Botelho et al., 2019). In China, only a few reports are available on the molecular epidemiology and genome information of *P. aeruginosa* high-risk clones. At the same time, research on Carbapenem-Resistant *P. aeruginosa* (CRPA) together with high-risk clones is even more limited.

Guangdong is the most populous province with the highest domestic migration in China (Xu et al., 2022). This study aims to characterize clinical *P. aeruginosa* strains isolated from two hospitals in Guangdong, in terms of clone diversity, antibiotics resistance profiles, and virulence determinants. Furthermore, the relationship between the carbapenem genes and *P. aeruginosa* high-risk clones will be explored. The outcome of this study will fill the gap related to cultures of *P. aeruginosa* and their associated genome sequences in China. The genetic and phenotypic characteristics of *P. aeruginosa* high-risk clones would provide novel insights into the epidemic and genomic mechanism background of this pathogen in China.

2. Materials and methods

2.1. Study area

The samples were collected from two-level III (tertiary care) hospitals in Guangdong Province, China, where the patients are heterogeneous. From 2018 to 2020, 212 nonduplicate *P. aeruginosa* isolates were obtained at The First Affiliated Hospital of Guangzhou Medical University and Zhujiang Hospital of Southern Medical University.

2.2. Ethics approval

The Ethics Committee of the Zhujiang Hospital of Southern Medical University approved the research on May 30, 2021 (approval no.2021-KY-046-01).

2.3. Isolate and clinical data collection

Over 3 years (2018–2020), a total of 212 non-duplicate isolates were selected and distributed over different years, with 55 isolates in 2018, 104 isolates in 2019, and 53 isolates in 2020 (Supplementary Table S1). All isolates were stored in 10% glycerol-containing media and were frozen at −80°C until further use. These isolates originate from different clinical specimens processed by the hospital's laboratory as part of the microbiological diagnostic routine.

2.4. Antibiotic susceptibility testing

Antibiotic susceptibility testing was performed with the Minimum Inhibitory Concentration (MIC) method according to the guidelines of the Clinical and Laboratory Standards Institute (Humphries et al., 2021). The antibiotics tested included piperacillin-tazobactam, ceftazidime, cefepime, imipenem, meropenem, aztreonam, ciprofloxacin, levofloxacin, amikacin, tobramycin and colistin. The minimum inhibitory concentrations (MICs μg/mL) were obtained for each isolate, and the isolates were classified as resistant (R), intermediate (I) and susceptible (S) according to the obtained MIC accordingly. Isolates were designated as multidrug-resistant (MDR) if they were resistant to an antibiotic from ≥3 classes and extensively drug-resistant (XDR) if they were resistant to an antibiotic from ≥6 classes tested following standardized criteria (Magiorakos et al., 2012). This study, Carbapenem-Resistance *P. aeruginosa* (CRPA) was defined as an isolate with imipenem and/or meropenem resistance.

2.5. Bacterial identification, DNA extraction and WGS sequencing

Stored strains were regrown from the −80°C stocks using Nutrient agar. The isolates were grown for DNA extraction in Nutrient agar medium overnight at 37°C. DNA was extracted from pure cultures using the MagAttract HMW DNA Kit (Qiagen, Germany) according to the kit protocol. The stLFR technology uses Tn5 transposase for the co-barcoding of DNA libraries. The stLFR library was constructed following the standard protocol using the MGIEasy stLFR Library Prep kit v1.1 (PN: 1000005622) with some process improvement for better assembly of bacterial genomes according to published methods (Wang et al., 2019). The WGS was performed on MGISEQ-2000 platform (DNBSEQ™, MGI, China) according to the manufacturer's instructions in paired-end mode (PE 100 bp).

2.6. Genome assembly and scaffolding

The sequencing reads were evaluated by FastQC 0.11.3, and the reads were filtered by SOAPfilter (v2.2 version) software with the

parameters set as follows: -q 33 -y -p -M 2 -f−1 -Q 10. Athena (v1.3) was used for genome assembly (Bishara et al., 2018). The completeness and pollution of the assembly results were counted by CheckM (v1.0.13), and the SLR-superscaffolder (v0.9.0) was used for the scaffolding. After selection, an average of 5.5 Gb of sequences were kept, with an average of 6.7 million pairs of reads and mean coverage >100X according to the expected genome size (approx. 6.4 Mb). Detailed sequencing statistics and quality results are summarized (Supplementary Table S2). The quality criteria were met for all sequences.

2.7. Genotype profiling

The Comprehensive Antibiotic Resistance Database (CARD)[1] was used as a reference for the antimicrobial resistance genes analysis, and Virulence Factor Database (VFDB)[2] was used for the virulence factors screening (Alcock et al., 2020; Liu et al., 2022). The serotype was identified by PAst 1.0.[3] The sequence types of all strains were determined from their whole genome sequence data using the MLST scheme[4] sited at the University of Oxford (Jolley et al., 2018). The population snapshot of *P. aeruginosa* was inferred by using Phyloviz 2.0. based on the MLST allelic profiles. The minimum spanning tree (MST) analysis was implemented in BioNumerics 8.0 software (Applied-Maths, Sint Maartens-Latem, Belgium). To detect the mutations of the *OprD* gene, the gene sequence encoding the reference protein OprD in *Pseudomonas aeruginosa* strain PAO1 (accession number NC_002516) was obtained and aligned with each target strain by BLASTp script.

3. Results

3.1. MIC phenotyping

The MIC values together with the interpretations are shown (Table 1). 117 strains were MDR (55.19%), and 13 strains were XDR (6.13%) (Figure 1). Among these tested strains, a total of 196 strains belonging to CRPA (92.45%). The identified CRPA strains displayed reduced susceptibilities to most of the tested antimicrobials, but remained highly susceptible to colistin (98.11%) and amikacin (87.74%).

3.2. Clinical *Pseudomonas aeruginosa* isolates from patients in Guangdong, China

From 2018 to 2020, a total of 212 strains were isolated from patients in Guangdong, China. Among all the isolated strains, the most frequently clinical specimen was from sputum ($n = 141$, 66.51%), then hydrothorax ($n = 11$, 5.19%), blood ($n = 10$, 4.72%) and bronchoalveolar lavage fluid (BALF) ($n = 9$, 4.25%), as shown in Figure 1. According to the ward distribution of the studied strains, the number of samples from the intensive care unit (ICU) was 76 (35.85%), the Respiratory Department samples numbered 34 (16.04%), and the Neurology Department samples numbered 15 (7.08%) (Supplementary Table S1).

Among the 76 strains isolated from ICU, 55 multi-locus sequence types (STs) were identified. The most prevalent STs were ST244 ($n = 6$, 7.89%), ST274 ($n = 4$, 5.26%), and ST381 ($n = 4$, 5.26%). The 34 strains sampled from the ward at the Respiratory Department were classified into 31 STs; only ST253, ST611 and ST703 were with double strains separately. Besides the ICU samples, the intra-hospital transmission, for example, in The First Affiliated Hospital of Guangzhou Medical University, 4 ST569 MDR (CRPA) strains were isolated from the Department of Neurology, and 4 ST1971 strains were purified from the Department of Internal Medicine, which indicates that attention should also be paid to outside-ICUs-acquired-infections.

3.3. High sequence type diversity among the studied isolates

Among the isolates studied here, 122 previously reported sequence types and seven novel sequencing types were detected. Most STs are represented only by single isolates ($n = 93$, 43.87%). The high diversity of STs found among the isolates is also evident from the phylogenetic analysis of the strains, as shown in Supplementary Figure S1. There are no clear predominant sequence types among the collected strains, even for the identified multidrug-resistance strains. The most common STs isolated were ST244 ($n = 9$, 4.25%) and ST1971 ($n = 7$, 3.30%), followed by ST381 ($n = 6$, 2.83%) (Supplementary Table S1).

3.4. Subclade of the high-risk clones with multidrug resistance

Six global high-risk clones (ST111, ST235, ST244, ST277, ST298 and ST357) (Oliver et al., 2015) were identified in this study (Figure 2). In addition, 12 widely disseminated clones (previously reported as potential high-risk clones) were also identified, including ST108 (Telling et al., 2018), ST179 (Moloney et al., 2020), ST253 (Fischer et al., 2020), ST260 (Martak et al., 2022), ST274 (Sewunet et al., 2022), ST348 (Dix et al., 2022), ST395 (Petitjean et al., 2017), ST446 (Pincus et al., 2019), ST463 (Hu et al., 2021), ST532 (Founou et al., 2020), ST699 (Rada et al., 2021) and ST773 (Singh et al., 2021). We observed that several MDR clones identified in this study were related to high-risk clones and were thus part of the same subclade as them (Table 2). For example, ST446 ($n = 1$) was part of the sub-clade with high-risk clone ST298. Additional examples include ST1338 ($n = 4$, ST244), ST3674 ($n = 2$, ST244), ST209 ($n = 1$, ST274), and ST2326 ($n = 2$, ST207) (Table 2).

Besides the reported high-risk clones, we found eight groups/subgroups in the population where the founder STs were represented by multiple isolates with MDR/XDR phenotype (Green circles in Figure 2). These including ST381 ($n = 6$, 2.83%), ST207 ($n = 5$, 2.36%), ST313 ($n = 4$, 1.89%), ST508 ($n = 4$, 1.89%), ST569 ($n = 4$, 1.89%), and ST611 ($n = 4$, 1.89%). These successful emerging *P. aeruginosa* clones with MDR/XDR phenotype are summarized (Table 2).

1 https://card.mcmaster.ca/
2 http://www.mgc.ac.cn/VFs/
3 https://cge.cbs.dtu.dk/services/PAst/
4 https://pubmlst.org/paeruginosa/; last accessed July 11, 2022.

10.3389/fmicb.2023.1117017

TABLE 1 Resistance rate (percentages) and minimum inhibitory concentrations (MICs) of *P. aeruginosa* isolates in Guangdong (*n*=212).

Antimicrobial agents	*Pseudomonas aeruginosa* (*n*=212)				
	S	*I*	*R*	MIC50 (μg/mL)	MIC90 (μg/mL)
Ceftazidime	48.58%	9.91%	41.51%	2	64
Cefepime	53.30%	25.47%	21.23%	8	32
Aztreonam	21.23%	24.06%	54.72%	32	64
Imipenem	11.32%	1.42%	87.26%	16	16
Meropenem	9.91%	13.68%	76.42%	16	16
Piperacillin-tazobactam	29.72%	37.74%	32.55%	32	128
Ciprofloxacin	40.57%	17.45%	41.98%	1	4
levofloxacin	19.81%	9.43%	70.75%	4	8
Amikacin	87.74%	3.30%	8.96%	2	32
Tobramycin	79.72%	0.94%	19.34%	1	16
Colistin	98.11%	0.00%	1.89%	0.5	0.5

MIC_{50} minimal inhibitory concentration inhibiting 50% of isolates, MIC_{90} minimal inhibitory concentration inhibiting 90% of isolates, S susceptible, I intermediately susceptible, R resistant.

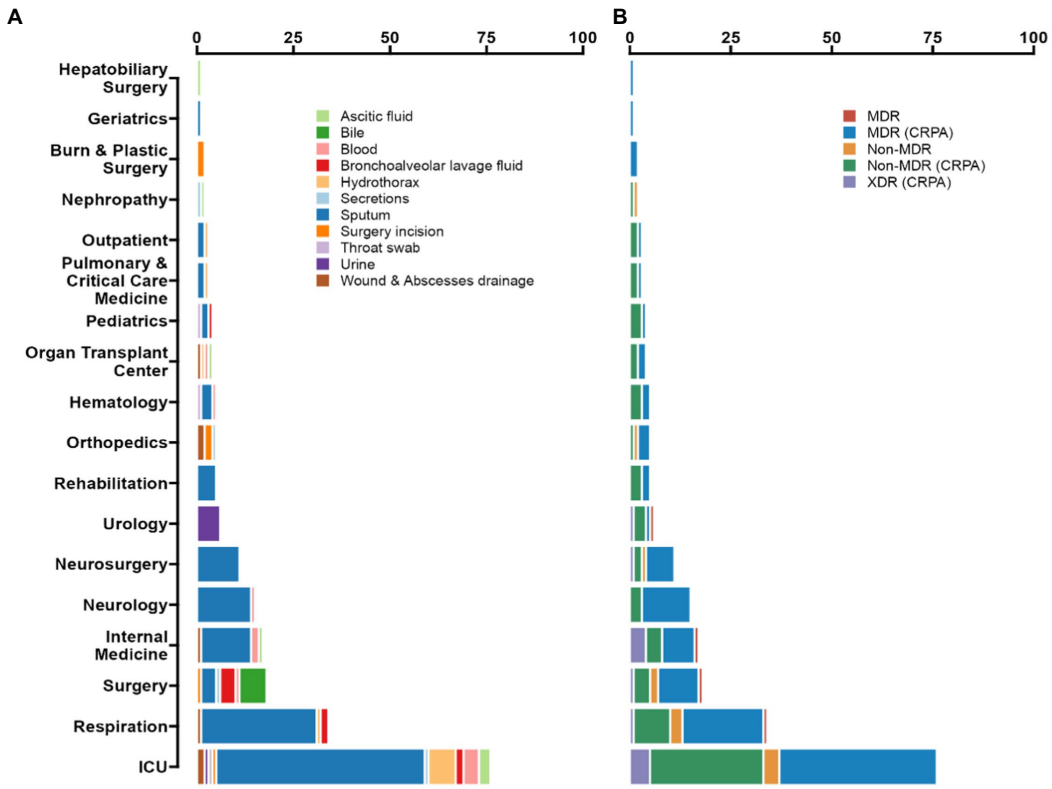


FIGURE 1
Prevalence of *P. aeruginosa* antimicrobial resistance in Guangdong province, China. **(A)** According to the ward distribution of the studied strains, the number of samples from ICU is 76 (35.85%), Respiratory Department samples numbered 34 (16.04%), and Neurology department samples numbered 15 (7.08%). Among all of the isolated strains, the most frequently isolated specimen was sputum (66.51%), then was hydrothorax (5.19%), blood (4.72%), and bronchoalveolar lavage fluid (BALF) (4.25%). **(B)** Up to 92.45% of the isolates were classified as CRPA, 55.19% as MDR (CRPA) and 6.13% as XDR (CRPA).

3.5. Antibiotics resistance genes profiling

In this study, 19 isolates harbored genes encoding aminoglycoside-modifying enzymes, including aminoglycoside acetyltransferase genes [*aac(6')-IIa*, *aac(6')-Ib4*, *aac(6')-Ib7*, *aac(6')-Ib9*, *aac(6')-Ib10*, and *aac(3)-IId*] and/or aminoglycoside adenylyltransferase genes (*aadA*, *aadA2*, *aadA3*, *aadA7* and *aadA13*). The identified bla_{OXA} type genes were composed of $bla_{OXA\text{-}50\text{-}Like}$ genes (*n*=212, 100%), $bla_{OXA\text{-}1\text{-}Like}$ genes (*n*=6, 2.83%), $bla_{OXA\text{-}10\text{-}Like}$ genes (*n*=4, 1.89%) and $bla_{OXA\text{-}21\text{-}Like}$ genes (*n*=1, 0.47%). In several cases, the strains were carrying multiple

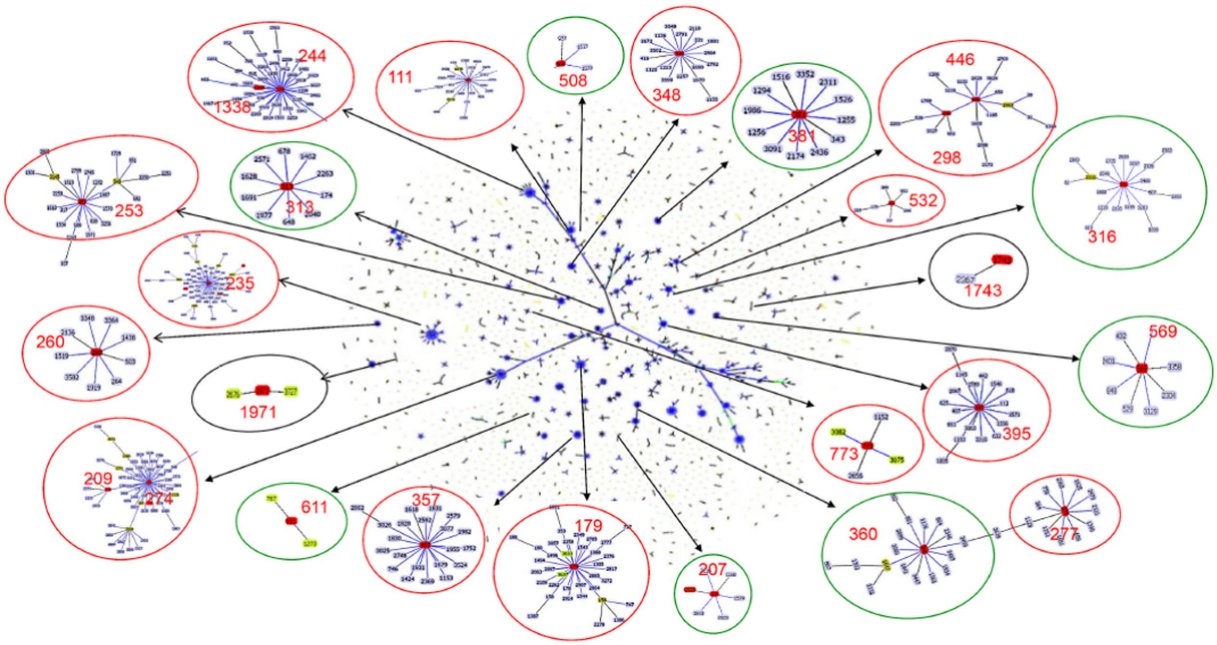


FIGURE 2
Population snapshot of *P. aeruginosa*. The 3,729 sequence types listed on the *P. aeruginosa* PubMLST database (http://pubmlst.org/paeruginosa, 2022/07/11) are displayed in a single eBURST diagram by setting the group definition to zero of seven shared alleles. Each dot represents a ST, and lines connect single-locus variants. The identified clones in this study were marked in red in each group of related STs. A red circle surrounded the groups reported as international high-risk clones, the groups with the new potential high-risk clones were surrounded by a green circle, and a black circle surrounded the local risk clones identified in this study.

β-lactamases. Among them, nine strains (4.25%) simultaneously carried two different bla_{OXA} type genes, and one ST1971 strain (0.47%) carried three different bla_{OXA} type genes. Furthermore, one bla_{VIM-2} gene and two different bla_{OXA} genes were found in each isolate of ST235 ($n = 2$) and ST244 ($n = 1$) strains. $bla_{IMP-9/45}$, bla_{KPC-2}, $bla_{CTX-M-13}$ and $bla_{CARB-1/3}$ co-existed with one/two bla_{OXA} type genes were also discovered in different strains (Table 3).

As with high prevalence rates of bla_{OXA} type genes, 211 isolates harbored *Pseudomonas*-derived cephalosporinase (PDC) β-lactamase genes, which includes 24 bla_{PDC} types. The dominant type was bla_{PDC-3} ($n = 51$), followed by bla_{PDC-5} ($n = 42$), which both belong to the extended-spectrum β-lactamase (ESBLs). We found bla_{PDC} types and STs had special correspondence relationships. Among them, 7 ST244 strains were carrying bla_{PDC-1} and bla_{PDC-5}, 5 ST357 strains were with bla_{PDC-11}, and 4 ST277 strains were with bla_{PDC-5}. Besides high-risk clones, 5 ST253 strains were with bla_{PDC-68}, 4 ST313 strains were with $bla_{PDC-105}$, 6 ST381 strains were with bla_{PDC-14}, 5 ST207 strains were with $bla_{PDC-337}$, and 7 ST1971 strains were with bla_{PDC-3}, and so on (Supplementary Table S3).

We did not find Class A and Class B antibiotic resistance genes (ARGs) in high-risk clone ST111, ST277 and ST298, but the *Klebsiella pneumoniae* carbapenemase (KPC) gene bla_{KPC-2} was found in one ST463 strain. The Verona Integron-encoded Metallo-β-lactamase (VIM) gene was detected in one ST244 strain, and two ST235 strains carried two bla_{VIM-2} genes in each of the isolates (Table 3). Furthermore, two types of *qnrVC* variants were detected in 6 CRPA strains (2.83%) with five different STs, which included five strains carrying *qnrVC1* and one strain carrying *qnrVC6*. The heat map showing unique resistance genes for identified *P. aeruginosa* high risk clone strains is illustrated (Figure 3). In order to further explore the reasons for the high proportion of CRPA in the collected strains, the *OprD* gene alignment was also performed. BLASTP analysis showed that the wild-type *OprD* gene was not detected in all collected strains (*oprD* mutation strains, $n = 212$).

The *exoU* and *exoS* genes, which encoding effector proteins of type III secretion system (T3SS) in *P. aeruginosa,* were investigated. The more frequent distribution of *exoS* ($n = 150$, 70.75%) was observed compared with *exoU* ($n = 57$, 26.89%). More interestingly, we found the co-presence of both genes in 5 strains (ST463, $n = 1$; ST273, $n = 1$; ST3658, $n = 1$; ST274 $n = 2$).

3.6. The antibiotics genes profiling and phenotype correlation

Several class A and class B β-lactamases were identified in 22 strains (10.37%), and all these were classified as CRPA and with multidrug resistance. 8 bla_{CARB-3} genes carried strains were identified, and 4 of these strains belonged to ST1338 with XDR phenotype. 5 strains harboring the $bla_{CTX-M-13}$ gene were identified and represented multidrug resistance. Isolates producing bla_{KPC} type enzymes often relate to the resistance to most of β-lactams and non-β-lactams antibiotics. 1 ST463 strain was carrying bla_{KPC-2} gene and represented XDR (CRPA) phenotype. 1 bla_{IMP-9} and 2 bla_{IMP-45} carried isolates were distributed into three dissimilar STs, including ST292, ST316 and ST381, which were isolated from the same hospital but in different wards. The MIC profiling showed that these three strains were all representing XDR (CRPA) phenotype. Three bla_{VIM-2} harboring

TABLE 2 The successful clones identified in this study with potential risks.

Clone types	Description of this study				Reported countries
	Strains No. and percentage	Serotype	MIC	Source	
ST381	6, 2.83%	O2	XDR(CRPA)	Sputum, Surgery incision	Brazil France Ivory Coast Malaysia Poland Russia Spain
ST207	5, 2.35%	O1	MDR/XDR(CRPA)	Sputum, Tissue	Australia Canada France Malaysia Poland Russia Spain
ST313	4, 1.89%	O1	MDR(CRPA)	Hydrothorax, Sputum, Blood	Armenia Australia France Hungary Poland Russia Spain
ST508	4, 1.89%	O3	MDR/XDR(CRPA)	Sputum, Surgery incision, Bile	Australia France The Netherlands
ST569	4, 1.89%	O3	MDR(CRPA)	Sputum	Australia The Netherlands Central African Republic
ST611	4, 1.89%	O5	MDR(CRPA)	Sputum, Blood, Bile	Australia Poland The Netherlands
ST316	2, 0.94%	O11	XDR(CRPA)	Urine, Sputum	Australia France Ivory Coast Senegal
ST360	2, 0.94%	O5	MDR(CRPA)	Sputum	Australia Senegal Ivory Coast Poland

high-risk clone strains were identified which were distributed in 2 ST235 and 1 ST244 strains. The responding phenotypes were multidrug resistance. Contrary to the high proportion of CRPA, class A and B carbapenemase genes identified in collected isolates are less frequent, which means their reduced susceptibility to carbapenems resulted from various non-enzymatic mechanisms (Figure 3).

4. Discussion

The emergence and increase of high-risk isolates, including their antibiotic resistance, has been recognized as a threat for human health (Oliver et al., 2015; del Barrio-Tofiño et al., 2020). However, our understanding of these clones' global distribution and evolution is still incomplete due to uneven sampling in different regions of the world. Here, we investigated clinical *P. aeruginosa* clones isolated in Guangdong, China. The MLST analysis showed high genetic diversity among collected isolates in agreement with prior studies in China (Fan et al., 2016). Importantly, some international high-risk clones as well as new widely distributed clones (regarded as potential risk clones), were identified in our study. The observations of potential high-risk and locally distributed clones underscore the relevance of continued inspection of the population structure in clinical settings in different parts of the world.

10.3389/fmicb.2023.1117017

TABLE 3 Horizontally-acquired β-lactamases identified in collected isolates.

β-lactamases			Clone (Strains)
Ambler class	Type	Enzyme	
Class A	CARB	CARB-1	ST1239(1), ST1338(1)
		CARB-3	ST357(1), ST1295(1), ST1338(3), NA(1)
	CTX-M	CTX-M-13	ST16(1), ST1971(2), ST3393(1), ST3660(1)
	KPC	KPC-2	ST463(1)
Class B	IMP	IMP-9	ST292(1)
		IMP-45	ST316(1), ST381(1)
	VIM	VIM-2	**ST235(2), ST244(1)**
Class D	OXA	OXA-1(OXA-1-like)	ST16(1), ST316(1), ST381(1), ST1971(1), ST3393(1), ST3360(1)
		OXA-10(OXA-10-like)	**ST244**(1), ST292(1)
		OXA-21(OXA-2-like)	ST1971(1)
		OXA-50(OXA-50-like)	ST16(3), ST260(2), ST270(2), ST287(1), ST319(1), ST360(2), ST381(6), ST389(1), ST408(1), ST611(4), ST617(1), ST698(2), ST699(1), ST708(1), ST903(1), ST1123(1), ST1129(1), ST1667(1), ST1718(2), ST1930(1), ST1966(1), ST2211(1), ST2447(1), ST2920(1), ST3393(1), ST3658(1), ST3660(1), ST3661(1), ST3664(1), ST3665(1), ST3667(1), ST3672(1), ST3679(1), ST3682(1), ST3684(2), ST3686(1), ST3690(1), ST3692(1), ST3693(2), ST3693(1), ST3694(1), ST3695(2), NA(3)
		OXA-246(OXA-10-like)	**ST235(2)**
		OXA-486(OXA-50-like)	ST252(1), ST266(2), ST274(5), ST412(1), ST463(1), ST508(4), ST553(2), ST567(1), ST644(1), ST703(2), ST807(1), ST1295(1), ST1655(1), 1748 (1), ST3680(1), NA(1)
		OXA-488(OXA-50-like)	ST235(2), ST253(5), ST313(3), ST377(1), ST560(1), ST606(1), ST3713(1), NA(1)
		OXA-846(OXA-50-like)	**ST111(2),** ST168(1), ST207(5),ST241(2), ST273(1), ST**298(1), ST**313(1), ST316(2), ST**357(5),** ST395(1), ST446(1), ST645(1), ST773(4), ST927(2), ST1021(1), ST1086(1), ST1129(1), ST1203(1), ST 1239 (1), ST1248(1), ST1743(4), ST1920(1), ST1971(7), ST2326(2), ST3659(1), ST3675(1), ST3683(1), NA(1)
		OXA-850(OXA-50-like)	ST27(1), ST108(1), ST155(1), ST179(1), ST209(1), ST**244(9),** ST245(1), ST261(1), ST**277(4),** ST292(1), ST348(1), ST385(1), ST455(1), ST554(1), ST569(4), ST769(1), ST798(1), ST980(1), ST1247(1), ST1338(3), ST1652(1), ST1684(1), ST2069(1), ST2957(1), ST3213(1), ST3666(1), ST3670(1), ST3674(2), ST3677(1), ST3685(1), ST3687(1), ST3688(1), ST3691(1)
		OXA-906(OXA-50-like)	ST532(3)

The bold values represent observed global *P. aeruginosa* high-risk clones.

Although previous studies had showed that high-risk clones were predominant among multidrug-resistance strains (del Barrio-Tofiño et al., 2020; Hu et al., 2021; Teo et al., 2021; Aguilar-Rodea et al., 2022), we found that 50/212 isolates (23.58%) belong to the 18 previously reported high-risk clones, and that the remaining isolates showed high diversity together with other reported STs in all strains.

Among the high-risk clones found in our study, we observed that ST111 included two MDR/non-MDR strains from different hospitals. ST235 is the most widely spread *P. aeruginosa* high-risk clone associated MDR/XDR isolates and carries various resistance genes (Treepong et al., 2018; Tahmasebi et al., 2022). Three ST235 MDR strains were isolated from the neurology ward, of which two were carrying the $bla_{\text{VIM-2}}$ gene. The most prevalent high-risk clone in all isolated strains was ST244 ($n = 9$); six multidrug resistance strains were collected from the ICU, of which one harbored the $bla_{\text{VIM-2}}$ gene. ST277 was reported to be highly prevalent in Brazil, and commonly associated with MBL SPM-1 (Galetti et al., 2019; Silveira et al., 2020). We found ST277 isolates composed of 4 MDR/XDR strains, but the specific MBLs genes were not identified. ST357 with bla_{IMP} and bla_{VIM} type MBLs was previously reported in Turkey (Çekin et al., 2021) and Czech (Papagiannitsis et al., 2017), three ST357 stains identified in our study showed MDR/XDR phenotype, and one of them harbored $bla_{\text{CARB-3}}$ gene. ST298 is the subclade of the CC446 clonal complex (Pincus et al., 2019). One ST298 isolate and one ST446 isolate, both with MDR/XDR phenotype, were identified in different hospitals. In some regional research projects, the clonal diversity of *P. aeruginosa* is much lower among multidrug-resistant isolates (del Barrio-Tofiño et al., 2020; Hu et al., 2021; Teo et al., 2021; Aguilar-Rodea et al., 2022). This is not entirely the case among our MDR subsets from 2018–2020. We have observed 84 (39.62%) different STs in the 130 MDR/XDR

FIGURE 3
Heat map showing the resistance genes for the collected *P. aeruginosa* high-risk clone strains. The heat map was formed by performing tblastn searches using the sequences present on the Comprehensive Antibiotic Resistance Database (CARD) as queries and the genomes as subjects. Grey indicates the absence, and green indicates the presence of the resistance genes.

strains. However, only 12.26% (n = 26) of MDR/XDR isolates are actually related to the reported international high-risk clones.

We note that our study is focused on the Guangdong Province. This region is the most populous province with the largest population inflow in China. We speculate that trade and mobility in this region have contributed to the high level of diversity observed among the isolated strains. Overall, our survey of regional *P. aeruginosa* high-risk clones in China may assist in a better understanding of the evolution of these important clone types.

Besides the high-risk clones, eight founder STs of clonal groups/subgroups have been identified (Figure 2), and these founder STs have been reported to be distributed in at least three different countries (Table 2). Furthermore, one special disseminated and multidrug-resistant clone, ST1971 (n = 7, 3.30%), has only been identified in China (Wang et al., 2017). This clone can be defined as a regional epidemic risk type for nosocomial healthcare. Some of the STs identified in this study have been previously described as new potential high-risk clones, including ST108 (Telling et al., 2018), ST179 (Moloney et al., 2020), ST253 (Fischer et al., 2020), ST260 (Martak et al., 2022), ST274 (Sewunet et al., 2022), ST348 (Dix et al., 2022), ST395 (Petitjean et al., 2017), ST446 (Pincus et al., 2019), ST463 (Hu et al., 2021), ST532 (Founou et al., 2020), ST699 (Rada et al., 2021) and ST773 (Singh et al., 2021).

The antibiotic profiles of all isolated *P. aeruginosa* strains were quantified by MIC assays. 196 strains (92.45%) were found to be CRPA, and most of them represented resistance to multiple antibiotics (126/196, 64.29%). It was previously reported that meropenem and imipenem resistance rates for *P. aeruginosa* range from 21.6 to 25.6% in China (Wang and Wang, 2020), and the detection rate of CRPA in Guangdong was 41.9, 17.9 and 19.2% from 2018 to 2020, respectively. The CRPA rates determined in this study were higher than the previous studies. Previous studies reported that the CRPA isolating rate of inpatients and those admitted to an ICU is higher than outpatients and those in non-ICU wards (Gill, Nicolau, and on behalf of the ERACE-PA Global Study Group, 2022). Here, 35.85% of strains were isolated from ICU patients, which might explain the high prevalence of CRPA in the collection. In future studies, collecting strains from more diverse wards and improving the randomness of the evaluation would be beneficial.

In contrast to the high rate of carbapenem resistance, the MBLs genes carrying strains were with very low proportion (n = 6, 2.83%) in all collected isolates. Only 3 bla_{IMP} harbored strains and 3 $bla_{\mathrm{VIM\text{-}2}}$ harbored strains were found. The *oprD* gene has been reported as the main determinant of imipenem resistance (Quinn et al., 2022). The MIC data show that the imipenem resistance rate is 87.26%. The wild-type *oprD* gene was not found in any newly sequenced strains, which implied that the mutational inactivation of the *oprD* gene may play a role in carbapenem resistance. However, the relationship between the mutations in the *oprD* gene and its function cannot be fully understood only by whole genome sequencing, and future investigations should focus on mechanistic studies of these mutations.

Some variants of PDC enzymes (class C β-lactamases) are responsible for the resistance to imipenem and newer β-lactam/β-lactamase inhibitor combinations such as ceftolozane-tazobactam (Papp-Wallace et al., 2020). Furthermore, the production of class D β-lactamases (oxacillinase; OXA) can lead to third-generation cephalosporins or multiple-drug resistance (Vrancianu et al., 2020). In this study, PDC β-lactamase and OXA β-lactamase genes were prevalent in collected strains, which can induce multidrug resistance to some extent by antibiotic inactivation. Previous studies showed that the presence of *qnrVC* genes plays important roles in quinolone resistance (Liu et al., 2018; Saki et al., 2022), and 2.3% of collected

P. aeruginosa isolates from Guangdong carried *qnrVC* genes (Lin et al., 2020), which is consistent with this study.

The type III secretion system is an important virulence factor of *P. aeruginosa*, contributing to microbial cytotoxicity and invasion (Horna and Ruiz, 2021). The *exoU* and *exoS* genes encode type III secretion virulence effector proteins. Previous studies have shown the mutual exclusion of the *exoU* and *exoS* genes (Bradbury et al., 2010; Juan et al., 2017). Through the VFDB analysis, we found the co-presence of both genes in 5 clone types strains (ST273, ST274, ST463, and ST3658). The $exoS^{+}$/$exoU^{+}$ strains have no significant differences in the resistance phenotype, but the strains carrying two virulence genes were all classified into XDR (CRPA).

5. Conclusion

*Pseudomonas aeruginos*a isolates from hospitals in this study exhibited multidrug resistance to various drugs and a high rate of resistance against carbapenem. MLST analysis further revealed the genetic diversity of these clinical isolates of *P. aeruginosa*. The stLFR technology was first applied for the bacteria antibiotic gene profiling, and the findings indicate that the sequence clone types are independent of carbapenem resistance or MDR/XDR phenotype. The strains harboring bla_{OXA} and bla_{PDC} genes mainly circulate in the hospital environments studied. The spectrum of high-risk clones is also different from previous reports and has obvious regional characteristics. This means that the control measures for the circulation and success of high-risk clones and CRPA must have a unified global standard and custom local strategies.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number (s) can be found in the article/Supplementary material.

Ethics statement

The study was approved by the Ethics Committee (full name: The Ethics Committee of the Zhujiang Hospital of Southern Medical University), the reference number 2021-KY-046-01. The samples were obtained with written informed consent and reviewed by the ethical board in accordance with the tenets of the Declaration of Helsinki.

Author contributions

YZ designed the experimental plan, analyzed WGS data, and finished the draft manuscript writing. DC collected strains and participated in project design. BJ prepared MLST distribution and digital chart. XZ contributed to manuscript revision. MA participated in project design. LJ contributed to conceptualization, project administration, and manuscript revision. All authors participated in the manuscript critical review and editing, and agreed with the above contribution details.

Funding

This work was supported by a research grant to LJ from the Independent Research Fund Denmark (9039-00350A).

Acknowledgments

The authors would like to thank all reviewers for their insightful comments and constructive suggestions to improve the manuscript.

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.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1117017/full#supplementary-material

References

Aguilar-Rodea, P., Estrada-Javier, E. L., Jiménez-Rojas, V., Gomez-Ramirez, U., Nolasco-Romero, C. G., Rodea, G. E., et al. (2022). New variants of *Pseudomonas aeruginosa* high-risk clone ST233 associated with an outbreak in a Mexican Paediatric Hospital. *Microorganisms* 10:1533. doi: 10.3390/microorganisms10081533

Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., et al. (2020). CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 48, D517–D525. doi: 10.1093/nar/gkz935

Bishara, A., Moss, E. L., Kolmogorov, M., Parada, A. E., Weng, Z., Sidow, A., et al. (2018). High-quality genome sequences of uncultured microbes by assembly of read clouds. *Nat. Biotechnol.* 36, 1067–1075. doi: 10.1038/nbt.4266

Botelho, J., Grosso, F., and Peixe, L. (2019). Antibiotic resistance in *Pseudomonas aeruginosa* – Mechanisms, epidemiology and evolution. *Drug Resist. Updat.* 44:100640. doi: 10.1016/j.drup.2019.07.002

Bradbury, R. S., Roddam, L. F., Merritt, A., Reid, D. W., and Champion, A. C. (2010). Virulence gene distribution in clinical, nosocomial and environmental isolates of *Pseudomonas aeruginosa*. *J. Med. Microbiol.* 59, 881–890. doi: 10.1099/jmm.0.018283-0

Çekin, Z. K., Dabos, L., Malkoçoğlu, G., Fortineau, N., Bayraktar, B., Iorga, B. I., et al. (2021). Carbapenemase-producing *Pseudomonas aeruginosa* isolates from Turkey: first report of *P. aeruginosa* high-risk clones with VIM-5-and IMP-7-type carbapenemases in a tertiary hospital. *Diagn. Microbiol. Infect. Dis.* 99:115174. doi: 10.1016/j.diagmicrobio.2020.115174

del Barrio-Tofiño, E., López-Causapé, C., and Oliver, A. (2020). *Pseudomonas aeruginosa* epidemic high-risk clones and their association with horizontally-acquired β-lactamases: 2020 update. *Int. J. Antimicrob. Agents* 56:106196. doi: 10.1016/j.ijantimicag.2020.106196

Dix, L. M. L., de Goeij, I., Manintveld, O. C., Severin, J. A., and Verkaik, N. J. (2022). *Pseudomonas aeruginosa* left ventricular assist device (LVAD) driveline infection acquired from the bathroom at home. *Am. J. Infect. Control* 50, 1392–1394. doi: 10.1016/j.ajic.2022.04.011

Fan, X., Wu, Y., Xiao, M., Xu, Z. P., Kudinha, T., Bazaj, A., et al. (2016). Diverse genetic background of multidrug-resistant *Pseudomonas aeruginosa* from mainland China and emergence of an extensively drug-resistant ST292 clone in Kunming. *Sci. Rep.* 6:26522. doi: 10.1038/srep26522

Fischer, S., Dethlefsen, S., Klockgether, J., and Tümmler, B. (2020). Phenotypic and genomic comparison of the two Most common ExoU-positive *Pseudomonas aeruginosa* clones, PA14 and ST235. *mSystems* 5, e01007–e01020. doi: 10.1128/mSystems.01007-20

Founou, R. C., Founou, L. L., Allam, M., Ismail, A., and Essack, S. Y. (2020). First report of a clinical multidrug-resistant *Pseudomonas aeruginosa* ST532 isolate harbouring a ciprofloxacin-modifying enzyme (CrpP) in South Africa. *J. Glob. Antimicrob. Resist.* 22, 145–146. doi: 10.1016/j.jgar.2020.05.012

Galetti, R., Andrade, L. N., Varani, A. M., and Darini, A. L. C. (2019). SPM-1-producing *Pseudomonas aeruginosa* ST277 carries a chromosomal pack of acquired resistance genes: an example of high-risk clone associated with "intrinsic resistome". *J. Glob. Antimicrob. Resist.* 16, 183–186. doi: 10.1016/j.jgar.2018.12.009

Gill, C. M., and Nicolau, D. P.on behalf of the ERACE-PA Global Study Group (2022). Carbapenem-resistant *Pseudomonas aeruginosa*: an assessment of frequency of isolation from ICU versus non-ICU, phenotypic and genotypic profiles in a multinational population of hospitalized patients. *Antimicrob. Resist Infect Control* 11:146. doi: 10.1186/s13756-022-01187-8

Gravningen, K., Kacelnik, O., Lingaas, E., Pedersen, T., Iversen, B. G., Pseudomonas outbreak groupet al. (2022). *Pseudomonas aeruginosa* countrywide outbreak in hospitals linked to pre-moistened non-sterile washcloths, Norway, October 2021 to April 2022. *Eur. Secur.* 27:2200312. doi: 10.2807/1560-7917.ES.2022.27.18.2200312

Horna, G., and Ruiz, J. (2021). Type 3 secretion system of *Pseudomonas aeruginosa*. *Microbiol. Res.* 246:126719. doi: 10.1016/j.micres.2021.126719

Hu, Y., Peng, W., Wu, Y., Li, H., Wang, Q., Yi, H., et al. (2021). A potential high-risk clone of *Pseudomonas aeruginosa* ST463. *Front. Microbiol.* 12:670202. doi: 10.3389/fmicb.2021.670202

Humphries, R., Bobenchik, A. M., Hindler, J. A., and Schuetz, A. N. (2021). Overview of changes to the clinical and laboratory standards institute performance standards for antimicrobial susceptibility testing, M100, 31st edition. *J. Clin. Microbiol.* 59, e00213–e00221. doi: 10.1128/JCM.00213-21

Jolley, K. A., Bray, J. E., and Maiden, M. C. J. (2018). Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* 3:124. doi: 10.12688/wellcomeopenres.14826.1

Juan, C., Peña, C., and Oliver, A. (2017). Host and pathogen biomarkers for severe *Pseudomonas aeruginosa* infections. *J. Infect. Dis.* 215, S44–S51. doi: 10.1093/infdis/jiw299

Kawalek, A., Kotecka, K., Modrzejewska, M., Gawor, J., Jagura-Burdzy, G., and Bartosik, A. A. (2020). Genome sequence of *Pseudomonas aeruginosa* PAO1161, a PAO1 derivative with the ICEPae1161 integrative and conjugative element. *BMC Genomics* 21:14. doi: 10.1186/s12864-019-6378-6

Lin, J., Chen, D. Q., Hong, J., Huang, H., and Xu, X. (2020). Prevalence of qnrVC genes in *Pseudomonas aeruginosa* clinical isolates from Guangdong, China. *Curr. Microbiol.* 77, 1532–1539. doi: 10.1007/s00284-020-01974-9

Liu, J., Yang, L., Chen, D., Peters, B. M., Li, L., Li, B., et al. (2018). Complete sequence of pBM413, a novel multidrug resistance megaplasmid carrying qnrVC6 and blaIMP-45 from *Pseudomonas aeruginosa*. *Int. J. Antimicrob. Agents* 51, 145–150. doi: 10.1016/j.ijantimicag.2017.09.008

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

Magiorakos, A.-P., Srinivasan, A., Carey, R. B., Carmeli, Y., Falagas, M. E., Giske, C. G., et al. (2012). Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin. Microbiol. Infect.* 18, 268–281. doi: 10.1111/j.1469-0691.2011.03570.x

Martak, D., Gbaguidi-Haore, H., Meunier, A., Valot, B., Conzelmann, N., Eib, M., et al. (2022). High prevalence of *Pseudomonas aeruginosa* carriage in residents of French and German long-term care facilities. *Clin. Microbiol. Infect.* 28, 1353–1358. doi: 10.1016/j.cmi.2022.05.004

Moloney, E. M., Deasy, E. C., Swan, J. S., Brennan, G. I., O'Donnell, M. J., and Coleman, D. C. (2020). Whole-genome sequencing identifies highly related *Pseudomonas aeruginosa* strains in multiple washbasin U-bends at several locations in one hospital: evidence for trafficking of potential pathogens via wastewater pipes. *J. Hosp. Infect.* 104, 484–491. doi: 10.1016/j.jhin.2019.11.005

Oliver, A., Mulet, X., López-Causapé, C., and Juan, C. (2015). The increasing threat of *Pseudomonas aeruginosa* high-risk clones. *Drug Resist. Updat.* 21–22, 41–59. doi: 10.1016/j.drup.2015.08.002

Pang, Z., Raudonis, R., Glick, B. R., Lin, T. J., and Cheng, Z. (2019). Antibiotic resistance in *Pseudomonas aeruginosa*: mechanisms and alternative therapeutic strategies. *Biotechnol. Adv.* 37, 177–192. doi: 10.1016/j.biotechadv.2018.11.013

Papagiannitsis, C. C., Medvecky, M., Chudejova, K., Skalova, A., Rotova, V., Spanelova, P., et al. (2017). Molecular characterization of carbapenemase-producing *Pseudomonas aeruginosa* of Czech origin and evidence for clonal spread of extensively resistant sequence type 357 expressing IMP-7 metallo-β-lactamase. *Antimicrob. Agents Chemother.* 61, e01811–e01817. doi: 10.1128/AAC.01811-17

Papp-Wallace, K. M., Mack, A. R., Taracila, M. A., and Bonomo, R. A. (2020). Resistance to novel β-lactam–β-lactamase inhibitor combinations. *Infect. Dis. Clin. N. Am.* 34, 773–819. doi: 10.1016/j.idc.2020.05.001

Petitjean, M., Martak, D., Silvant, A., Bertrand, X., Valot, B., and Hocquet, D. (2017). Genomic characterization of a local epidemic *Pseudomonas aeruginosa* reveals specific features of the widespread clone ST395. *Microb. Genom.* 3:e000129. doi: 10.1099/mgen.0.000129

Pincus, N. B., Bachta, K. E. R., Ozer, E. A., Allen, J. P., Pura, O. N., Qi, C., et al. (2019). Long-term persistence of an extensively drug-resistant subclade of globally distributed *Pseudomonas aeruginosa* clonal complex 446 in an Academic Medical Center. *Clin. Infect. Dis.* 71, 1524–1531. doi: 10.1093/cid/ciz973

Quinn, A. M., Bottery, M. J., Thompson, H., and Friman, V. P. (2022). Resistance evolution can disrupt antibiotic exposure protection through competitive exclusion of the protective species. *ISME J.* 16, 2433–2447. doi: 10.1038/s41396-022-01285-w

Rada, A. M., de la Cadena, E., Agudelo, C. A., Pallares, C., Restrepo, E., Correa, A., et al. (2021). Genetic diversity of multidrug-resistant *Pseudomonas aeruginosa* isolates carrying blaVIM–2 and blaKPC–2 genes that spread on different genetic environment in Colombia. *Front. Microbiol.* 12:663020. doi: 10.3389/fmicb.2021.663020

Saki, M., Farajzadeh Sheikh, A., Seyed-Mohammadi, S., Asareh Zadegan Dezfuli, A., Shahin, M., Tabasi, M., et al. (2022). Occurrence of plasmid-mediated quinolone resistance genes in *Pseudomonas aeruginosa* strains isolated from clinical specimens in Southwest Iran: a multicentral study. *Sci. Rep.* 12:2296. doi: 10.1038/s41598-022-06128-4

Sewunet, T., Asrat, D., Woldeamanuel, Y., Aseffa, A., and Giske, C. G. (2022). Molecular epidemiology and antimicrobial susceptibility of *Pseudomonas* spp. and *Acinetobacter* spp. from clinical samples at Jimma medical center, Ethiopia. *Front. Microbiol.* 13:951857. doi: 10.3389/fmicb.2022.951857

Silveira, M. C., Rocha-de-Souza, C. M., Albano, R. M., de Oliveira Santos, I. C., and Carvalho-Assef, A. P. D'. A. (2020). Exploring the success of Brazilian endemic clone *Pseudomonas aeruginosa* ST277 and its association with the CRISPR-Cas system type I-C. *BMC Genomics* 21:255. doi: 10.1186/s12864-020-6650-9

Singh, S., Pulusu, C. P., Pathak, A., Pradeep, B. E., and Prasad, K. N. (2021). Complete genome sequence of an extensively drug-resistant *Pseudomonas aeruginosa* ST773 clinical isolate from North India. *J. Glob. Antimicrob. Res.* 27, 244–246. doi: 10.1016/j.jgar.2021.10.010

Tahmasebi, H., Dehbashi, S., Nasaj, M., and Arabestani, M. R. (2022). Molecular epidemiology and collaboration of siderophore-based iron acquisition with surface adhesion in hypervirulent *Pseudomonas aeruginosa* isolates from wound infections. *Sci. Rep.* 12:7791. doi: 10.1038/s41598-022-11984-1

Talebi Bezmin Abadi, A., Rizvanov, A. A., Haertlé, T., and Blatt, N. L. (2019). World Health Organization report: current crisis of antibiotic resistance. *Bio Nano Sci.* 9, 778–788. doi: 10.1007/s12668-019-00658-4

Telling, K., Laht, M., Brauer, A., Remm, M., Kisand, V., Maimets, M., et al. (2018). Multidrug resistant *Pseudomonas aeruginosa* in Estonian hospitals. *BMC Infect. Dis.* 18:513. doi: 10.1186/s12879-018-3421-1

Teo, J. Q.-M., Tang, C. Y., Lim, J. C., Lee, S. J. Y., Tan, S. H., Koh, T. H., et al. (2021). Genomic characterization of carbapenem-non-susceptible *Pseudomonas aeruginosa* in Singapore. *Emerg. Microbes. Infect.* 10, 1706–1716. doi: 10.1080/22221751.2021.1968318

Thornton, C. S., and Parkins, M. D. (2023). Microbial epidemiology of the cystic fibrosis airways: past, present, and future. *Semin. Respir. Crit. Care Med.* 44, 269–286. doi: 10.1055/s-0042-1758732

Treepong, P., Kos, V. N., Guyeux, C., Blanc, D. S., Bertrand, X., Valot, B., et al. (2018). Global emergence of the widespread *Pseudomonas aeruginosa* ST235 clone. *Clin. Microbiol. Infect.* 24, 258–266. doi: 10.1016/j.cmi.2017.06.018

Vrancianu, C. O., Gheorghe, I., Czobor, I. B., and Chifiriuc, M. C. (2020). Antibiotic resistance profiles, molecular mechanisms and innovative treatment strategies of *Acinetobacter baumannii*. *Microorganisms* 8:935. doi: 10.3390/microorganisms8060935

Wang, K., Chen, Y. Q., Salido, M. M., Kohli, G. S., Kong, J. L., Liang, H. J., et al. (2017). The rapid in vivo evolution of *Pseudomonas aeruginosa* in ventilator-associated pneumonia patients leads to attenuated virulence. *Open Biol.* 7:170029. doi: 10.1098/rsob.170029

Wang, O., Chin, R., Cheng, X., Wu, M. K. Y., Mao, Q., Tang, J., et al. (2019). Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. *Genome Res.* 29, 798–808. doi: 10.1101/gr.245126.118

Wang, W., and Wang, X. (2020). Prevalence of metallo-β-lactamase genes among *Pseudomonas aeruginosa* isolated from various clinical samples in China. *J. Lab. Med.* 44, 197–203. doi: 10.1515/labmed-2019-0162

Xu, J., Deng, Y., Yang, J., Huang, W., Yan, Y., Xie, Y., et al. (2022). Effect of population migration and socioeconomic factors on the COVID-19 epidemic at county level in Guangdong, China. *Front. Environ. Sci.* 10:841996. doi: 10.3389/fenvs.2022.841996

Yoon, E.-J., and Jeong, S. H. (2021). Mobile carbapenemase genes in *Pseudomonas aeruginosa*. *Front. Microbiol.* 12:614058. doi: 10.3389/fmicb.2021.614058

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 17 April 2023
DOI 10.3389/fmicb.2023.1160433

OPEN ACCESS

EDITED BY
Ernesto Perez-Rueda,
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas,
Universidad Nacional Autónoma de México, Mexico

REVIEWED BY
Changyong Cheng,
Zhejiang A &F University, China
Xin-jun Du,
Tianjin University of Science and Technology, China

*CORRESPONDENCE
Yin Wang
✉ 10334@sicau.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 07 February 2023
ACCEPTED 28 March 2023
PUBLISHED 17 April 2023

CITATION
Zhou Y, Jiang D, Yao X, Luo Y, Yang Z, Ren M, Zhang G, Yu Y, Lu A and Wang Y (2023) Pan-genome wide association study of *Glaesserella parasuis* highlights genes associated with virulence and biofilm formation.
Front. Microbiol. 14:1160433.
doi: 10.3389/fmicb.2023.1160433

Pan-genome wide association study of *Glaesserella parasuis* highlights genes associated with virulence and biofilm formation

You Zhou[1†], Dike Jiang[1†], Xueping Yao[1], Yan Luo[1], Zexiao Yang[1], Meishen Ren[1,2,3,4], Ge Zhang[2,3,4], Yuanyuan Yu[2,3,4], Aiping Lu[2,3,4] and Yin Wang[1*]

[1]Key Laboratory of Animal Diseases and Human Health of Sichuan Province, College of Veterinary Medicine, Sichuan Agricultural University, Chengdu, China, [2]Law Sau Fai Institute for Advancing Translational Medicine in Bone and Joint Diseases (TMBJ), School of Chinese Medicine, Hong Kong Baptist University, Kowloon Tong, Hong Kong SAR, China, [3]Guangdong-Hong Kong-Macao Greater Bay Area International Research Platform for Aptamer-Based Translational Medicine and Drug Discovery (HKAP), Hong Kong, Hong Kong SAR, China, [4]Institute of Integrated Bioinformedicine and Translational Science (IBTS), School of Chinese Medicine, Hong Kong Baptist University, Kowloon Tong, Hong Kong SAR, China

Glaesserella parasuis is a gram-negative bacterium that causes fibrotic polyserositis and arthritis in pig, significantly affecting the pig industry. The pan-genome of *G. parasuis* is open. As the number of genes increases, the core and accessory genomes may show more pronounced differences. The genes associated with virulence and biofilm formation are also still unclear due to the diversity of *G. parasuis*. Therefore, we have applied a pan-genome-wide association study (Pan-GWAS) to 121 strains *G. parasuis*. Our analysis revealed that the core genome consists of 1,133 genes associated with the cytoskeleton, virulence, and basic biological processes. The accessory genome is highly variable and is a major cause of genetic diversity in *G. parasuis*. Furthermore, two biologically important traits (virulence, biofilm formation) of *G. parasuis* were studied *via* pan-GWAS to search for genes associated with the traits. A total of 142 genes were associated with strong virulence traits. By affecting metabolic pathways and capturing the host nutrients, these genes are involved in signal pathways and virulence factors, which are beneficial for bacterial survival and biofilm formation. This research lays the foundation for further studies on virulence and biofilm formation and provides potential new drug and vaccine targets against *G. parasuis*.

KEYWORDS

***Glaesserella parasuis*, pan-genome, GWAS, the core genome, the accessory genome, virulence, biofilm formation**

1. Introduction

Glaesserella parasuis (formerly known as Haemophilus parasuis) is the etiologic agent of Glässer's disease in pigs. Glässer's disease is characterized by fibrinous polyserositis and meningitis (Oliveira and Pijoan, 2004). Glässer's disease is an important infectious disease in the world pig industry (Sun et al., 2022). Glässer's disease is not only the main infectious disease of nursery pigs but also affects the health of fattening pigs and sows. *G. parasuis* has been reported to bring substantial economic losses to the pig industry in various countries and is the

third most crucial bacterial pathogen affecting finisher herds (Rivera-Benítez et al., 2021).

The pan-genome is the assemblage of all the genes in a given species (Barh et al., 2020; Sherman and Salzberg, 2020; Her et al., 2021). Generally, the pan-genome consists of the core genome and the accessory genome. The boundaries of the core genome can be extrapolated from highly conserved genes (Iranzadeh and Mulder, 2019). Importantly, the core genome is the backbones of all strains of a specie (Segerman, 2012). The accessory genome assemblages of genes unique to different individuals (Kung et al., 2010). The accessory genome reflects bacteria resistance and adaptation to different environmental conditions (Jackson et al., 2011). Moreover, some studies have reported that the pan-genome of *G. parasuis* is open (Li et al., 2013). The openness of *G. parasuis* pan-genome reflects the diversity of the gene pool among the strains of *G. parasuis* (Bentley, 2009). Open pan-genome is the main reason for the low cross-protection of existing vaccines. As the number of genes increases, the core genome and the accessory genome may show further differences in virulence, function, metabolic pathways, etc. However, the specific differences between the core genome and the accessory genome are uncertain in *G. parasuis*.

To date, *G. parasuis* can be divided into 15 serovars *via* a gel immuno-diffusion assay. Moreover, a molecular serotyping method is established according to the polysaccharide capsule of the standard strains (Nedbalcova et al., 2006). Generally, some serovars, such as 4, 5, and 13, are more pathogenic than others (Rapp-Gabrielson and Gabrielson, 1992). However, specific serovars previously considered nontoxic can still exhibit pathogenicity. These cases may reveal no absolute relationship between virulence and serovars (Aragon et al., 2010). In clinical cases, some clinical symptoms presenting as acute or chronic may result from the virulence of different strains. Although the genome of *G. parasuis* has long been obtained, the function and essentiality of the annotated genes remain largely unknown, especially virulence factors (Costa-Hurtado and Aragon, 2013; Zhang et al., 2014). In pathogenic bacteria, virulence factors play a critical role in pathogenesis. For example, the VtaA8 and VtaA9 proteins, belonging to the member of trimeric autotransporters, have phagocytosis resistance to evade the host innate immune responses (Costa-Hurtado et al., 2012). However, only very few virulence factors were identified in previous studies. The role of most virulence factors in adhesion to and invasion of host cells, resistance to phagocytosis by macrophages, resistance to serum complement, and induction of inflammation remains uncertain.

Biofilm formation is central to promoting colonization, changing strains' adhesion and even increasing drug resistance (Watnick and Kolter, 2000; Lewis, 2001; Kokare et al., 2009). Moreover, biofilm formation also plays an essential role in the phenotype switch from commensal to pathogen. In previous studies, *G. parasuis* had the ability to form biofilm (Jin et al., 2006). The part gene expression levels of *G. parasuis* changed significantly before and after the formation of the biofilm (Jiang R. et al., 2021). For instance, the cytoplasmic heme-binding protein HutX, characterized by transporting the heme-degrading enzyme HutZ, can affect the formation of the biofilm and result in chronic infection of *G. parasuis* (Sekine et al., 2016). This may provide a clue to the pathogenic mechanism. Glycogen operon protein GlgX, playing an essential role in starch and sucrose metabolism pathways, can maintain the structural integrity of the biofilm, suggesting that metabolism-related genes are similarly closely related to biofilm formation (Sheppard and Howell, 2016; Jiang R. et al., 2021). However, these genes related to biofilm formation have still been explored.

In the present study, we reported a pan-genome-wide association study of genome sequence from 121 *G. parasuis* natural population isolates. This study revealed the major difference between the core genome and the accessory genome. Our results highlight that these genes with causal relationships to virulence and biofilm formation capacity were deeply excavated in the association of traits and genes.

2. Materials and methods

2.1. Bacterial strains

In this study, the complete genomes of 105 *G. parasuis* strains with geographical, virulent and serological diversity were retrieved. These strains were available in March 2020 from NCBI.[1] Information about the 105 strains is summarized in Supplementary Table S1.

Additionally, 16 clinical strains isolated in Sichuan between 2015 and 2020 were sequenced in our previous study. Information about the 16 strains were available from China National GeneBank DataBase (CNGBdb).[2] Project number is CNP0002150. Sequencing information about the 16 strains is summarized in Supplementary Table S2. And the background information of *G. parasuis* isolates showed in Table 1.

2.2. Pan-genomic construction

To sustain the consistency and reliability of gene prediction and annotation, the Prokaryotic Genome Annotation System (Prokka) pipeline (v1.14.5)[3] was uniformly applied to all the 121 *G. parasuis* genomes. Based on the GFF3 files produced by Prokka, the Roary program[4] was used to construct the pan-genome with a minimum percentage identity of 95% between each predicted protein homolog.

2.3. Association of virulence with pan-genome

Serotypes were determined *via* PCR simulation amplification. *G. parasuis* serotypes PCR identified primers were used for PCR simulation amplification the genome of 121 *G. parasuis* to identify strains serotypes, as the basis for the virulence determination. Moreover, combining the strain background information recorded in the literature and NCBI, the 121 isolates were classified as either virulent and avirulent strains or virulent, moderate virulent and avirulent strains (Table 2). Furthermore, according to literature, serotype 7 strains previously considered nontoxic were classified in the virulence strain category. Subsequently, virulence traits were

1 ftp://ftp.ncbi.nih.gov/genomes/all/
2 https://db.cngb.org/
3 https://github.com/tseemann/prokka, accessed on 5 September 2021.
4 https://github.com/sanger-pathogens/Roary, accessed on 6 October 2021.

10.3389/fmicb.2023.1160433

TABLE 1 The background information of *G. parasuis* isolates.

Name	Species identification	Isolated regions	Serotype	Isolated area (Sichuan Province)
GP 01	*Glaesserella parasuis*	Lung lesion tissue	5	Suining city
GP 02	*Glaesserella parasuis*	Lung lesion tissue	5	Mianyang city
GP 03	*Glaesserella parasuis*	Lung lesion tissue	5	Suining city
GP 04	*Glaesserella parasuis*	Lung lesion tissue	10	Nanchong city
GP 05	*Glaesserella parasuis*	Lung lesion tissue	1	Pujiang county
GP 06	*Glaesserella parasuis*	Lung lesion tissue	5	Pujiang county
GP 07	*Glaesserella parasuis*	Joint tissue effusion	Unknow	Nanchong city
GP 08	*Glaesserella parasuis*	Lung lesion tissue	5	Suining city
GP 09	*Glaesserella parasuis*	Lung lesion tissue	5	Pujiang county
GP 10	*Glaesserella parasuis*	Joint tissue effusion	Unknow	Nanchong city
GP 11	*Glaesserella parasuis*	Lung lesion tissue	7	Guangan city
GP 12	*Glaesserella parasuis*	Lung lesion tissue	Unknow	Suining city
GP 13	*Glaesserella parasuis*	Lung lesion tissue	10	Suining city
GP 14	*Glaesserella parasuis*	Lung lesion tissue	7	Xichang city
GP 15	*Glaesserella parasuis*	Joint tissue effusion	unknow	Pi county
GP 16	*Glaesserella parasuis*	Lung lesion tissue	5	Suining city

TABLE 2 Classification of the virulence of 121 *G. parasuis* strains.

Virulence classification	Serotype	Number of strains	Total
Virulent	1	6	67
	5	20	
	10	6	
	12	4	
	13	10	
	14	8	
	7	13	
Moderate virulent	2	10	35
	4	19	
	8	2	
	15	4	
Avirulent	3	1	6
	6	2	
	9	1	
	11	2	
Unknow	Unknow	13	13

translated into binary (such as "1" is a virulent strain, and "0" is an avirulent strain) and inputted Scoary.[5] Combined with the information on the presence or absence of genes in each isolate obtained by pan-genome analysis, the association between genes and virulence traits was judged based on Fisher's exact test. *P* values determined significance between genes and traits. And Odds Radio (OR) greater than 1 as a criterion for judging the strength of the positive association. Moreover, the basis on pairwise comparison algorithms in the unweighted pairs group using mean average (UPGMA) trees, the number of genes and phenotypes in evolutionary history is taken as strong evidence for a causal association.

The selected virulence trait association genes were used for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, and the virulence factors were annotated. The final results were visualized using R 4.1.

2.4. Biofilm quantification with microtiter plate assay and pan-GWAS

The mediums used to culture bacteria were Trypticase Soy Agar (TSA, OXOID, Hampshire, England) media supplemented with 5% (v/v) calf serum and nicotinamide adenine dinucleotide (NAD, JS, Nanjing, China). Bacteria were routinely streaked from −80°C stocks onto TSA and incubated at 36 h at 37°C before each experiment. Then 20 μL of the bacterial cell suspension were inoculated in 2 mL of Trypticase Soy Broth (TSB, OXOID, Hampshire, England) medium containing 5% calf serum with NAD and added to 24-well plates. Sterile media were added as a negative control. Microplates were incubated statically for 48 h at 37°C. The media and planktonic cells were removed by gentle tapping *via* inverted microplates. To remove loosely attached bacteria, wells were washed twice with 300 μL of sterile PBS. Then biofilms were fixed with 500 μL of methanol for 30 min and air-dried completely at room temperature after removal of the methanol. In order to stain bacterial biofilm, 500 μL of 1% crystal violet (CV) solution was added per well and plates were incubated statically for 30 min at room temperature.

5 https://github.com/AdmiralenOla/Scoary, accessed on 8 September 2021.

Then the solution was removed after the bacterial biofilm was completely dyed. Wells were washed 3 times with 300 μL of saline and air-dried at 37°C before taking photos of the record. Wells were added 200 μL of 33% v/v acetic acid to dissolve CV in biofilm completely. The amount of destained CV was determined by reading OD_{630} in a microplate reader.

The basis on the comparison of OD_{630} value, 16 isolates were classified into strong biofilm-forming strains and weak biofilm-forming strains. Then, the biofilm forming ability of strains was translated into binary and inputted Scoary (see footnote 5). Combined with the information on the presence or absence of genes in each isolate obtained by pan-genome analysis, the association between genes and biofilm forming ability was judged based on Fisher's exact test. The significance between genes and traits was determined by *p* values. And Odds Radio (OR) greater than 1 as a criterion for judging the strength of the positive association. Moreover, the basis of pairwise comparison algorithms in the unweighted pairs group using mean average (UPGMA) trees, the number of genes and phenotypes in evolutionary history is taken as strong evidence for a causal association.

The selected biofilm forming association genes were used for GO and KEGG enrichment. The final results were visualized using R 4.1.

2.5. Antimicrobial agent susceptibility testing

The minimum inhibitory concentrations (MICs) of 16 strains were tested as described previously studies (Zhang et al., 2019). Six drugs frequently were used to against *G. parasuis* were selected, namely, amoxicillin, ampicillin, gentamycin, kanamycin, and tetracycline. *Escherichia coli* ATCC 25922 was used as a quality control strain for drug susceptibility.

2.6. Cell adhesion experiments

Adhesion experiments were performed using PK-15 cells. The experiments were performed based on a previously described (Jiang C. et al., 2021) with some minor modifications. Briefly, the 5×10^5 cells were seeded into 24-well plates in dulbecco's modified eagle medium (DMEM, Gibco, California, United States) medium containing 10% fetal bovine serum (FBS, Gibco, California, United States) at 37°C in a humidified incubator containing 5% CO_2. After the cells have grown to 100% confluence, the cells were washed thrice with PBS and infected with approximately 1×10^7 CFU *G. parasuis*. The culture plates were thereafter incubated at 37°C for 2 h to allow bacterial adhesion. The plates were then washed five times with sterile PBS to eliminate non-specific bacterial attachment and then incubation with 200 μL 0.25% trypsin/EDTA at 37°C for 10 min. After incubation, the cells were resuspended from the bottom of every well. The cell suspensions with adherent bacteria were diluted by 10 times and placed onto the TSA (OXOID, Hampshire, United Kingdom) plates supplemented with NAD and serum and then incubation at 37°C for 12 h. And then cell counting was performed, and the adhesion rate was calculated.

3. Results

3.1. Separation and composition of the core genome and the accessory genome

In this study, the pan-genome was classified into the core genome and the accessory genome: 1,133 genes were included in the core genome (i.e., present in more than 95% of the strains, core and soft-core genes) and 7,752 genes in the accessory genome (i.e., less than 95% of the strains, shell and cloud genes). More detailed information is presented in Table 3.

Our results were consistent with previous studies that *G. parasuis* has an open pan-genome. The core genome of *G. parasuis* is small, accounting for only 4.39% of the pan-genome. The rest of the genome is highly variable, continuously obtaining foreign genes to adapt to different environments.

3.2. The COG functional classification of the core genome and the accessory genome

In order to study the functions of the different gene sets, we used an eggNOG-mapper to align and annotate with clusters of orthologous groups (COG). By clustering the analysis of the COG function, we found that the backbone of the genome and the basic biological phenotypes consisted of the core genome (Table 4 and Figure 1). The core genome participates in all link of the bacterial life process and all function classification includes some proportion of the core genome. The primary function of the core genome focuses on the E (Amino-acid transport and metabolism), J (Translation, ribosomal structure, and biogenesis), and T (Signal transduction mechanisms) of COG functional classifications. Particularly, the cytoskeleton was only found in the core genes, suggesting that the core genes play an important role in maintaining the cell morphology and the internal cell structure. The Accessory genome includes all COG functional classifications except cytoskeleton (Table 4 and Figure 1). It means that the accessory genome will also participate in a part of the life process. Moreover, the major function of the accessory genome focuses on the X (Mobilome: prophages, transposons), V (Defense mechanisms), and G (Carbohydrate transport and metabolism) of COG functional classifications. These functions may confer selective advantages to *G. parasuis* and enrich population diversity (Supplementary Table S3). Notably, up to ~17.11% of the core and accessory genome still have unknown functions in the COG database (Table 4).

3.3. The PHI annotation of the core genome and the accessory genome

Pathogen-host interactions are usually the basis of the occurrence of infectious diseases. In this study, BLASTp was used to confirm genes related to pathogen-host interactions (PHI). Four classes of gene sets were aligned to the pathogen-host interaction database and annotated. Our results showed that although pathogenic genes are widely found in both core and accessory genome, the accessory genome has a higher abundance than the core genome. All related genes annotated to Glasser's disease are mainly present in the

10.3389/fmicb.2023.1160433

TABLE 3 Characteristics of the core genome and the accessory genome.

Pan-genome	Classification	Number of genes	Max length (bp)	Min length (bp)	Average length (bp)
Core genome	Core genes	390	4,491	132	821.93
	Soft-core genes	743	4,029	123	952.79
Accessory genome	Shell genes	1880	17,223	123	890.51
	Cloud genes	5,872	11,295	123	715.25

TABLE 4 COG annotation of the core genome and the accessory genome.

Classifications	Functions	Core	Soft_core	Shell	Cloud
A	RNA processing and modification	1 (0.24%)	NA	NA	1 (0.08%)
C	Energy production and conversion	23 (5.58%)	58 (7.47%)	33 (4.7%)	68 (5.69%)
D	Cell-cycle control, cell division, and chromosome assignment	11 (2.67%)	7 (0.9%)	14 (1.99%)	20 (1.67%)
E	Amino acid transport and metabolism	38 (9.22%)	81 (10.44%)	55 (7.83%)	86 (7.19%)
F	Nucleotide transport and metabolism	10 (2.43%)	38 (4.9%)	16 (2.28%)	43 (3.6%)
G	Carbohydrate transport and metabolism	23 (5.58%)	53 (6.83%)	61 (8.69%)	102 (8.53%)
H	Coenzyme transport and metabolism	20 (4.85%)	60 (7.73%)	41 (5.84%)	68 (5.69%)
I	Lipid transport and metabolism	12 (2.91%)	19 (2.45%)	16 (2.28%)	27 (2.26%)
J	Translation, ribosomal structure and biogenesis	55 (13.35%)	107 (13.79%)	50 (7.12%)	109 (9.11%)
K	Transcription	18 (4.37%)	36 (4.64%)	57 (8.12%)	63 (5.27%)
L	Replication, recombination and repair	19 (4.61%)	47 (6.06%)	47 (6.7%)	71 (5.94%)
M	Cell wall/membrane/envelope biogenesis	40 (9.71%)	58 (7.47%)	48 (6.84%)	104 (8.7%)
N	Cell motility	2 (0.49%)	5 (0.64%)	4 (0.57%)	9 (0.75%)
O	Posttranslational modification, protein turnover, chaperones	29 (7.04%)	48 (6.19%)	33 (4.7%)	68 (5.69%)
P	Inorganic ion transport and metabolism	26 (6.31%)	44 (5.67%)	31 (4.42%)	61 (5.1%)
Q	Secondary metabolites biosynthesis, transport and catabolism	3 (0.73%)	3 (0.39%)	2 (0.28%)	3 (0.25%)
R	General function prediction only	10 (2.43%)	36 (4.64%)	44 (6.27%)	69 (5.77%)
S	Function unknown	24 (5.83%)	21 (2.71%)	32 (4.56%)	48 (4.01%)
T	Signal transduction mechanisms	21 (5.1%)	24 (3.09%)	15 (2.14%)	29 (2.42%)
U	Intracellular trafficking, secretion, and vesicular transport	8 (1.94%)	16 (2.06%)	11 (1.57%)	30 (2.51%)
V	Defense mechanisms	14 (3.4%)	11 (1.42%)	31 (4.42%)	41 (3.43%)
W	Extracellular structures	2 (0.49%)	2 (0.26%)	6 (0.85%)	12 (1%)
X	Mobilome: prophages, transposons	1 (0.24%)	2 (0.26%)	55 (7.83%)	64 (5.35%)
Z	Cytoskeleton	2 (0.49%)	NA	NA	NA

accessory genome. These genes relate to bacteria's serum resistance, adhesion, and invasion capabilities. Additionally, lethal factors were not found in the core gene. This result also provided evidence that the primary function of the core genes is to perform the necessary biological processes (as shown in Figure 2 and Table 5). And more detailed information are summarized in Supplementary Table S4.

3.4. The VFDB annotation of the core genome and the accessory genome

The virulence genes of pathogenic bacteria encode toxins, adhesins or other virulence factors. Four gene sets were compared with Virulence factors of pathogenic bacteria (VFDB) databases and annotated to identify potential virulence factors. In VFDB-annotation genes, the core genome was found to exist in all classifications. Interestingly, compared with the accessory genome, the core genome has the highest proportion of nutritional/metabolic factors, biofilm formation, and effector delivery system (Table 6). These cases supported our previous conclusion. Namely, the functions of the core genome focus on maintaining survival. Additionally, the core genome is essential for bacteria to interfere with host defenses or regulate their invasion (Figure 3).

In contrast, virulence factors of the accessory genome mainly focus on B (Biofilm formation), C (Adherence), D (Immune modulation), F (Invasion), and K (Capsule). The core and accessory genomes showed differences in the composition of adherence-associated genes. For example, bartonella adhesin A (BadA)/

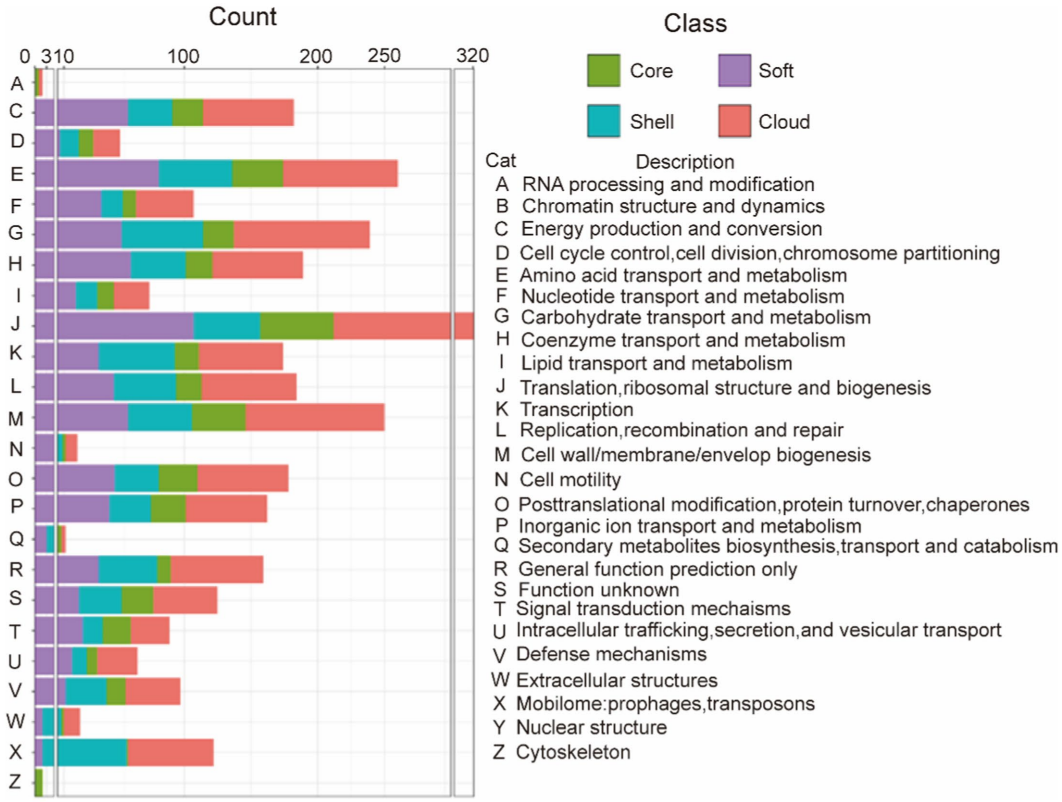


FIGURE 1
COG annotation of the core genome and the accessory genome. The category is shown in Table 2, green represents the core gene, purple represents the soft-core gene, blue represents the shell gene, and red represents cloud genes.

variably expressed outer-membrane proteins (Vomp), belonging to cloud genes, can promote bacterial autoaggregation as well as the adhesion of extracellular matrix proteins, resulting in activating hypoxia-inducible factor-1 (HIF-1) and inducing the secretion of vascular endothelial growth factor (VEGF) secretion in infected host cells (Riess et al., 2004; Zhang et al., 2004; Kaiser et al., 2008). Additionally, BadA/Vomp also is necessary to format optimal biofilm (Okaro et al., 2019). Interestingly, BadA/Vomp is not included in the core genome, but the reason for the difference is unknown.

Differences in immune modulation and stress survival between the core and accessory genome included virulence factors such as neisserial surface protein A (NspA), RecN. NspA play an essential role in binding complement factor H to inhibit host innate immune defenses (Lewis et al., 2010, 2012). RecN, a recombinational repair protein, protects against ROS and non-oxidative killing by neutrophils (Stohl et al., 2005; Criss et al., 2009). These virulence factors are beneficial for bacterial survival in the host to trigger subsequent infection.

Virulence factors for exoenzyme were also found. IgA1 protease can interfere with the barrier functions of mucosal IgA antibodies *via* cleaving secretory IgA1 in the hinge region (Rao et al., 1999). Hyaluronidase, an essential pathogenic bacterial spreading factor, can cleave hyaluronan to help bacterial spread (Baker and Pritchard, 2000; Li and Jedrzejas, 2001). Interestingly, these virulence factors classed into exoenzyme were only found in the accessory genome. Our results could indicate exoenzyme produced by the accessory genome is the dominant extracellular digestive function. This process is beneficial for the survival and spread of bacteria in the host.

Thus, various types of virulence factors were found in the core gene. The avirulent strains considered in earlier studies may also contribute to the infection process. Moreover, some avirulent strains may transfer into virulent strains *via* obtaining foreign genetic material. Virulence factors in the accessory genome are complement to the core genome. These virulence factors confer *G. parasuis* various pathogenic mechanisms to provide the opportunity for bacterial survival, resulting in more severe pathological responses in the host (Figure 3 and Table 6).

3.5. The CARD annotation of the core genome and the accessory genome

In relation to the resistance genes, the resistance gene identifier (RGI) was used to analyze four classes of gene sets. Figure 4 shows the proportion of different resistance mechanisms in perfect or strict hits or loose hits for the four gene sets.

The resistance genes were completely missing in the core genes in perfect or strict hits. Similarly, we only identified one gene about cephalosporin in the soft-core gene of the core genome. However, the accessory genome abundantly enriched other resistance genes (diaminopyrimidine, cephalosporin, peptide, sulfonamide, tetracycline, and lincosamide). Moreover, these resistance genes characteristic of the accessory genome was classified into four

10.3389/fmicb.2023.1160433

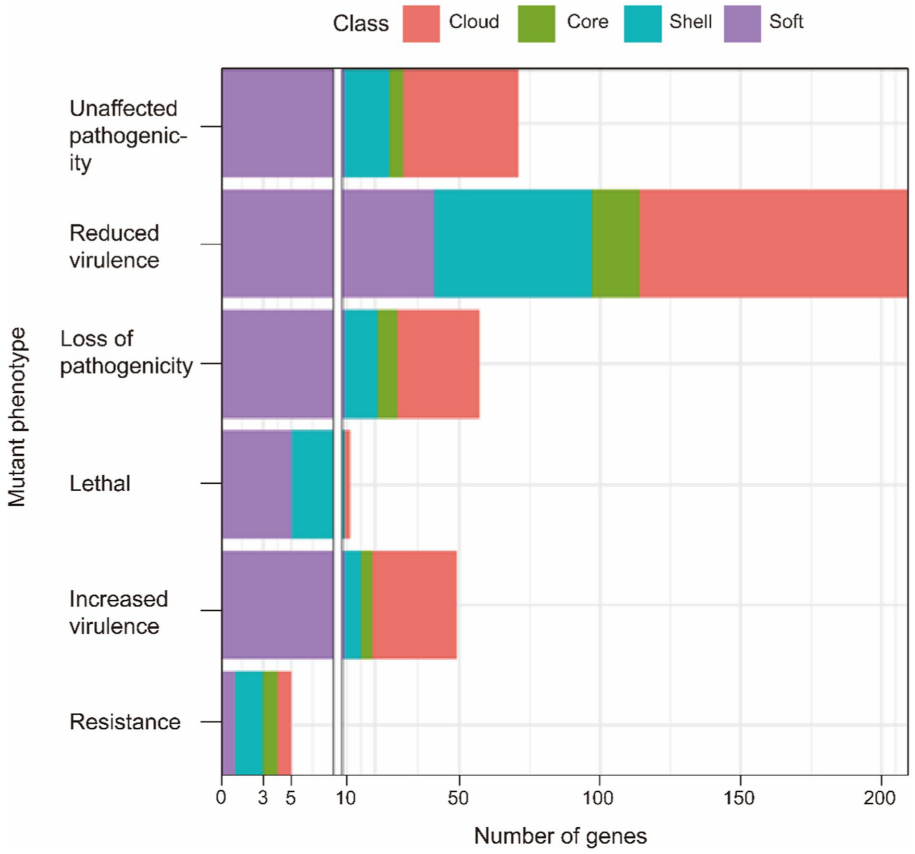


FIGURE 2
PHI annotation of the core genome and the accessory genome.

TABLE 5 PHI annotation of the core genome and the accessory genome.

Mutant phenotype	Core	Soft	Shell	Cloud
Resistance	1	1	2	1
Increased virulence	4	9	6	30
Loss of pathogenicity	7	9	12	29
Reduced virulence	17	41	56	96
Unaffected pathogenicity	5	9	16	41
Lethal	0	5	4	2

resistance mechanisms: A (antibiotic target replacement), B (antibiotic target alteration), C (antibiotic inactivation), and D (antibiotic efflux) (Table 7).

In loose hits, although resistance genes involved in six resistance mechanisms were found in both core and soft-core gene, the proportion of resistance genes is still lower than in the accessory genome (Figure 4). As in the results in perfect or strict hits, the resistance genes are mainly concentrated in the accessory genome. Our results showed that resistance genes of *G. parasuis* mostly come from the accessory genome. The genome is a dynamic repository of genetic information. The constant gain and loss of resistance genes are beneficial for bacteria to adapt to environmental stress and maintain survival in the host. Furthermore, the dynamic resistance profile also makes the clinical treatment of *G. parasuis* difficult.

3.6. The GO annotation and enrichment of the core genome and the accessory genome

BLASTx was used to compare all genes in the pan-genome with the Swiss-Prot database. The GO annotation of all genes was completed by comparing the above results with bacterial GO mapping background information. According to GO annotation files, an enricher was called to achieve the enrichment analysis of the four gene sets. *P* value (<0.05) was used to judge the significance.

Our results showed that the major structures of bacteria consist of the core genome (Supplementary Figure S1). These structures include cytoderm, cytomembrane, membrane protein complex, ribosome, etc. The core genome also participates in various biological processes of bacterial survival, including DNA replication, transcription, and translation. Moreover, the synthesis and metabolism of various necessary substances are involved in the core genome, such as RNA, amino acid, lipid, organic acid, etc. Overall, the core genome is an essential part of the basic structure and life activities of *G. parasuis* (Figure 5).

Compared to the core genome, the majority of genes characteristic of the accessory genome focus on the host cells binding, extracellular matrix binding, quinone binding, proteasome activity and so on. In addition, the core genome and the accessory genome also showed differences in biological processes. Compared to the core genome, the accessory genome pays attention to dispensable biological processes

10.3389/fmicb.2023.1160433

TABLE 6 VFDB annotation of the core genome and the accessory genome.

Classfications	Core	Soft	Shell	Cloud
Nutritional/Metabolic factor	16	32	46	38
	(23%)	(19%)	(19%)	(16%)
Biofilm formation	7	5	11	12
	(10%)	(3%)	(5%)	(5%)
Adherence	4	8	37	53
	(6%)	(5%)	(15%)	(22%)
Immune modulation	12	39	44	76
	(17%)	(23%)	(18%)	(31%)
Exotoxin	2	6	12	9
	(3%)	(4%)	(5%)	(4%)
Invasion	1	1	1	4
	(1%)	(1%)	(1%)	(2%)
Effector delivery system	7	43	32	36
	(10%)	(26%)	(13%)	(15%)
Flagella mediated movement	5	8	13	17
	(7%)	(5%)	(5%)	(7%)
Stress survival	2	9	7	11
	(3%)	(5%)	(3%)	(5%)
Regulation	1	1	2	1
	(1%)	(1%)	(1%)	(1%)
Capsule	12	13	34	46
	(17%)	(8%)	(14%)	(19%)
Exoenzyme	NA	1	5	5
		(1%)	(2%)	(2%)

including colonization, parasitism, defense reaction, anaerobic respiration and so on. This study showed that bacterial and host interrelation of abilities are given by the accessory genome. Furthermore, genes for defense against phage infection were also found in the accessory genome. The accessory genome renders bacteria to adapt the metabolic way to survive in different environments.

3.7. The KEGG annotation and enrichment of the core genome and the accessory genome

Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation and enrichment is an effective method to search for the correlation between genes and signaling pathways. In this study, clusterProfiler was used for KEGG enrichment of the core and the accessory genome.

The core genome is mainly involved in metabolism, genetic information processing, environmental information processing, cellular processes, and other signaling pathways (Figures 6A,B). Notably, the quorum sensing system and biofilm formation are one of the functional pathways involved in the core genome of *G. parasuis*. It showed that all *G. parasuis* strains had the ability to form biofilms. This ability helps bacteria resist stress and maximize energy to ensure colony survival. Furthermore, the core genome also showed resistance to cationic antimicrobial peptides and vancomycin. These resistances are beneficial for the bacteria to overcome the innate immune system in the host engraftment.

In relation to the accessory genome, this study also found specific signal pathways, including ethylbenzene degradation, naphthalene degradation, chlorinated dilute hydrocarbon degradation, arabinogalactan biosynthesis, and sphingolipid metabolism (Figure 6C). Functional genes related-pathways may be obtained from the environment by bacteria. The abundant carbon source utilization pathway facilitates the survival of *H.parasuis* in the low-oxygen lung environment or inside the biofilm in pathological conditions, resulting in long-term infection. Arabinogalactan participates in cell wall assembly to provide an antibiotic penetration barrier to bacterial (Degnan and Macfarlane, 1995). In addition, sphingolipid metabolism is beneficial for *G. parasuis* to interrupt host's sphingolipid balance leading to colonization, invasion, and intracellular survival of bacterial.

3.8. Virulence and pan-genome-wide association

A total of 121 strains were classified into virulent strains and avirulent strains. Virulence traits were grouped into binary categories for Pan-GWAS. We resulted that constant virulence factor was not found in virulence strains. This result supports that the bacterial genome is a viewpoint of a dynamic gene repository. A total of 142 genes were found to be associated with strong virulence traits in this study, involved in signal pathways such as metabolism, genetic information processing, environmental information processing, and cellular processes (Supplementary Table S5 and Figure 7). Metabolic pathways support the maintenance of normal bacterial life activities, including carbohydrate metabolism, energy metabolism, amino acid metabolism, etc. Differences in the functional pathways of genetic information processing include DNA repair, mismatch repair, homologous recombination, etc. These functional pathways showed that exogenous genetic material might be easily obtained by virulent strains. Moreover, virulent strains may have stronger adaptability to harsh environment to decrease survival pressure. Differences in the functional pathways of environmental information processing include phosphotransferase (PTS) systems and the two-component systems (TCS), which can be used for signal transduction and sensing of environmental stimuli. Other pathways such as the secretory system, quorum sensing system, and ABC transporter system take effect in bacterial colonization and acquisition of niche competitive advantage.

This study also showed the main functional categories of the virulence factors (Table 8). (1) Adherence is helpful for virulent strains colonization in the host. (2) Nutritional/metabolic, essential nutrients are necessary to colonize the host successfully. For instance, bacteria can compete or usurp iron ion from the host by utilizing heme, transferrin, and lactoferrin (Faraldo-Gómez and Sansom, 2003). (3) Exotoxin is beneficial to resist the clearance of the host immune system *via* inducing host cell apoptosis and cleaving of macrophages. (4) Regulation. For example, bacteria can regulate their growth and metabolism in response to oxygen starvation. (5) Invasion. (6) Immune modulation. It is mediated by the capsule, lipo-oligosaccharide (LOS), lipopolysaccharide (LPS) and so on. The

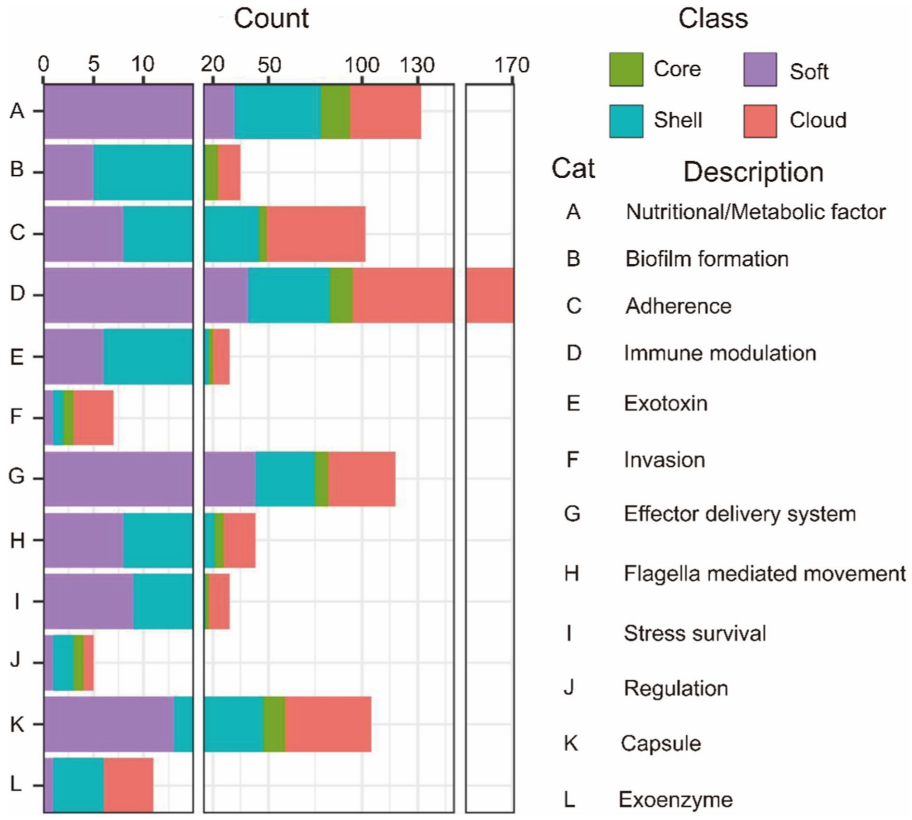


FIGURE 3
VFDB annotation of the core genome and the accessory genome.

capsule has anti-phagocytosis and anti-complement effects of maintaining bacterial long-term survival in the host (Schembri et al., 2004). LOS participates in immune evasion and is a major proinflammatory factor. LOS is helpful for bacterial survival in different microenvironments within the host (Preston et al., 1996). (7) Secretory system can lead bacterial effector proteins to enter the host cells to manipulate their cellular processes.

3.9. Biofilm formation ability and pan-genome-wide association

The biofilm formed by 16 *G. parasuis* strains on polystyrene plates is shown in Figure 8A. The OD_{630} values of biofilms of each strain were compared with those of the control group (TSB medium) by one-tailed hypothesis tests, and the p values were all lower than the test level ($\alpha = 0.05$). Our results showed that all strains could form biofilm, which also supports previous conclusions on KEGG enrichment (Figure 8 and Table 9). According to the OD630 threshold of 0.5, 16 *G. parasuis* strains were divided into strong biofilm forming strains and weak biofilm forming strains. The results were classified into the binary category for Pan-GWAS (Figures 8B,C).

Pan-GWAS results showed that 76 genes were associated with *G. parasuis* strong biofilm formation capacity, involved in signal pathways such as metabolism, genetic information processing, and environmental information processing (Figure 9). However, these genes associated with strong biofilm formation that could specifically identify the strong biofilm formation strains were not found. Metabolic pathways include carbohydrate, energy, amino acids, vitamins, etc. Importantly, the metabolism of nitrogen and nitrate may induce the denitrification process of the bacteria inside the biofilm in low dissolved oxygen conditions. Genetic information processing functional pathways focus on the translation process. Environmental information processing includes signal transduction and membrane transport. Furthermore, the functions of virulence factors concentrate on movement, immune modulation, nutritional metabolism, and adherence. For instance, the flagellum, which participates in movement and adherence, plays an essential role in the initial formation of biofilm and the dissipation process of the mature stage. In addition, lipoid A is modified by phosphoethanolamine transferase to reduce repulsion between the individuals *via* decreasing overall net negative charge of the bacterial outer membrane (Samantha and Vrielink, 2020). This way makes bacteria more easily clustered to form biofilm. Copper-transporting ATPases can regulate the copper balance and prevent copper poisoning while maintaining its nutrition (Migocka, 2015).

3.10. Antimicrobial susceptibility profiles

The MICs of six antimicrobial agents tested against 16 strains are summarized in Table 10. It can be seen from the results that 16 strains had varying degrees of resistance to aminoglycoside, β-lactam, tetracycline antibiotics. Overall, 16 strains showed the highest degree

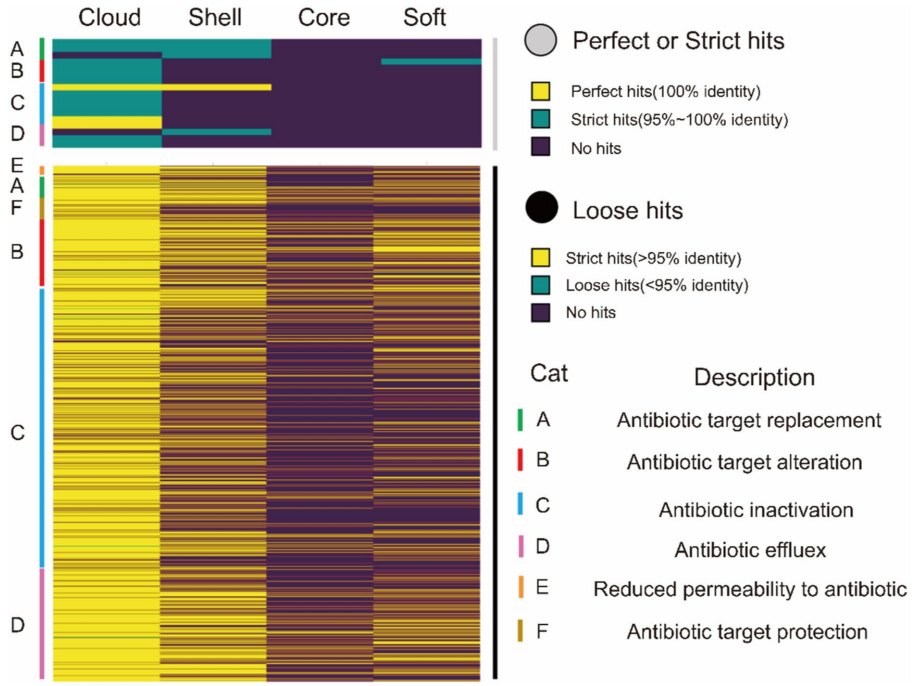


FIGURE 4
CARD annotation of the core genome and the accessory genome. The upper heatmap (gray) is the annotation result obtained using perfect or strict mode, yellow represents 100% agreement, green represents 95~100% agreement, and purple represents no match. The lower heatmap shows the annotation results obtained using loose mode, with >95% agreement in yellow, <95% agreement in green, and no match in purple. The right side of the heat map is marked as A~F, and there are 6 types of drug resistance mechanisms.

of the MIC of aminoglycoside antibiotics. This phenomenon also supported our results of CARD associated genes (Table 7). Namely, a large number of genes that against aminoglycoside antibiotics were found in cloud genes. Additionally, tetracycline also had a certain inhibitory effect on the isolated strain. A small number of genes involved in tetracycline also were found in shell and cloud genes. Of the six antibiotics tested, amoxicillin and ampicillin had the best bacteriostatic effect. And we did not find any genes related to β-lactam class in CARD associated genes. Hence, these results indicated the close association of antibiotic resistance phenotype and the composition of the CARD associated genes.

3.11. Cell adhesion experiments

The results of cell adhesion experiments are shown in Figure 10. In a previous study, we classified 1,5,7,10 serotypes of 16 strains as virulent strains in order to search for genes associated with virulence (Table 2). In cell adhesion experiments, we also used this approach to differentiate between the virulence of the 16 strains. Our results showed that strains of serotype 5 exhibited higher adhesion rates in virulent strains. However, the reasons for the difference in adhesion rates between the different serotype strains were still unknown. Although the GP 07 shows the highest adhesion rate in 16 strains, the adhesion rate between different strains is large in the unknown serum type (Figure 10). Confusingly, we did not find a clear relationship between virulence and adhesion rate. However, *via* differentiating the biofilm-forming capacity of the 16 strains, we seem to find that there may be a positive correlation between the adhesion ability of *G. parasuis* and the ability to form biofilms (Table 9, Figure 10). Namely, the adhesion rates of these strong biofilm-forming strains were higher than those with weak biofilm-forming ability.

4. Discussion

Glaesserella parasuis is rich in genetic diversity, with a strong ability to acquire foreign genes. The phenotype (virulence, biofilm formation) varies greatly between strains. Pan-GWAS is an effective method for studying traits and gene associations (Gori et al., 2020). In this study, *G. parasuis* pan-genome analysis showed that only 390 genes were shared by different individuals, namely core genes. The core gene is a basic skeleton for supporting remainder of the genome, rather than the minimal genome necessary for bacterial survival. If the definition of the core is enlarged (including genes missing only in small parts of the genome), core genomes consist of 1,133 genes (including core genes and soft genes). These genes were present in at least 99% of the sample genomes. The number of core genome also were consistent with previous study (Howell et al., 2014). Our result indicated that *G. parasuis* has opening pan-genome, and the number of core genome did not change significantly with the increase in the number of strains. In contrast, genes in some strains and specific genes only remaining in a single strain constitute an accessory genome, including 7,752 genes.

Core genes not only included all COG classifications, but also majorly focused on basic and necessary biological functions in comparison to accessory genome, such as energy production and conversion, essential substances transport and metabolism, translation

10.3389/fmicb.2023.1160433

TABLE 7 CARD annotation of the core genome and the accessory genome (Perfect or Strict hits).

Class	Gene	Drug class	Resistance mechanism	AMR gene family
Core	NA	NA	NA	NA
Soft-core	*ftsI*	Cephalosporin	Antibiotic target alteration	Penicillin-binding protein mutations
Shell	*dhfrIII*	Diaminopyrimidine	Antibiotic target replacement	Trimethoprim resistant dihydrofolate reductase dfr
	GPS_2460	Diaminopyrimidine	Antibiotic target replacement	Trimethoprim resistant dihydrofolate reductase dfr
	penP	Cephalosporin	Antibiotic inactivation	ROB beta-lactamase
	folP_2	Sulfonamide	Antibiotic target replacement	Sulfonamide resistant dihydropteroate synthase folP
	tetR	Tetracycline	Antibiotic target alteration	Major facilitator superfamily (MFS) antibiotic efflux pump
			Antibiotic efflux	
Cloud	*GPS_8640*	Cephalosporin	Antibiotic target alteration	Penicillin-binding protein mutations
	GPS_8664	Cephalosporin	Antibiotic target alteration	Penicillin-binding protein mutations
	GPS_2847	Peptide	Antibiotic target alteration	MCR phosphoethanolamine transferase
	tetA	Tetracycline	Antibiotic efflux	Major facilitator superfamily (MFS) antibiotic efflux pump
	GPS_4453	Tetracycline	Antibiotic efflux	Major facilitator superfamily (MFS) antibiotic efflux pump
	GPS_4456	Tetracycline	Antibiotic efflux	Major facilitator superfamily (MFS) antibiotic efflux pump
	GPS_4455	Tetracycline	Antibiotic efflux	Major facilitator superfamily (MFS) antibiotic efflux pump
	linA	Lincosamide	Antibiotic inactivation	Lincosamide nucleotidyltransferase (LNU)
	folP_1	Sulfonamide	Antibiotic target replacement	Sulfonamide resistant dihydropteroate synthase folP
	GPS_5385	Sulfonamide	Antibiotic target replacement	Sulfonamide resistant dihydropteroate synthase folP
	GPS_5386	Sulfonamide	Antibiotic target replacement	Sulfonamide resistant dihydropteroate synthase folP
	cat	Phenicol	Antibiotic inactivation	Chloramphenicol acetyltransferase (CAT)
	bcr_2	Phenicol	Antibiotic efflux	Major facilitator superfamily (MFS) antibiotic efflux pump
	GPS_6310	Diaminopyrimidine	Antibiotic target replacement	Trimethoprim resistant dihydrofolate reductase dfr
	dhfrIII_2	Diaminopyrimidine	Antibiotic target replacement	Trimethoprim resistant dihydrofolate reductase dfr
	GPS_8666	Diaminopyrimidine	Antibiotic target replacement	ROB beta-lactamase
	GPS_8892	Diaminopyrimidine	Antibiotic target replacement	ROB beta-lactamase
	GPS_7895	Aminoglycoside	Antibiotic inactivation	APH(6)
	neo	Aminoglycoside	Antibiotic inactivation	APH(3″)
	aacA-aphD	Aminoglycoside	Antibiotic inactivation	APH(2″); AAC(6″)
	GPS_7899	Aminoglycoside	Antibiotic inactivation	APH(3″)
	GPS_7898	Aminoglycoside	Antibiotic inactivation	APH(3″)
	GPS_7894	Aminoglycoside	Antibiotic inactivation	APH(6)
	GPS_7901	Aminoglycoside	Antibiotic inactivation	APH(3″)
	ant1	Aminoglycoside	Antibiotic inactivation	ANT(3″)
	yokD	Aminoglycoside	Antibiotic inactivation	AAC(3)
	GPS_6438	Aminoglycoside	Antibiotic inactivation	APH(3″)
	GPS_6441	Aminoglycoside	Antibiotic inactivation	APH(3″)
	GPS_7896	Aminoglycoside	Antibiotic inactivation	APH(6)
	GPS_7897	Aminoglycoside	Antibiotic inactivation	APH(6)
	ftsI_2	Cephalosporin	Antibiotic target alteration	Penicillin-binding protein mutations
	GPS_8575	Cephalosporin	Antibiotic target alteration	Penicillin-binding protein mutations
	GPS_8640	Cephalosporin	Antibiotic target alteration	Penicillin-binding protein mutations
	GPS_8664	Cephalosporin	Antibiotic target alteration	Penicillin-binding protein mutations

NA stands for missing values.

and so on. Importantly, cytoskeleton was found only in the core genes. Bacterial cytoskeleton structures are a filamentous structure which based on polymers of a single class of protein (Shih and Rothfield, 2006). Bacterial cytoskeleton plays an essential role in cell division, cell polarity, cell shape regulation, plasmid partition, and other functions (Michie and Löwe, 2006). Presumably, the core genes may play a key

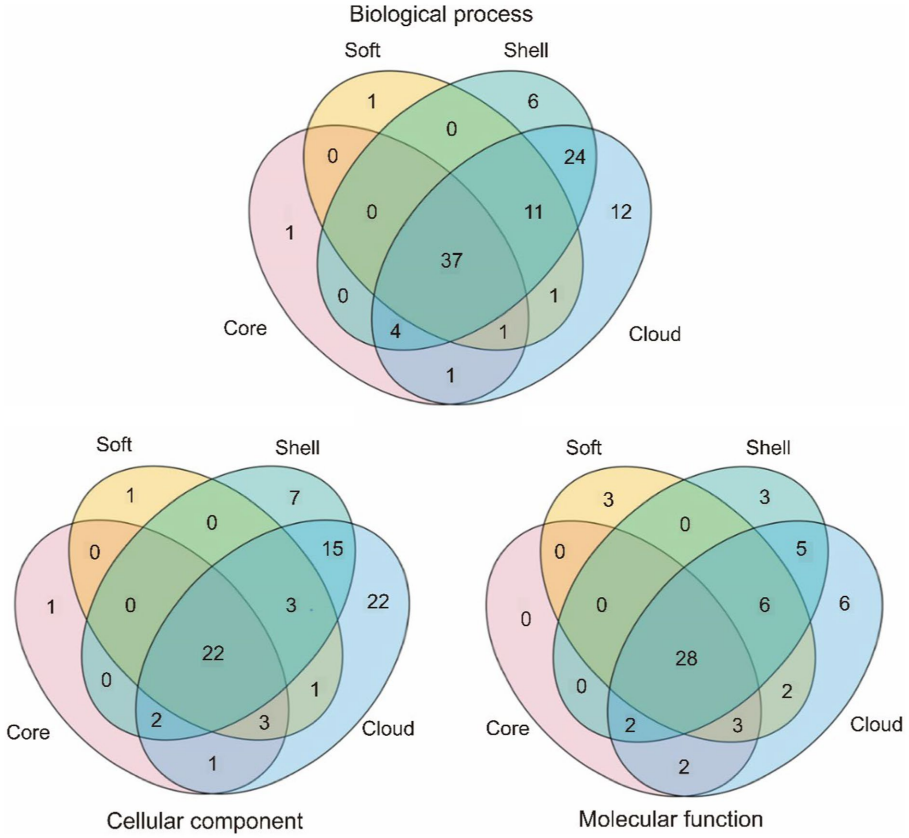


FIGURE 5
Comparison of GO annotation of the core genome and the accessory genome.

role in maintaining shape, multiplication and heterogeneous resistance of bacteria. Moreover, lethal factors were not found in core genes. These cases reflected that the essence of core genes is to control bacterial normal morphology, reproduction and to perform basic biological functions. In addition, the accessory genome also found genes that can participate in *G. parasuis* basic biological process. Presumably, the accessory genome may have an alternative function to maintain part of the biological process of bacterial when damaged core genes.

Pathogenic genes were largely found in accessory genome. These genes participated in bacterial resistance, colonization and invasion. Interestingly, the strains with these genes are mainly isolated from the pericardium and heart blood, rather than the joints and nasal cavity. Our results were consistent with previous studies (Yu et al., 2014; Van et al., 2019). Namely, the avirulent strains mainly adhere to and colonize the upper respiratory tract. Therefore, accessory genome was presumed to be the main source of bacterial pathogenicity.

Interestingly, a proportion of the virulence factors were found in core genes and had various pathogenic mechanisms. Thus, even *G. parasuis* strains considered avirulent are equally potentially pathogenic, and these strains should be included in the prevention and control scope. Virulence factors of the accessary genome major focus on biofilm forming, adherence, immune modulation, invasion, capsule etc. These functions confer the bacteria a greater resistance to the host immune system and contribute to triggering further infection.

Additionally, antimicrobial resistance (AMR) genes were widespread in accessory genomes rather than core genes. The horizontal transfer of AMR genes was the major way bacteria obtain drug resistance in previous studies (Kung et al., 2010; Croll and McDonald, 2012). Presumably, with the increase of genome number, AMR genes will be widespread in *G. parasuis* strains, due to *G. parasuis* had an open pan-genome. Moreover, combined with antibacterial activity test results, it was found that *G. parasuis* has strong resistance to aminoglycoside antibiotics and has a tendency to tolerate tetracycline antibiotics. Presumably, the acquisition and loss of AMR genes did affect the antibiotic phenotype of *G. parasuis* to certain extent. Hence, the misuse of antibiotics to treat *G. parasuis* infections likely contributed to the development and spread of antibiotic resistance in *G. parasuis*.

Although 142 genes were linked to *G. parasuis* hypervirulence traits in this study, the virulence factors that could specifically identify the hypervirulent strains were not found. However, some virulence genes associated with traits were found *via* Pan-GWAS. For instance, the glycosyltransferase that participates in LPS synthesis and modification is essential for adherence and invasion in the pig host. The glycosyltransferase has been shown to inactivate host proteins *via* glycosylation. This process will disturb signal transduction and immune response to achieve immune evasion (DeAngelis, 2002; Zhang et al., 2014). The *IbeA* invasion protein, mediating bacteria in crossing the blood–brain barrier, is essential for invasion (Huang et al., 1995). Colibacillus lacking the

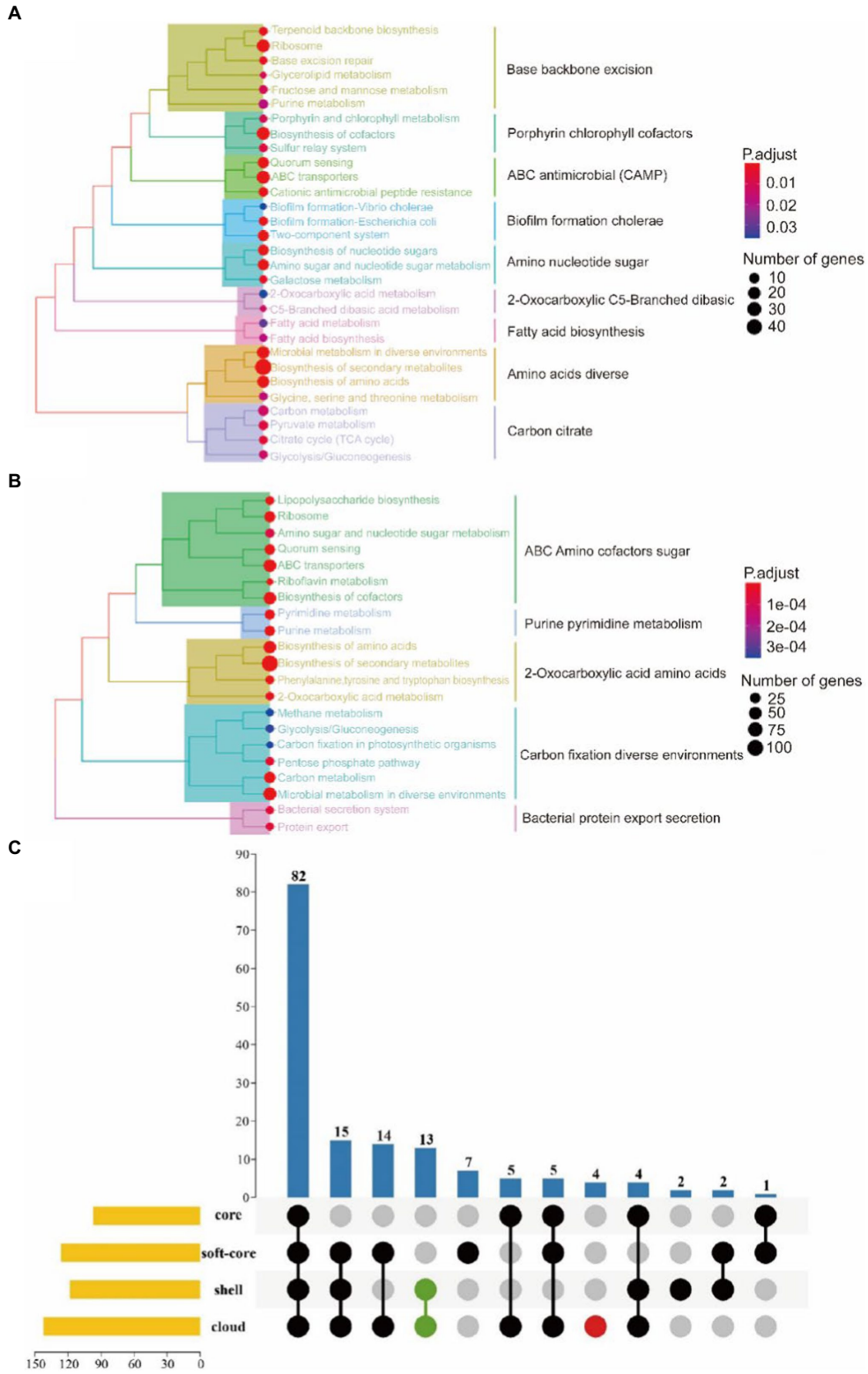


FIGURE 6
KEGG enrichment of the core genome. **(A)** Represents the core gene. **(B)** Represents the soft-core gene. **(C)** Represents KEGG enrichment for the four classes of gene sets. The yellow bar chart on the left represented the number of KEGG metabolic pathways involved in each of the four gene sets. The bar chart at the top and the dot chart at the bottom together explained the number of KEGG metabolic pathways involved in different groups of the four gene sets. Additionally, green and red were used to represent the accessory genomes and cloud genes, respectively.

IbeA protein are unable to invade cerebral microvascular endothelial cells (Huang et al., 2001). In some acute cases, central nervous system symptoms caused by *G. parasuis* infection may be implicated in invasion protein represented by *IbaA*. Moreover, *IbeA* was presumed to locate in the extracellular membrane. And it can be used as a potential vaccine antigen target. Trimeric

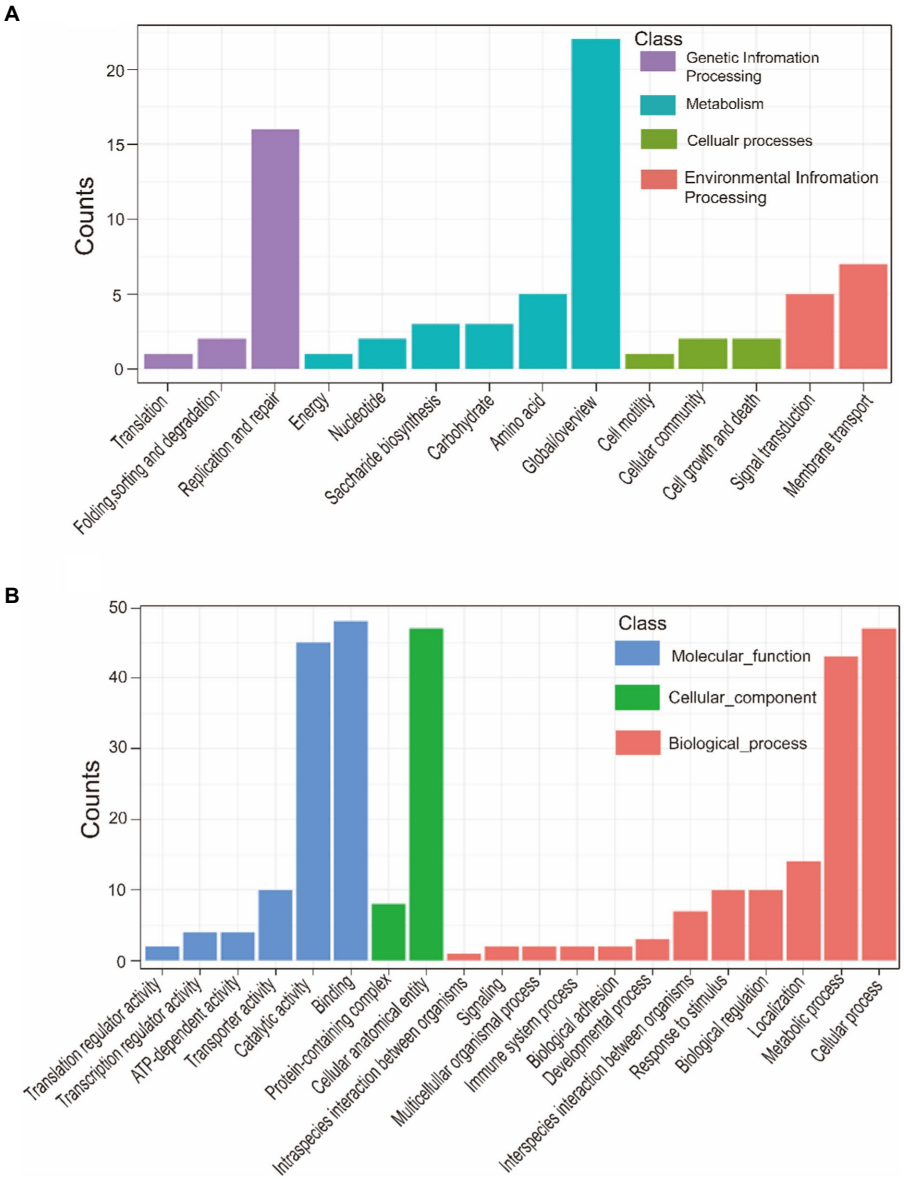


FIGURE 7
KEGG annotation of strong virulence-associated genes **(A)**. GO annotation of strong virulence-associated genes **(B)**.

autotransporters (VtaAs) are only found in virulence strains and play an essential role in adherence and anti-phagocytosis (Laarmann et al., 2002; Meng et al., 2006). Our study also supports this view. Virulence-associated VtaAs can interfere with the phagocytosis of the host leukocytes and promise vaccine candidates. Sialyltransferase (lsgB) is involved in sialic acid utilization (Bouchet et al., 2003). Huan Wang found that *G. parasuis* strains lacking the lsgB are unable to effectively invade porcine iliac artery endothelial cells and porcine kidney epithelial cells due to the decreased autoaggregation ability of bacteria (Wang et al., 2021). Moreover, LOS mediated by lsgB is associated with resistance to the bactericidal effects of complement in the blood (Lewis et al., 2012). These data suggest the difference between avirulent and virulent strains is mainly reflected in entering the host, escaping host defense, bacteria multiplication, and damaging tissues.

Biofilm forming is critical for a successful bacteria survival and infection host. Especially, biofilm forming had been shown to be associated with virulence, antibiotic resistance and genetic typing (Dufour et al., 2010). In this study, although 76 genes were found that are associated with strong biofilm-forming ability, the strains with a strong biofilm-forming phenotype showed differences in the composition of these 76 genes. Moreover, some of the 76 genes also were found in some strains with a weak biofilm-forming phenotype. Presumably, the strain biofilm phenotype is the dynamic result of multiple gene regulation. Even between strains of the same serotype, the ability to form biofilms varies greatly. This may be due to the fact that *G. parasuis* has an open pan-genome, and isolates in different regions differ in the composition of the accessory genome, which in turn affects the biofilm phenotype. In addition, these 76 genes involved in signaling pathways such as metabolism, genetic

TABLE 8 VFDB annotation of strong virulence-associated genes.

Gene	Production	Function
upaG_10	Autotransporter adhesin	Adherence
group_8590	SiM protein	Adherence
elfC	F17 fimbrial uscher	Adherence
metR	Transcriptional regulator	Nutritional/Metabolic-iron acquisition
fhuB	Fe (3+)-hydroxamate ABC transporter permease	Nutritional/Metabolic-iron acquisition
hemY	Heme biosynthesis protein	Nutritional/Metabolic-iron acquisition
hxuC	Heme-hemopexin utilization protein	Nutritional/Metabolic-iron acquisition
group_3905	Transferrin-binding protein 2 precursor	Nutritional/Metabolic-iron acquisition
uhpT	Hexose phosphate transport protein	Nutritional/Metabolic-carbohydrate
uhpC	Hexose phosphate transport protein	Nutritional/Metabolic-glucose
glnA	Type I glutamate-ammonia ligase	Nutritional/Metabolic
oppA_1	Peptide ABC transporter substrate	Nutritional/Metabolic
lacE	ABC transporter substrate	Nutritional/Metabolic
group_5852	Hemolysin A	Exotoxin
group_2488	RTX toxin transporter, ATPase protein	Exotoxin
uup	ABC-type transporter	Exotoxin
prpC	Serine–threonine phosphatase	Regulation
uhpA	Response regulator transcription factor	Regulation-oxidative stress
group_6939	HigA family addiction module antidote protein	Regulation
group_1744	Invasion protein IbeA	Invasion
group_5547	Membrane protein	Immune modulation -capsule
group_7550	Sialyltransferase	Immune modulation-LOS
pglH	Glycosyltransferase family 4 protein	Immune modulation -capsule
group_4165	Hypothetical protein	Immune modulation -capsule
cpsY	Capsular polysaccharide phosphotransferase	Immune modulation -capsule
group_6341	Glycosyltransferase	Immune modulation -capsule
wbbD_1	Glycosyltransferase	Immune modulation -capsule
group_1181	Polysaccharide biosynthesis protein	Immune modulation -capsule
group_4155	Glycosyltransferase family 4 protein	Immune modulation-LPS
group_576	Pyridoxal phosphate	Immune modulation-LPS
arnB	Lipopolysaccharide biosynthesis protein	Immune modulation-LPS
epsL	Probable sugar transferase	Immune modulation-LPS
lpxK	Tetraacyldisaccharide 4′-kinase	Immune modulation-LOS
group_4170	Sialyltransferase	Immune modulation-LOS
lst_2	Sialyltransferase	Immune modulation-LOS
group_158	Hyperosmolarity resistance protein	Immune modulation
rbsB_2	Sugar ABC transporter substrate	Secrete system
tsf	Type IV secretion system effector	Secrete system

information processing, and environmental information processing. Nitrogen and nitrate metabolism associated with denitrification play an important role in anoxia conditions (Richardson and Watmough, 1999). Bacterial located in a hypoxic microenvironment within a biofilm can utilize nitrate as the electron acceptors to complete cellular respiration. Furthermore, complex dynamic and metabolic heterogeneity were found in biofilm in the previous study (Flemming et al., 2016; Karygianni et al., 2020). Anaerobic metabolic processes may have local effects on the microenvironment *via* by-products to maintain biofilm homeostasis. Interestingly, pathways related to ethylbenzene degradation, naphthalene degradation, and degradation of chloroalkene were found in the accessory genome (Rabus and Widdel, 1995). These data showed that certain strains might obtain fitness advantages under specific environmental conditions. For

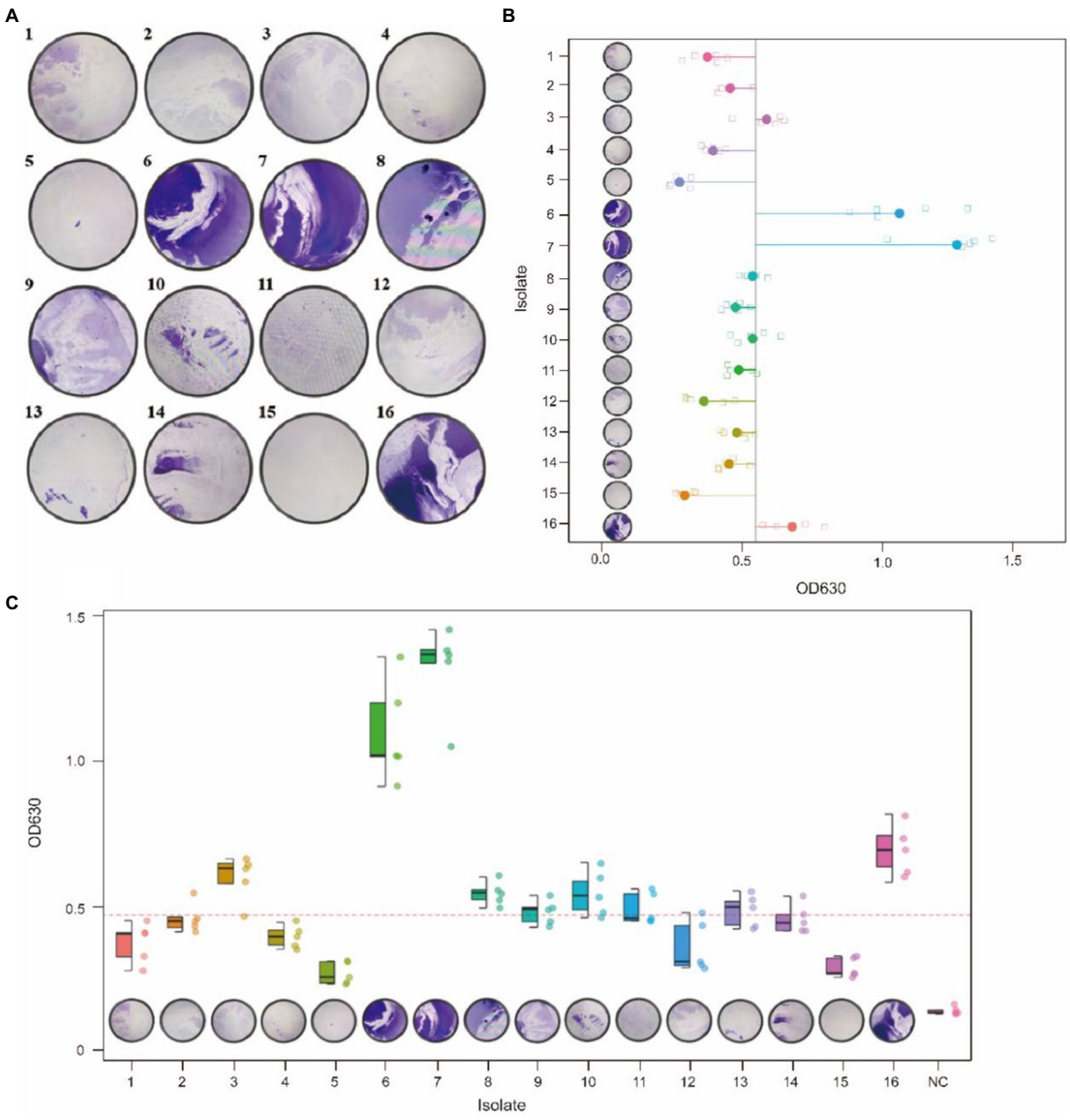


FIGURE 8
Crystal violet staining of biofilms of 16 *G. parasuis* strains **(A)**. One represents strain GP 01, others are analogous; the picture was taken from the biofilm formed on a 9cm×9cm polystyrene petri dish. Comparison of biofilm forming ability of 16 *G. parasuis* strains **(B,C)**. **(B,C)** Described the minimum, 25th percentile, median, 75th percentile, and maximum values of OD 630 for 16 strains using scatter plots and boxplots, respectively. And jitters were added to prevent data overlap.

example, certain *G. parasuis* strains can survive long-term and cause persistent infection in anaerobic environments such as host tissue or within the biofilm (Dar et al., 2021). Lipid A phosphoethanolamine transferase, coded by *opgE_2* gene, can reduce the overall net-negative charge of the outer membrane of some gram-negative bacteria *via* modification of lipid A. This process will confer resistance to polymyxin (Samantha and Vrielink, 2020). Hence, our results also supported an intimate connection between biofilm presence and antibiotic resistance (Mah and O'Toole, 2001). Furthermore, the decrease of net-negative charge is putatively helpful to abate repulsion between bacteria individuals. Bacteria can autoaggregate *via* this way. Autoaggregation and microcolony formation are among the first steps in building a biofilm (Trunk et al., 2018). Thus, lipid A phosphoethanolamine transferase may be a potential molecular target for antimicrobial agents and vaccine design.

Generally, the high-biofilm production phenotype linked to virulence is blurry (Jin et al., 2006; Bello-Ortí et al., 2014). In the present study, virulence and biofilm forming may have potential relation. Our results showed that the same protein products were found in strong virulence association genes and strong biofilm formation association genes, including transferrin-binding protein two and fibrinogen binding M-like protein (SiM protein) (Table 8 and

10.3389/fmicb.2023.1160433

TABLE 9 Comparison of biofilm forming ability of 16 *G. parasuis* strains.

Strains	OD630	Mean±Deviation var	*p*-value	Strains	OD630	Mean±Deviation var	*p*-value
GP 01	0.466	0.3904 ± 0.069	3.92512E-05	GP 09	0.511	0.4944 ± 0.043	8.40868E-08
	0.343				0.443		
	0.296				0.504		
	0.425				0.552		
	0.422				0.462		
GP 02	0.478	0.475 ± 0.052	5.53427E-07	GP 10	0.503	0.5588 ± 0.043	1.32444E-06
	0.442				0.6		
	0.562				0.551		
	0.465				0.664		
	0.428				0.476		
GP 03	0.661	0.6112 ± 0.078	6.82489E-07	GP 11	0.557	0.5072 ± 0.054	3.14836E-07
	0.592				0.464		
	0.483				0.466		
	0.644				0.475		
	0.676				0.574		
GP 04	0.434	0.4114 ± 0.037	2.70049E-07	GP 12	0.306	0.3776 ± 0.087	0.000266419
	0.46				0.313		
	0.383				0.327		
	0.369				0.493		
	0.411				0.449		
GP 05	0.274	0.2866 ± 0.038	5.47184E-05	GP 13	0.451	0.4998 ± 0.054	4.07263E-07
	0.254				0.531		
	0.25				0.567		
	0.329				0.512		
	0.326				0.438		
GP 06	1.361	1.1066 ± 0.175	1.01956E-06	GP 14	0.431	0.4714 ± 0.049	3.77525E-07
	0.921				0.549		
	1.025				0.432		
	1.204				0.487		
	1.022				0.458		
GP 07	1.061	1.3214 ± 0.151	6.947E-08	GP 15	0.345	0.3054 ± 0.033	8.50377E-06
	1.339				0.338		
	1.385				0.273		
	1.369				0.287		
	1.453				0.284		
GP 08	0.561	0.5586 ± 0.04	1.19678E-08	GP 16	0.649	0.7068 ± 0.09	4.54057E-07
	0.615				0.596		
	0.571				0.755		
	0.537				0.828		
	0.509				0.706		

Table 11). Transferrin-binding protein plays an essential role in iron acquisition (Moraes et al., 2009; Calmettes et al., 2012). Pathogenic bacteria can use high-affinity iron uptake systems, such as transferrin-binding proteins, to capture the iron of the host (del Río et al., 2005; Curran et al., 2015). Similarly, transferrin-binding protein expression is beneficial to maintaining bacteria survival under low iron conditions, such as within biofilm. SiM protein is the dominant virulence factor in some streptococcal species. SiM protein can confer

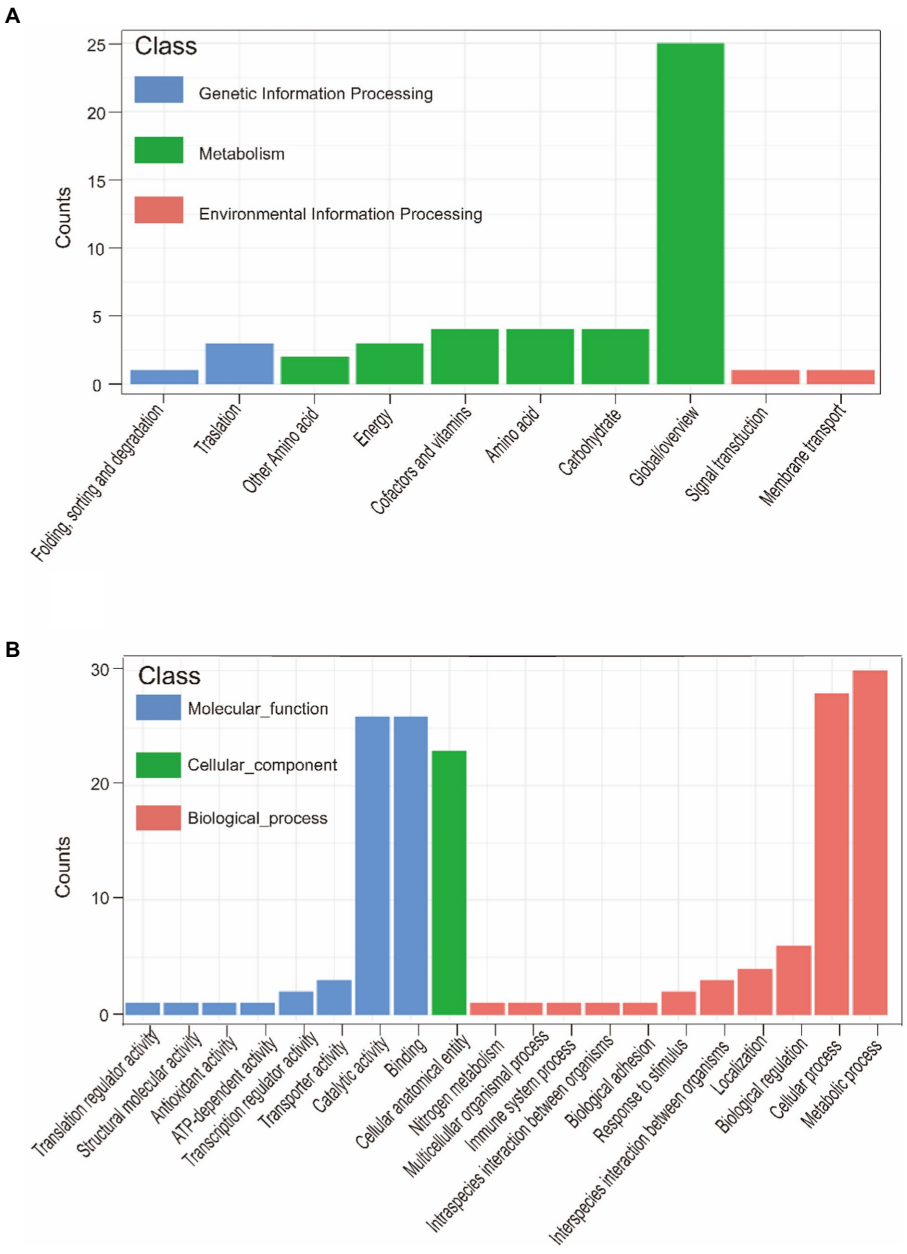


FIGURE 9
KEGG annotation of biofilm formation-associated genes **(A)**. GO annotation of biofilm formation-associated genes **(B)**.

bacterial resistance to phagocytosis *via* binding fibrinogen (Baiano et al., 2008). Expression of SiM protein in virulence strains may block the deposition of complement on bacterial surface. This process will confer bacterial evasion of phagocytosis and multiply within the host (Geyer and Schmidt, 2000). Additionally, SiM proteins were found to play a role in bacterial autoaggregation (Trunk et al., 2018). On the one hand, autoaggregation may represent an additional virulence mechanism. Virulence is enhanced by the formation of aggregates (Bieber et al., 1998). On the other hand, autoaggregation is more beneficial for bacterial adhesion to the tissue surfaces to form the biofilm (Sherlock et al., 2005). Hence, we postulated that these two products may become a marker of the relationship between virulence and biofilm forming.

Generally, some virulence factors play a crucial role in the adhesion and colonization of bacteria. In our study, some virulence factors that favor bacterial adhesion also were found (Table 6). For example, the fimbrial protein, encoded by the *PilE* gene, has been proven to mediate bacterial adherence to mucosal epithelia (Xu et al., 2011). LOS is not only involved in immune regulation, but also acts as an adhesion factor in *G. parasuis*-triggered meningitis (Wang et al., 2021). However, in cell adhesion experiments, we did not observe certain exact relationship between the different serotypes and cell adhesion rates. This may be due to annotation to adherent virulence factors predominantly present in the accessory genome. Even with strains of the same serotype, there are still large differences in genetic composition, which leads to differences in adhesion ability between

TABLE 10 Antibacterial activity test results.

Strains	MIC (μg/mL)					
	Amoxicillin	Ampicillin	Gentamicin	Kanamycin	Streptomycin	Tetracycline
GP 01	0.125	0.125	1	32	4	0.5
GP 02	0.062	0.062	4	32	8	2
GP 03	0.062	0.125	2	8	1	0.5
GP 04	0.062	0.5	2	8	32	16
GP 05	8	4	8	64	64	32
GP 06	0.031	0.031	4	8	0.5	0.125
GP 07	0.031	0.031	2	8	0.5	0.125
GP 08	0.125	0.062	8	16	1	1
GP 09	0.062	0.062	2	16	8	0.5
GP 10	0.062	0.125	4	8	2	0.5
GP 11	0.031	0.125	4	16	1	1
GP 12	0.5	0.5	8	32	16	8
GP 13	0.031	0.031	2	8	8	1
GP 14	1	0.5	1	32	32	16
GP 15	1	1	16	32	64	32
GP 16	0.062	0.062	2	8	0.5	0.5

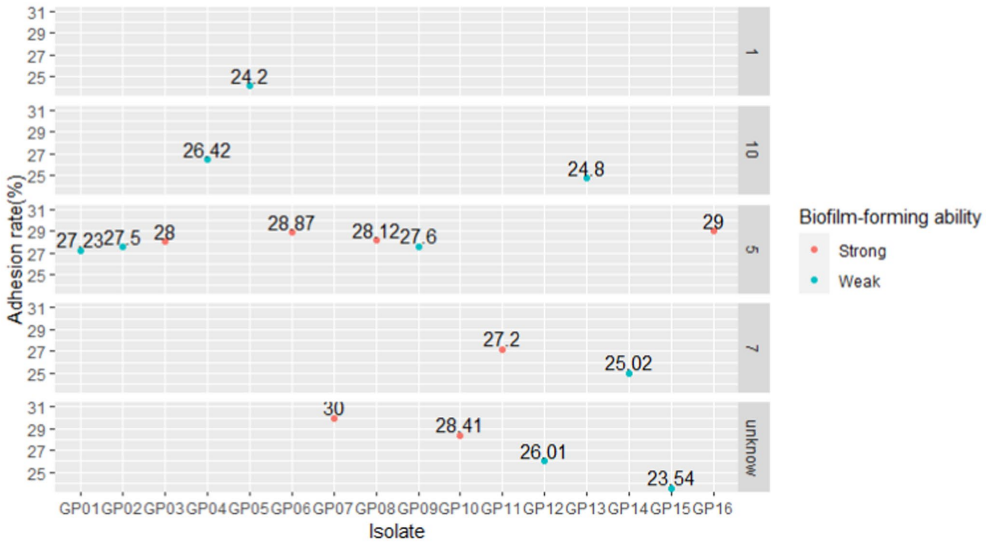


FIGURE 10
Comparison of adhesion rates of different isolates. We faceted the graphs according to the serotypes of the 16 isolates. According to the absorbance of different isolates (Table 9), with an OD630 value threshold of 0.5, it was divided into strong biofilm-forming strains and weak biofilm-forming strains, and the color was distinguished.

strains of the same serotype (Figure 10). Although the biofilm formation process is extremely complex, initial adhesion is necessary for bacteria to form biofilm (Arciola et al., 2018). Our results found that a correlation between adhesion and biofilm. Namely, the cell adhesion ability of the *G. parasuis* can be directly linked to its biofilm-forming ability. However, limited to only 16 strains, this connection still needs to be verified in a large number of samples.

Our analysis has some limitations. First, we were confined to the current publicly available *G. parasuis* genomes retrieved from NCBI (a total of 106 isolates) and 16 genomes from CNSA. Second, our study was limited to the biofilm formation process *in vitro*. In fact, biofilm formation is a complex process engaging various steps and physiologically diverse cellular states. The expression of some genes may differ during the biofilm formation process *in vitro* or vivo. Third, as our analysis is confined to putative protein in the database, further experiments will be needed to verify these supposed genes linked to the traits. Finally, due to the small sample size, experiments using gene-knockout may be useful for

TABLE 11 VFDB annotation of strong biofilm formation-associated genes.

Gene	Production	Function
group_979	RNA polymerase factor	Flagellum mediates movement
group_320	M-related protein Enn	Immune modulation
group_421	PE domain-containing protein	Immune modulation
group_1171	Nitrate reductase	Nutritional/Metabolic
opgE_2	Phosphoethanolamine transferase	Nutritional/Metabolic
yiaJ	DNA-binding transcriptional repressor	Nutritional/Metabolic
group_829	Transferrin-binding protein 2 precursor	Nutritional/Metabolic
copA	Copper-translocating P-type ATPase	Nutritional/Metabolic
bioD	Dethiobiotin synthetase	Nutritional/Metabolic
bioC	Biotin synthesis protein	Nutritional/Metabolic
fucO	Adhesion protein Lap	Adherence
group_279	Surface-exposed protein, autotransporter	Adherence
group_2798	Elongation factor Tu	Adherence
group_4247	SiM protein	Adherence

evaluating the hypothesis that virulence and biofilms may have potential links.

In conclusion, we have shown the characteristic differences in the core genome and the accessory genome of *G. parasuis*. Moreover, we screened out 142 genes with an association with strong virulence. This may establish the basis for the search for new virulence factors. In biofilm formation, multiple genes are involved in prophase adhesion, and autoaggregation processes to influence biofilm formation by regulating metabolism, and may contribute to the pathogenicity of *G. parasuis* pathogenicity. Additionally, we speculated that virulence and biofilms may have potential links. Finally, we assumed a positive correlation between the adhesion rate of *G. parasuis* on the cell and biofilm-forming ability of *G. parasuis*.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Author contributions

YZ, MR, and DJ: conceptualization. YZ and DJ: software. YZ and MR: methodology. YZ, MR, and YY: formal analysis. YZ and GZ: investigation. ZY, YL, and YW: resources. DJ and MR: data curation. YZ and AL: writing—original draft preparation. DJ, ZY, MR, and YW: writing—review and editing. YZ and YW: visualization. MR and YW: supervision. YZ, MR, ZY, and XY: project administration. YL and YW: funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding

This project was supported by the Sichuan Province Science and Technology Planning Program (2021ZDZX0010, 2021YJ0270, 2020YJ0345, and 2021YFSY0005), Provincial Natural Science Foundation of Sichuan (2023NSFSC1216), China Postdoctoral Science Foundation (2022M722300), and the Hong Kong Scholars Program 2022, China Postdoctoral Science Foundation (XJ2022047).

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.

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.2023.1160433/full#supplementary-material

References

Aragon, V., Cerdà-Cuéllar, M., Fraile, L., Mombarg, M., Nofrarías, M., Olvera, A., et al. (2010). Correlation between clinico-pathological outcome and typing of Haemophilus parasuis field strains. *Vet. Microbiol.* 142, 387–393. doi: 10.1016/j.vetmic.2009.10.025

Arciola, C. R., Campoccia, D., and Montanaro, L. (2018). Implant infections: Adhesion, biofilm formation and immune evasion. *Nat. Rev. Microbiol.* 16, 397–409. doi: 10.1038/s41579-018-0019-y

Baiano, J. C., Tumbol, R. A., Umapathy, A., and Barnes, A. C. (2008). Identification and molecular characterisation of a fibrinogen binding protein from *Streptococcus iniae*. *BMC Microbiol.* 8:67. doi: 10.1186/1471-2180-8-67

Baker, J. R., and Pritchard, D. G. (2000). Action pattern and substrate specificity of the hyaluronan lyase from group B streptococci. *Biochem. J.* 348, 465–471. doi: 10.1042/bj3480465

Barh, D., Soares, S. C., Tiwari, S., and Azevedo, V. (2020). *Pan-genomics: Applications, challenges, and future prospects*. Academic Press: Cambridge, MA.

Bello-Ortí, B., Deslandes, V., Tremblay, Y. D., Labrie, J., Howell, K. J., Tucker, A. W., et al. (2014). Biofilm formation by virulent and non-virulent strains of Haemophilus parasuis. *Vet. Res.* 45:104. doi: 10.1186/s13567-014-0104-9

Bentley, S. (2009). Sequencing the species pan-genome. *Nat. Rev. Microbiol.* 7, 258–259. doi: 10.1038/nrmicro2123

Bieber, D., Ramer, S. W., Wu, C. Y., Murray, W. J., Tobe, T., Fernandez, R., et al. (1998). Type IV pili, transient bacterial aggregates, and virulence of enteropathogenic *Escherichia coli*. *Science* 280, 2114–2118. doi: 10.1126/science.280.5372.2114

Bouchet, V., Hood, D. W., Li, J., Brisson, J. R., Randle, G. A., Martin, A., et al. (2003). Host-derived sialic acid is incorporated into *Haemophilus influenzae* lipopolysaccharide and is a major virulence factor in experimental otitis media. *Proc. Natl. Acad. Sci. U. S. A.* 100, 8898–8903. doi: 10.1073/pnas.1432026100

Calmettes, C., Alcantara, J., Yu, R. H., Schryvers, A. B., and Moraes, T. F. (2012). The structural basis of transferrin sequestration by transferrin-binding protein b. *Nat. Struct. Mol. Biol.* 19, 358–360. doi: 10.1038/nsmb.2251

Costa-Hurtado, M., and Aragon, V. (2013). Advances in the quest for virulence factors of Haemophilus parasuis. *Vet. J.* 198, 571–576. doi: 10.1016/j.tvjl.2013.08.027

Costa-Hurtado, M., Ballester, M., Galofre-Mila, N., Darji, A., and Aragon, V. (2012). Vtaa8 and vtaa9 from Haemophilus parasuis delay phagocytosis by alveolar macrophages. *Vet. Res.* 43:57. doi: 10.1186/1297-9716-43-57

Criss, A. K., Katz, B. Z., and Seifert, H. S. (2009). Resistance of neisseria gonorrhoeae to non-oxidative killing by adherent human polymorphonuclear leucocytes. *Cell. Microbiol.* 11, 1074–1087. doi: 10.1111/j.1462-5822.2009.01308.x

Croll, D., and McDonald, B. A. (2012). The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathog.* 8:e1002608. doi: 10.1371/journal.ppat.1002608

Curran, D. M., Adamiak, P. J., Fegan, J. E., Qian, C., Yu, R. H., and Schryvers, A. B. (2015). Sequence and structural diversity of transferrin receptors in gram-negative porcine pathogens. *Vaccine* 33, 5700–5707. doi: 10.1016/j.vaccine.2015.07.097

Dar, D., Dar, N., Cai, L., and Newman, D. K. (2021). Spatial transcriptomics of planktonic and sessile bacterial populations at single-cell resolution. *Science* 373:eabi4882. doi: 10.1126/science.abi4882

DeAngelis, P. L. (2002). Evolution of glycosaminoglycans and their glycosyltransferases: Implications for the extracellular matrices of animals and the capsules of pathogenic bacteria. *Anat. Rec.* 268, 317–326. doi: 10.1002/ar.10163

Degnan, B. A., and Macfarlane, G. T. (1995). Arabinogalactan utilization in continuous cultures of *Bifidobacterium longum*: Effect of co-culture with bacteroides thetaiotaomicron. *Anaerobe* 1, 103–112. doi: 10.1006/anae.1995.1005

del Río, M. L., Gutierrez-Martin, C. B., Rodriguez-Barbosa, J. I., Navas, J., and Rodriguez-Ferri, E. F. (2005). Identification and characterization of the TonB region and its role in transferrin-mediated iron acquisition in Haemophilus parasuis. *FEMS Immunol. Med. Microbiol.* 45, 75–86. doi: 10.1016/j.femsim.2005.02.008

Dufour, D., Leung, V., and Lévesque, C. M. (2010). Bacterial biofilm: Structure, function, and antimicrobial resistance. *Endod. Top.* 22, 2–16. doi: 10.1111/j.1601-1546.2012.00277.x

Faraldo-Gómez, J. D., and Sansom, M. S. (2003). Acquisition of siderophores in gram-negative bacteria. *Nat. Rev. Mol. Cell Biol.* 4, 105–116. doi: 10.1038/nrm1015

Flemming, H. C., Wingender, J., Szewzyk, U., Steinberg, P., Rice, S. A., and Kjelleberg, S. (2016). Biofilms: An emergent form of bacterial life. *Nat. Rev. Microbiol.* 14, 563–575. doi: 10.1038/nrmicro.2016.94

Geyer, A., and Schmidt, K. H. (2000). Genetic organisation of the m protein region in human isolates of group c and g streptococci: Two types of multigene regulator-like (mgrc) regions. *Mol. Gen. Genet.* 262, 965–976. doi: 10.1007/pl00008665

Gori, A., Harrison, O. B., Mlia, E., Nishihara, Y., Chan, J. M., Msefula, J., et al. (2020). Pan-gwas of *Streptococcus agalactiae* highlights lineage-specific genes associated with virulence and niche adaptation. *mBio* 11:e00728-20. doi: 10.1128/mBio.00728-20

Her, H. L., Lin, P. T., and Wu, Y. W. (2021). Pangenomenet: A pan-genome-based network reveals functional modules on antimicrobial resistome for *Escherichia coli* strains. *BMC Bioinform.* 22:548. doi: 10.1186/s12859-021-04459-z

Howell, K. J., Weinert, L. A., Chaudhuri, R. R., Luan, S. L., Peters, S. E., Corander, J., et al. (2014). The use of genome wide association methods to investigate pathogenicity, population structure and serovar in Haemophilus parasuis. *BMC Genomics* 15:1179. doi: 10.1186/1471-2164-15-1179

Huang, S. H., Wan, Z. S., Chen, Y. H., Jong, A. Y., and Kim, K. S. (2001). Further characterization of *Escherichia coli* brain microvascular endothelial cell invasion gene ibeA by deletion, complementation, and protein expression. *J. Infect. Dis.* 183, 1071–1078. doi: 10.1086/319290

Huang, S. H., Wass, C., Fu, Q., Prasadarao, N. V., Stins, M., and Kim, K. S. (1995). *Escherichia coli* invasion of brain microvascular endothelial cells in vitro and in vivo: Molecular cloning and characterization of invasion gene ibe10. *Infect. Immun.* 63, 4470–4475. doi: 10.1128/iai.63.11.4470-4475.1995

Iranzadeh, A., and Mulder, N. J. (2019). *Microbial genomics in sustainable agroecosystems*. Singapore: Springer, 21–38.

Jackson, R. W., Vinatzer, B., Arnold, D. L., Dorus, S., and Murillo, J. (2011). The influence of the accessory genome on bacterial pathogen evolution. *Mob. Genet. Elements* 1, 55–65. doi: 10.4161/mge.1.1.16432

Jiang, C., Ren, J., Zhang, X., Li, C., Hu, Y., Cao, H., et al. (2021). Deletion of the crp gene affects the virulence and the activation of the NF-kappaB and MAPK signaling pathways in PK-15 and iPAM cells derived from G. parasuis serovar 5. *Vet. Microbiol.* 261:109198.

Jiang, R., Xiang, M., Chen, W., Zhang, P., Wu, X., Zhu, G., et al. (2021). Biofilm characteristics and transcriptomic analysis of Haemophilus parasuis. *Vet. Microbiol.* 258:109073. doi: 10.1016/j.vetmic.2021.109073

Jin, H., Zhou, R., Kang, M., Luo, R., Cai, X., and Chen, H. (2006). Biofilm formation by field isolates and reference strains of Haemophilus parasuis. *Vet. Microbiol.* 118, 117–123. doi: 10.1016/j.vetmic.2006.07.009

Kaiser, P. O., Riess, T., Wagner, C. L., Linke, D., Lupas, A. N., Schwarz, H., et al. (2008). The head of bartonella adhesin a is crucial for host cell interaction of *Bartonella henselae*. *Cell. Microbiol.* 10, 2223–2234. doi: 10.1111/j.1462-5822.2008.01201.x

Karygianni, L., Ren, Z., Koo, H., and Thurnheer, T. (2020). Biofilm matrixome: Extracellular components in structured microbial communities. *Trends Microbiol.* 28, 668–681. doi: 10.1016/j.tim.2020.03.016

Kokare, C., Chakraborty, S., Khopade, A., and Mahadik, K. R. (2009). Biofilm: Importance and applications. *Indian J. Biotechnol.* 8, 159–168.

Kung, V. L., Ozer, E. A., and Hauser, A. R. (2010). The accessory genome of pseudomonas aeruginosa. *Microbiol. Mol. Biol. Rev.* 74, 621–641. doi: 10.1128/MMBR.00027-10

Laarmann, S., Cutter, D., Juehne, T., Barenkamp, S. J., and St, G. J. (2002). The *Haemophilus influenzae* hia autotransporter harbours two adhesive pockets that reside in the passenger domain and recognize the same host cell receptor. *Mol. Microbiol.* 46, 731–743. doi: 10.1046/j.1365-2958.2002.03189.x

Lewis, K. (2001). Riddle of biofilm resistance. *Antimicrob. Agents Chemother.* 45, 999–1007. doi: 10.1128/AAC.45.4.999-1007.2001

Lewis, L. A., Carter, M., and Ram, S. (2012). The relative roles of factor h binding protein, neisserial surface protein a, and lipooligosaccharide sialylation in regulation of the alternative pathway of complement on meningococci. *J. Immunol.* 188, 5063–5072. doi: 10.4049/jimmunol.1103748

Lewis, L. A., Ngampasutadol, J., Wallace, R., Reid, J. E., Vogel, U., and Ram, S. (2010). The meningococcal vaccine candidate neisserial surface protein a (nspa) binds to factor h and enhances meningococcal resistance to complement. *PLoS Pathog.* 6:e1001027. doi: 10.1371/journal.ppat.1001027

Li, S., and Jedrzejas, M. J. (2001). Hyaluronan binding and degradation by streptococcus agalactiae hyaluronate lyase. *J. Biol. Chem.* 276, 41407–41416. doi: 10.1074/jbc.M106634200

Li, Y., Kwok, A. H., Jiang, J., Zou, Y., Zheng, F., Chen, P., et al. (2013). Complete genome analysis of a Haemophilus parasuis serovar 12 strain from China. *PLoS One* 8:e68350. doi: 10.1371/journal.pone.0068350

Mah, T. F., and O'Toole, G. A. (2001). Mechanisms of biofilm resistance to antimicrobial agents. *Trends Microbiol.* 9, 34–39. doi: 10.1016/s0966-842x(00)01913-2

Meng, G., Surana, N. K., St, G. J. R., and Waksman, G. (2006). Structure of the outer membrane translocator domain of the *Haemophilus influenzae* hia trimeric autotransporter. *EMBO J.* 25, 2297–2304. doi: 10.1038/sj.emboj.7601132

Michie, K. A., and Löwe, J. (2006). Dynamic filaments of the bacterial cytoskeleton. *Annu. Rev. Biochem.* 75, 467–492. doi: 10.1146/annurev.biochem.75.103004.142452

Migocka, M. (2015). Copper-transporting atpases: The evolutionarily conserved machineries for balancing copper in living systems. *IUBMB Life* 67, 737–745. doi: 10.1002/iub.1437

Moraes, T. F., Yu, R. H., Strynadka, N. C., and Schryvers, A. B. (2009). Insights into the bacterial transferrin receptor: The structure of transferrin-binding protein b from actinobacillus pleuropneumoniae. *Mol. Cell* 35, 523–533. doi: 10.1016/j.molcel.2009.06.029

Nedbalcova, K., Satran, P., Jaglic, Z., Ondriasova, R., and Kucerova, Z. (2006). Haemophilus parasuis and Glässer's disease in pigs: A review. *Vet. Med.* 51, 168–179. doi: 10.17221/5537-VETMED

Okaro, U., Green, R., Mohapatra, S., and Anderson, B. (2019). The trimeric autotransporter adhesin bada is required for in vitro biofilm formation by bartonella henselae. *NPJ Biofilms Microbiomes.* 5:10. doi: 10.1038/s41522-019-0083-8

Oliveira, S., and Pijoan, C. (2004). Haemophilus parasuis: New trends on diagnosis, epidemiology and control. *Vet. Microbiol.* 99, 1–12. doi: 10.1016/j.vetmic.2003.12.001

Preston, A., Mandrell, R. E., Gibson, B. W., and Apicella, M. A. (1996). The lipooligosaccharides of pathogenic gram-negative bacteria. *Crit. Rev. Microbiol.* 22, 139–180. doi: 10.3109/10408419609106458

Rabus, R., and Widdel, F. (1995). Anaerobic degradation of ethylbenzene and other aromatic hydrocarbons by new denitrifying bacteria. *Arch. Microbiol.* 163, 96–103. doi: 10.1007/BF00381782

Rao, V. K., Krasan, G. P., Hendrixson, D. R., Dawid, S., and St, G. J. R. (1999). Molecular determinants of the pathogenesis of disease due to non-typable *Haemophilus influenzae*. *FEMS Microbiol. Rev.* 23, 99–129. doi: 10.1111/j.1574-6976.1999.tb00393.x

Rapp-Gabrielson, V. J., and Gabrielson, D. A. (1992). Prevalence of Haemophilus parasuis serovars among isolates from swine. *Am. J. Vet. Res.* 53, 659–664.

Richardson, D. J., and Watmough, N. J. (1999). Inorganic nitrogen metabolism in bacteria. *Curr. Opin. Chem. Biol.* 3, 207–219. doi: 10.1016/S1367-5931(99)80034-9

Riess, T., Andersson, S. G., Lupas, A., Schaller, M., Schäfer, A., Kyme, P., et al. (2004). Bartonella adhesin a mediates a proangiogenic host cell response. *J. Exp. Med.* 200, 1267–1278. doi: 10.1084/jem.20040500

Rivera-Benítez, J. F., Luz-Armendáriz, J. D. L., Gómez-Núñez, L., Diosdado Vargas, F., Escatell Socci, G., Ramírez-Medina, E., et al. (2021). Swine health: History, challenges and prospects. *Rev. Mex. Cienc. Pecuarias* 12, 149–185. doi: 10.22319/rmcp.v12s3.5879

Samantha, A., and Vrielink, A. (2020). Lipid a phosphoethanolamine transferase: Regulation, structure and immune response. *J. Mol. Biol.* 432, 5184–5196. doi: 10.1016/j.jmb.2020.04.022

Schembri, M. A., Dalsgaard, D., and Klemm, P. (2004). Capsule shields the function of short bacterial adhesins. *J. Bacteriol.* 186, 1249–1257. doi: 10.1128/JB.186.5.1249-1257.2004

Segerman, B. (2012). The genetic integrity of bacterial species: The core genome and the accessory genome, two different stories. *Front. Cell. Infect. Microbiol.* 2:116. doi: 10.3389/fcimb.2012.00116

Sekine, Y., Tanzawa, T., Tanaka, Y., Ishimori, K., and Uchida, T. (2016). Cytoplasmic heme-binding protein (hutx) from vibrio cholerae is an intracellular heme transport protein for the heme-degrading enzyme, hutz. *Biochemistry* 55, 884–893. doi: 10.1021/acs.biochem.5b01273

Sheppard, D. C., and Howell, P. L. (2016). Biofilm exopolysaccharides of pathogenic fungi: Lessons from bacteria. *J. Biol. Chem.* 291, 12529–12537. doi: 10.1074/jbc.R116.720995

Sherlock, O., Vejborg, R. M., and Klemm, P. (2005). The tiba adhesin/invasin from enterotoxigenic escherichia coli is self recognizing and induces bacterial aggregation and biofilm formation. *Infect. Immun.* 73, 1954–1963. doi: 10.1128/IAI.73.4.1954-1963.2005

Sherman, R. M., and Salzberg, S. L. (2020). Pan-genomics in the human genome era. *Nat. Rev. Genet.* 21, 243–254. doi: 10.1038/s41576-020-0210-7

Shih, Y., and Rothfield, L. (2006). The bacterial cytoskeleton. *Microbiol. Mol. Biol. R.* 70, 729–754. doi: 10.1128/MMBR.00017-06

Stohl, E. A., Criss, A. K., and Seifert, H. S. (2005). The transcriptome response of neisseria gonorrhoeae to hydrogen peroxide reveals genes with previously uncharacterized roles in oxidative damage protection. *Mol. Microbiol.* 58, 520–532. doi: 10.1111/j.1365-2958.2005.04839.x

Sun, Q., Yu, X., He, D., Ku, X., Hong, B., Zeng, W., et al. (2022). Investigation and analysis of etiology associated with porcine respiratory disease complex in China from 2017 to 2021. *Front. Vet. Sci.* 9:960033. doi: 10.3389/fvets.2022.960033

Trunk, T., Khalil, H. S., and Leo, J. C. (2018). Bacterial autoaggregation. *AIMS Microbiol.* 4, 140–164. doi: 10.3934/microbiol.2018.1.140

Van, C. N., Thanh, T. V. T., Zou, G., Jia, M., Wang, Q., Zhang, L., et al. (2019). Characterization of serotypes and virulence genes of Haemophilus parasuis isolates from Central Vietnam. *Vet. Microbiol.* 230, 117–122. doi: 10.1016/j.vetmic.2019.02.008

Wang, H., Wei, W., Cao, Q., Xu, M., Chen, Q., Lv, Y., et al. (2021). Sialylated lipooligosaccharide contributes to Glaesserella parasuis penetration of porcine respiratory epithelial barrier. *ACS Infect. Diseas.* 7, 661–671. doi: 10.1021/acsinfecdis.0c00850

Watnick, P., and Kolter, R. (2000). Biofilm, city of microbes. *J. Bacteriol.* 182, 2675–2679. doi: 10.1128/JB.182.10.2675-2679.2000

Xu, Z., Yue, M., Zhou, R., Jin, Q., Fan, Y., Bei, W., et al. (2011). Genomic characterization of Haemophilus parasuis SH0165, a highly virulent strain of serovar 5 prevalent in China. *PLoS One* 6:e19631. doi: 10.1371/journal.pone.0019631

Yu, J., Wu, J., Zhang, Y., du, Y., Peng, J., Chen, L., et al. (2014). Identification of putative virulence-associated genes among Haemophilus parasuis strains and the virulence difference of different serovars. *Microb. Pathog.* 77, 17–23. doi: 10.1016/j.micpath.2014.10.001

Zhang, P., Chomel, B. B., Schau, M. K., Goo, J. S., Droz, S., Kelminson, K. L., et al. (2004). A family of variably expressed outer-membrane proteins (Vomp) mediates adhesion and autoaggregation in *Bartonella quintana*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 13630–13635. doi: 10.1073/pnas.0405284101

Zhang, B., Ku, X., Zhang, X., Zhang, Y., Chen, G., Chen, F., et al. (2019). The AI-2/luxS quorum sensing system affects the growth characteristics, biofilm formation, and virulence of Haemophilus parasuis. *Front. Cell. Infect. Microbiol.* 9:62. doi: 10.3389/fcimb.2019.00062

Zhang, B., Tang, C., Liao, M., and Yue, H. (2014). Update on the pathogenesis of Haemophilus parasuis infection and virulence factors. *Vet. Microbiol.* 168, 1–7. doi: 10.1016/j.vetmic.2013.07.027

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 12 May 2023
DOI 10.3389/fmicb.2023.1160913

OPEN ACCESS

EDITED BY
Ernesto Perez-Rueda,
Universidad Nacional Autónoma de México, Mexico

REVIEWED BY
Yantao Jia,
Chinese Academy of Sciences (CAS), China
Rikky Rai,
University of Allahabad, India

*CORRESPONDENCE
Hongyou Zhou
✉ hongyouzhou2002@aliyun.com.cn
Liqun Zhang
✉ zhanglq@cau.edu.cn

†These authors have contributed equally to this work

RECEIVED 07 February 2023
ACCEPTED 26 April 2023
PUBLISHED 12 May 2023

CITATION
Wei Y, Dong B, Wu X, Zhao M, Wang D, Li N, Zhang Q, Zhang L and Zhou H (2023) *RpoZ* regulates 2,4-DAPG production and quorum sensing system in *Pseudomonas fluorescens* 2P24.
Front. Microbiol. 14:1160913.
doi: 10.3389/fmicb.2023.1160913

RpoZ regulates 2,4-DAPG production and quorum sensing system in *Pseudomonas fluorescens* 2P24

Yarui Wei[1†], Baozhu Dong[1†], Xiaogang Wu[2], Mingmin Zhao[1], Dong Wang[1], Na Li[3], Qian Zhang[4], Liqun Zhang[5*] and Hongyou Zhou[1*]

[1]College of Horticulture and Plant Protection, Inner Mongolia Agricultural University, Hohhot, Inner Mongolia, China, [2]College of Agriculture, Guangxi University, Nanning, China, [3]Erdos Agricultural and Animal Husbandry Technology Promotion Center, Erdos, Inner Mongolia, China, [4]Bayannaoer Agriculture and Animal Husbandry Technology Promotion Center, Bayannaoer, Inner Mongolia, China, [5]Department of Plant Pathology, China Agricultural University, Beijing, China

Introduction: *Pseudomonas fluorescens* 2P24 was isolated from soil of natural decay associated with wheat take-all and it can effectively control soil-borne diseases caused by a variety of plant pathogens. 2,4-diacetylphloroglucinol (2,4-DAPG), is produced by *P. fluorescens* 2P24 and plays an important role in the prevention and control of plant diseases. To understand the resistant mechanism, in this study, we conducted experiments to explore the regulation role of *rpoZ* in the synthesis of the antibiotic 2,4-DAPG and regulation of QS system.

Methods: A random mini-Tn5 mutagenesis procedure was used to screen regulators for *phlA* transcription in stain PM901, which containing a phlA::lacZ transcriptional fusion reporter plasmid. We identified 12 insertion mutants could significantly change *phlA* gene expression. By analyzing the amino acid sequences of the interrupted gene, we obtained a mutant strain Aa4-29 destroyed the *rpoZ* gene, which encodes the omiga subunit. We constructed the plasmid of *rpoZ* mutant (pBBR-△rpoZ) transformed into competent cells of *P. fluorescens* 2P24 by electro-transformation assay. The strains of *P. fluorescens* 2P24/pBBR, 2P24-△rpoZ/pBBR, 2P24-△rpoZ/pBBR-rpoZ were used to evaluate the regulation role of *rpoZ* in 2,4-DAPG production and quorum sensing system.

Results: According to β-galactosidase activity, we found that *rpoZ* positively regulated the expression of *phlA* (a synthesis gene of 2,4-DAPG) and *PcoI* (a synthesis gene of PcoI/PcoR QS signal system) at the transcriptional level. The production of 2,4-DAPG antibiotic and signal molecule AHL was influenced by *rpoZ*. Further, *rpoZ* was involved in regulating *rsmA* expression. *RpoZ* also has a certain regulatory effect on *rpoS* transcription, but no effect on the transcription of *phlF*, *emhABC* and *emhR*. According to the biocontrol assay, *P. fluorescens* 2P24 strains with *rpoZ* showed obvious antagonism ability against the *Rhizoctonia solani* in cotton, while the mutant strain of *rpoZ* lost the biocontrol effect. *RpoZ* had a significant effect on the swimming and biofilm formation in *P. fluorescens* 2P24.

Conclusion: Our data showed that *rpoZ* was an important regulator of QS system, 2,4-DAPG in *P. fluorescens* 2P24. This may imply that *P. fluorescens 2P24* has evolved different regulatory features to adapt to different environmental threats.

KEYWORDS

***Pseudomonas fluorescen* 2P24, *rpoZ*, 2,4-DAPG, quorum sensing, biocontrol**

1. Introduction

Plant growth-promoting rhizobacteria (PGPR) are closely related to plant roots, which can affect plant health and soil fertility. *Pseudomonas fluorescens* 2P24 generally colonizes in the root, inhibits plant soil borne diseases through producing 2,4-diacetylphloroglucinol (2,4-DAPG; Wei et al., 2004). The biosynthetic pathway of 2,4-DAPG has been clarified in several *Pseudomonas* strains (Vincent et al., 1991; Fenton et al., 1992). The 2,4-DAPG locus includes the four biosynthetic genes *phlACBD* that are transcribed as a single operon and is directly involved in the catalytic process of 2,4-DAPG production (Bangera and Thomashow, 1999). Among of them, *PhlA*, *phlC* and *phlB* are required for transacetylation of the monoacetylphloroglucinol (MAPG) precursor to generate DAPG and *phlD* is critical for the biosynthesis of MAPG (Bangera and Thomashow, 1999; Schnider-Keel et al., 2000). The *phlF* gene, which encodes a TetR-family transcriptional regulator, is located upstream of the *phlA* gene and blocks *phlACBD* transcription by binding to the *phlA* promoter region (Schnider-Keel et al., 2000; Li et al., 2018). Whereas, *phlG,* a gene located between the *phlF* and *phlH* genes (upstream of the *phlACBD* biosynthetic operon), mediates the conversion of DAPG to MAPG in *P. fluorescens* 2P24 (Zhao et al., 2020).

Except for the antibiotic production of *P. fluorescens* 2P24, another factor, quorum sensing (QS) regulation, should also pay attention in bacteria. QS system plays an important role in a diverse array of physiological activities, including symbiosis, virulence, competence, conjugation, antibiotic production, swarming, sporulation and biofilm formation (Kievit and Iglewski, 2000; Gonzalez and Keshavan, 2006; Wei and Zhang, 2006; Sakuragi and Kolter, 2007; Waters et al., 2008). The discovery of new regulators of QS system will help to further elucidate the signal transduction mechanism that bacteria survive under various environmental conditions. In *Pseudomonas*, regulatory elements of QS system, the stationary-phase sigma factor RpoS (Bertani and Venturi, 2004), the two-component regulatory system GacS/GacA (Reimmann et al., 1997), the small RNA-binding regulator RsmA (Pessi et al., 2001), the LuxR family member VqsR (Juhas et al., 2004) and the tetrahelical H-T-H superclass member RsaL (Rampioni et al., 2007) were identified. In *P. fluorescens* 2P24, the *GacS-GacA* system controls its target *phlACBD* by inducing four sRNAs (*RsmX*, *RsmX1*, *RsmY*, and *RmZ*) and repressing the levels of another sRNA, *RgsA* (Zhang et al., 2020a). The RsmA and RsmE proteins directly repress the translation of *phlACBD mRNA*, whereas four sRNAs (*RsmX*, *RsmX1*, *RsmY*, and *RsmZ*) depress the translation of *phlACBD* mRNA by sequestering the *RsmA* and *RsmE* proteins, thereby inducing the production of 2,4-DAPG (Zhang et al., 2020a,b). It was also reported that a quorum-sensing locus, *pcoI*/*pcoR*, which is involved in the regulation of root colonization and plant disease-suppressive ability in *P. fluorescens* 2P24, (Yan et al., 2009a).

DNA-dependent RNA polymerase (RNAP) is the central enzyme involved in gene expression and also constitutes a major target for genetic regulation (Darst et al., 1989, 1991; Schultz et al., 1993). The bacterial RNAP core enzyme consists of four subunits: alpha (α), beta (β), beta' (β') and omega (ω) subunit (Lonetto et al., 1992; Zhang et al., 1999). Among of them, the ω subunit encoded by the *rpoZ* gene was proposed to be an integral part of the core RNAP and is not essential for RNAP activity and cell survival, but can assist RNA polymerase assembly, help β 'subunits fold and protect β' subunits (Burgess, 1969; Dove and Hochschild, 1998; Ghosh et al., 2001; Mathew and Chatterji, 2006). *Streptomyces kasugaensis* produces an antibiotic called primathromycin (KSM), an aminoglycoside antibiotic, to control *Pyricularia oryzae* in rice. Kojima et al. showed that the production of primathromycin in the *rpoZ* mutant strain was reduced, and the formation of aerogenic mycelia was blocked, and these phenotypes could be restored to the wild type by the transfer of *rpoZ* complementary plasmid into the mutant strain (Kojima et al., 2002). This is indicating that *rpoZ* in *S. primaviae* regulated the production of primavithromycin. Transcription analysis of KSM synthesis genes showed that the expression of *kasT* (a specific transcriptional activator synthesized by KSM) was significantly decreased in *rpoZ* mutants, and the expression of *kasT* may require the participation of *rpoZ* or RNAP (including ω subunits). When the *rpoZ* gene of *Mycobacterium smegmatis* was mutated, the colony morphology of the *rpoZ* mutant strain was changed, and swimming ability, the biofilm formation, the strain cell growth were affected as well (Mathew et al., 2006).

As we have discussed above, *P. fluorescens* and its secondary metabolites play a very important role in biocontrol strategies. In this study, we identified a regulator of *rpoZ* in *P. fluorescens* 2P24, which is similar to *rpoZ* of several bacteria. The results showed that *rpoZ* regulates several genes including *phlA*, *pcoI*, *rpoS*, which indicated it might be an important upstream regulator of QS in *P. fluorescens* 2P24. We also found that strains with *rpoZ* showed obvious antagonism ability against the *Rhizoctonia solani* in cotton, and had significant effect on the swimming and biofilm formation in *P. fluorescens* 2P24.

2. Materials and methods

2.1. Bacterial strains and growth condition

Bacterial strains and plasmids used in this study are listed in Supplementary Table S1. *Escherichia coli* and *P. fluorescens* were cultured as described previously in (Wu et al., 2012; Zhao et al., 2020). *E. coli* was grown in Lysogenic broth (LB) medium at 37°C and *P. fluorescens* strains were grown at 28°C in LB medium， KB (King's B medium; King et al., 1954) or ABM medium (Chilton et al., 1974).

2.2. Construction of *rpoZ* mutant and complementation strain

According to the flanking sequence of *rpoZ* gene of strain *P. fluorescens* 2P24, two pairs of primers, rpoZ 29,729/rpoZ 30,442 and rpoZ 30,518/rpoZ 31,424, were designed to amplify *rpoZ* gene (Supplementary Table S2). Using the genome of wild bacterium 2P24 as template, the upstream and downstream *rpoZ* genes were amplified with the length of 907 bp and 714 bp, respectively. The PCR products were treated with the restriction enzymes, *EcoR* I/*Kpn* I and *Kpn*I/*Hind* III respectively, and then cloned into the vector pBLR digested with the corresponding *EcoR*I/*Hind* III restriction enzyme to obtain the suicide vector pBLR- △rpoZ. The construct was verified by diagnostic PCR by primer pair of G1/G2 and Ga/Gd using 2P24 genomic DNA and plasmid p299△G. A 378 bp fragment was lost in p299△G comparing with the wild type *rpoZ* gene according to the confirmation PCR and sequencing.

The plasmids of pBLR-△rpoZ were transformed into competent cells of *P. fluorescens* 2P24 by electro-transformation assay under screening of Km resistance in LB liquid medium. After 7 generations, bacteria grown on ABM medium containing Km and X-gal under the condition of 28°C for 24 h. The white clones were verified by PCR amplification.

The *rpoZ* gene was cloned into the shuttle vector pBBR, which was used for complementary experiments. Using the genome DNA of *P. fluorescens* 2P24 as template, primers rpoZ 30,059/rpoZ 30,712 were used to obtain the fragment of *rpoZ* gene. The fragment was digested by *Hind* III - *Kpn*I and connected with pBBR to obtain the complementary vector, named as plasmid pBBR-rpoZ.

In order to amplify the complete rpoZ gene, PCR primers rpoZ 29,729/rpoZ 30,442 and rpoZ 30,518/rpoZ 31,424 were designed based on the gene sequence of *Pseudomonas fluorescens* 2P24. Using *P. fluorescens* 2P24 genome as a template, two amplified fragments were digested by EcoRI/KpnI and KpnI/HindIII restriction endonuclease enzyme, respectively, and then linked to the corresponding enzyme digested vector pBLR, and a suicidal deletion vector pBLR-△rpoZ was obtained. Using PCR primer rpoZ 30,059/rpoZ 30,712 and wild bacterium 2P24 as template, the 654 bp target fragment containing the complete rpoZ gene was amplified and then connected to the shuttle vector pBBR after being digested by Hind-III/Kpn I restriction enzyme. The complementary vector pBBR-rpoZ was obtained. The deletion vector and complementary vector were transferred into wild-type 2P24 strain to obtain the deletion mutant strain and complementary strain of rpoZ (Supplementary Figures S1A,B).

Each strain was cultured in ABM and LB liquid medium overnight, and the culture concentration was adjusted to the same OD_{600} = 0.8 with ABM and LB liquid medium, and then transferred to 40 ml ABM and LB liquid medium containing corresponding Amp and Gm antibiotics at a ratio of 1:1000, respectively. Then, placed in a 28°C incubator for shaking with120 r/min, and collected samples and measured OD_{600} every 3 h. Each sample was repeated 3 times.

2.3. Detection of QS signal molecule (AHL) in *Pseudomonas fluorescens* 2P24 and its derived strains

Strains of *P. fluorescens* 2P24/pBBR, 2P24-△rpoZ/pBBR, 2P24-△rpoZ/pBBR-rpoZ were inoculated into 5 ml LB medium containing Amp and Gm antibiotics, placed in a shaker with 120 r/min for 36 h at 28°C. 30 μl was inoculated into 30 ml LB medium containing Amp and Gm antibiotics and placed in a shaker for 120 r/min at 28°C. 800 μl of each bacterial solution to be tested was added into the same volume of ethyl acetate, and signals were extracted by extraction method. The organic phase was air-dried and dissolved in 100 μl methanol, diluted 10 times, and stored at −20°C. The reported strain *A. tumefaciens* NTL4 (pZLR4) was inoculated into ABM liquid medium, incubated in a shaker with 120 r/min for 24 h at 28°C, and stored at 4°C. 200 μl newly cultured reporter bacteria (*A. tumefaciens* NTL4) was measured after being cultured with 5 μl of each signal molecule for 4 h. Reporter bacteria added with 5 μl methanol were used as the control.

The QS signals were detected by β-galactosidase activities as described previously in (Miller, 1972). All experiments were performed in triplicate (Wu et al., 2012). Each strain was set 3 replicates.

2.4. Determination of 2,4 DAPG production

Quantification of 2,4-DAPG was done as described previously (Shanahan et al., 1992). *P. fluorescens* 2P24/pBBR, 2P24-△rpoZ/pBBR, 2P24-△rpoZ/pBBR-rpoZ was inoculated into 40 ml King's B liquid medium and incubated in a shaker at 28°C with 120 r/min for 40 h to stationary stage. After centrifugation at 8,000 r/min for 10 min, the supernatant was taken out and acidified to pH 2.0 with 1 mol/l HC1. Equal volume of ethyl acetate was added for extraction, and the organic phase was extracted by rotary evaporation. The dry matter was dissolved with 50 ml methanol and determined by HPLC (UV2002 high performance Liquid chromatograph). C_{18} reverse column was used for HPLC: diameter: 150 × 4.6 mm; Detection wavelength: 270 nm; sample volume: 5 μl; mobile phase: water: ethyleye (V: V) = 45: 55, 0.1% H3P04; Flow rate: l.0 ml/min; retention time: 5.12 min. 2,4-DAPG chromatographic standard sample was purchased from Toronoto Research Chemicals Inc., (D365500).

2.5. Antagonism test of *rpoZ* in *Pseudomonas fluorescens* 2P24 against *Rhizoctonia solani*

Cultivation of *Rhizoctonia solani* and dual-culture confrontation assay was performed on PDA medium (Zhao et al., 2020). The fungus disk of strains with a diameter of 6 mm was placed in the center of the PDA plate, and the tested biocontrol bacteria were inoculated 2.5 cm away from the fungus disk. The plates were placed in an incubator at 25°C. The size of the inhibition zone was measured when the control fungus grew to the edge of the petri dish. Three replicates per treatment.

2.6. Motility tests

Motility tests were conducted as described by Rashid and Kornberg (2000), with slight modification (Wang et al., 2021). For swimming tests, we used water medium that contained 0.2% agar. Freshly cultured strains were dipped using pipette tips and inoculated on the surface of the center of plates. Then, plates were placed stably in the incubator and cultured at 28°C. After 24 h, the diameter of motility of clones was measured.

2.7. The biofilm formation of *Pseudomonas fluorescens* 2P24 and its derived strains

To test whether *rpoZ* affect the biofilm formation of *P. fluorescens* 2P24, the quantitative determination of biofilms is performed as described (Wei and Zhang, 2006) with slight modification. The strain of *ropZ* mutant was inoculated into LB liquid medium and cultured until saturated, then diluted into fresh LB liquid medium at the volume ratio of 1:1000. The formation of bioflim at the junction of solid and liquid was measured after 500 μl was added into a 2 ml centrifugation tube and incubated at 28°C for 24 h. Add 100 μl crystal violet with a concentration of 0.1% (W/V) to each tube. After 20–30 min staining at room temperature, rinse the centrifuge tube with strong distilled water. In the inside of the centrifuge tube, it can

be observed to form a strong biofilm at the junction of the liquid level and the tube wall. The crystal violet combined with biofilm was fully dissolved by adding 1,200 μl of 95% ethanol, and the absorption value of OD_{570} was measured with 1,000 μl.

2.8. Statistical analysis

GraphPad Prism software version 5.01 (Graphpad Software, Inc.) was used for analysis of variance, followed by multiple comparisons using one way ANOVA, and $p < 0.05$ was considered statistically significant.

3. Results

3.1. Identification and characteristic of *rpoZ* gene in *Pseudomonas fluorescens* 2P24

The random insertion method of Tn5 transposon was used to mutate the transcriptional fusion plasmid pGm-phlA contained in *P. fluorecens* 2P24 and screened the regulatory factors affecting the production of 2,4-DAPG. We identified 21 mutant strains that significantly reduced *phlA* expression (Supplementary Table S3). Because of Tn5 transposon is carrying Gm antibiotic resistance gene, the genes damaged by Tn5 transposons in mutant strains could be determined by analysis of the flanking sequence of Tn5. Restriction enzymes (*Sal*I/*Eco*RI) were used for enzyme digestion and self-ligated. Plasmids were transformed into *E. coli* DH5α and inoculated on LB solid medium containing 30 μg/ml antibiotics Gm. The single colony containing the flanking sequence of Tn5 transposon was selected and examined by sequencing. The mutant Aa4-29 with the most reduced *phlA* gene expression was purified and further investigated. Sequence analysis showed that Tn5 was inserted into *rpoZ* gene of mutant strain Aa4-29, which was named as 2P24-△rpoZ (Figure 1A). We constructed the deletion mutant and complementary strain of *rpoZ* as shown in Supplementary Figures S1A,B.

This gene is the ω-subunit of the encoding RNA polymerase, which encodes the synthesis of a 90 amino acid peptide chain with a molecular weight of about 10,105 Da. The RpoZ protein sequence in strain 2P24 is very similar to that of other RpoZ in *Pseudomonas*, among which has a similarity of 90% with the RpoZ sequence in *Pseudomonas brassicacearum* NFM421, has a similarity of 93% with the RpoZ sequence of *Pseudomonas fluorescens* PF0-1 and has a similarity of 86% with the RpoZ sequence of *Pseudomonas syringae* pv.*tomato* str.DC3000.

Analysis of the laterals of the Tn5 transposon revealed that there was a *Gmk* gene in the same transcriptional direction upstream of the *rpoZ* gene, which reportedly encodes the guanylate kinase (Supplementary Figure S1A). The downstream of the *rpoZ* gene is *spoT* gene, whose transcription direction is the same as *rpoZ*. *SpoT* gene encodes pyrophosphatase and is involved in regulating (p)ppGpp level in cells. The sequences of these three genes are also much conserved to upstream and downstream genes of *rpoZ* in *E. coli* and *Streptomyces cerulosus* (Santos-Beneit et al., 2011).

To test the effect of *rpoZ* mutant strain on the growth of *P. fluoresens* 2P24, we compared the wild-type strain *P. fluoresens* 2P24, rpoZ-deficient mutant strain △rpoZ/pBBR and complementary strain △rpoZ/PBBR-rpoZ in LB and ABM medium. The growth curve was drawn according to OD_{600} value at each time point (Figure 1C). The results showed that the growth curves of *P. fluorescens* 2P24/pBBR, △rpoZ/pBBR and △rpoZ/PBBR-rpoZ in LB and ABA medium were consistent (Figure 1B). The growth rate of the *rpoZ* mutant strain was significantly slower than that of the wild-type strain. This growth of the complementary strain carrying the rpoZ plasmid was recovered. These results indicated that the *rpoZ* gene played a role in the regulation of bacterial growth.

3.2. *RpoZ* regulates transcription of signal synthesis *pcoI* gene and production of signal molecule AHL

A promoter-free *lacZ* gene into the *pcoI* gene of the QS system on the genome of wild-type strain 2P24 was inserted and constructed pcoI::lacZ as a marker gene for transcription fusion previously by (Yan et al., 2009b). The transcription of *lacZ* was driven by the *pcoI* gene promoter, and its schematic structure was shown in Figure 2A, by which we determined the level of *pcoI* gene transcription by detecting the β-galactosidase (LacZ) activity of the strain.

The β-galactosidase activity of wild-type strain of *P. fluorescens* 2P24/pBBR and *rpoZ* gene mutant △rpoZ/pBBR was measured and the growth curve was drawn by the value of OD_{600} in the culture medium (Figure 2B). The growth curve showed that the mutant *rpoZ* gene delayed the growth of the bacteria compared with the wild-type strain. The β-galactosidase activity of △rpoZ/pBBR was significantly lower than that of wild-type strain 2P24/pBBR during the whole growth process, which indicated that the transcription activity of *pcoI* gene was greatly reduced after *rpoZ* gene mutation. We suggested that *rpoZ* positively regulated the expression of *pcoI* gene in strain *P. fluoresens* 2P24 at the transcriptional level.

In *P. fluoresens* 2P24, the *pcoI* gene is a synthase that is responsible for synthesizing quorum sensing signalling molecules. We also confirmed *rpoZ* gene also affects the synthesis of signal molecules in the QS system through regulation of *pcoI* expression. The signal molecules produced by the wild strains and the derived strains were extracted and interacted with the reported strain *A. tumefaciens* NTL4 (pZLR4). NTL4 is a modified engineering strain that cannot generate signal molecule itself, but it carries TraG::LacZ fusion gene to detect exogenous QS signal molecule and promote the expression of LacZ (Chai et al., 2001). Therefore, the LacZ activity level of strain *A. tumefaciens* NTL4 can be reported to compare the content of signal molecules in each sample to be tested.

Compared with the wild strain 2P24, the β-galactosidase activity of the reported strain was significantly decreased by the extract of the signal molecule of △rpoZ/pBBR, which meant that the production of signal molecule of △rpoZ/pBBR was smaller than that of wild type 2P24. And this change can be restored by the intact *rpoZ* gene carried by the plasmid PBBR-rpoZ. Therefore, we believed that *rpoZ* gene positively regulated the synthesis of signal molecules in *P. fluoresens* 2P24. This was consistent with the positive regulation of *pcoI* gene transcription by *rpoZ*, which was proved that *rpoZ* played a positive role in quorum sensing system.

In addition, the signal molecules produced by wild-type strain 2P24 have obvious characteristics on different culture time points (Figure 2C). In the early stage of culture, the synthesis amount of signal molecules was very low. However, with the increase of culture time, the

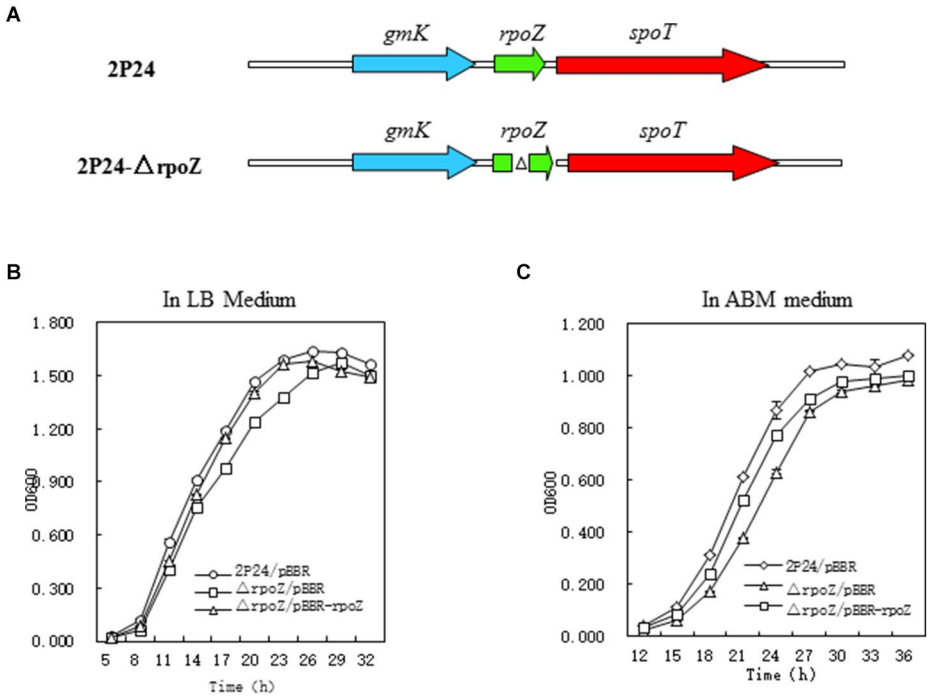


FIGURE 1
Structure and growth curve of rpoZ gene mutation. **(A)** Schematic diagram of the construction of the *ropZ* deletion mutant in *Pseudomonas fluorescens* 2P24. **(B,C)** Growth rate of wild-type (WT) *P. fluorescens* 2P24 and the *ropZ* deletion mutant in LB and ABM liquid medium. The error bars represent standard deviations and the statistical analysis was performed using a two-tailed *t*-test. Three replicates per treatment.

accumulation of signal molecules reached the peak at 24 h, and after the highest accumulation maintained for a period of time, the content of signal molecules in the culture medium began to decline again.

3.3. Effects of *rpoZ* gene on transcription of antibiotic synthesis gene *phlA* and 2,4-DAPG production

It was reported that extracellular secondary metabolite 2,4-DAPG from *P. fluoresens* 2P24 played an important role in controlling effects on soil borne disease in wheat (Zhou et al., 2005). The 2,4-DAPG production was regulated by antibiotic synthesis gene *phlA*. To investigate whether *rpoZ* gene ws involved in transcription of *phlA* and 2,4-DAPG production, the activity of β-galactosidase in PhlA-Lacz transcription fusion was detected in wild-type strain 2P24 and △rpoZ/pBBR. The β-galactosidase activity of each strain was measured and the growth curve was drawn by the value of OD_{600}. The results showed that the growth of the bacteria containing the mutant *rpoZ* gene significantly delayed compared with that of the wild-type strain (Figure 3A). The expression of *phlA* gene in △rpoZ/pBBR was significantly lower than that of wild-type strain 2P24/pBBR during the whole growth process, which indicated that the transcription activity of *phlA* gene was greatly reduced in △rpoZ/pBBR. This was suggested that *rpoZ* positively regulated the expression of *phlA* gene in strain *P. fluoresens* 2P24 at the transcriptional level.

Correspondingly, HPLC was used to detect whether the yield of 2,4-DAPG was consistent with the positive regulatory effect of *rpoZ* gene on the expression of *phlA* gene. The 2,4-DAPG production of △rpoZ/pBBR was minimal compared with that of wild strain 2P24, and this change was recovered by the intact *rpoZ* gene carried by the plasmid △rpoZ/pBBR (Figure 3B). Therefore, we believed that *rpoZ* gene in *P. fluoresens* 2P24 has an effect on the production of 2,4-DAPG antibiotics.

3.4. *RpoZ* affects 2,4 DAPG production through positive regulation of *rsmA*

In *P. fluorescens* 2P24, the RsmA and RsmE proteins directly repress the translation of phlACBD mRNA, whereas four sRNAs RsmX, RsmX1, RsmY, and RsmZ derepress the translation of phlACBD by sequestering the RsmA and RsmE proteins, thereby inducing the production of 2,4-DAPG (Zhang et al., 2020b). To examine how influence of *rpoZ* on those genes of *rsmA*,*rsmE*, three sRNAs *rsmX*, *rsmY*, and *rsmZ*, the different transcription LacZ fusion reporter plasmids of small RNA molecules (rsmX-lacZ, rsmY-LacZ, rsmZ-lacZ) were constructed and electrocuted into wild bacteria *P. fluorescens* 2P24/pBBR and △rpoZ/pBBR. The β-galactosidase activity of each bacterium was measured and profiled as shown in Figure 4.

The results showed that *rpoZ* gene has no effect on the transcription expression of *rsmE* gene, but *rpoZ* gene has a significant effect on the transcription expression of *rsmA* (about 4 times) (Figures 4A,B). *rpoZ* gene mutation had no effect on the transcription of *rsmX*, *rsmY* and *rsmZ* genes, which indicated that *rpoZ* was not involved in the expression of r*smX*, *rsmY* and *rsmZ* genes at the transcription level in strain 2P24 (Figures 4C–E). *RsmX*, *rsmY* and *rsmZ* genes are regulated by the GacS/GacA two-factor regulatory

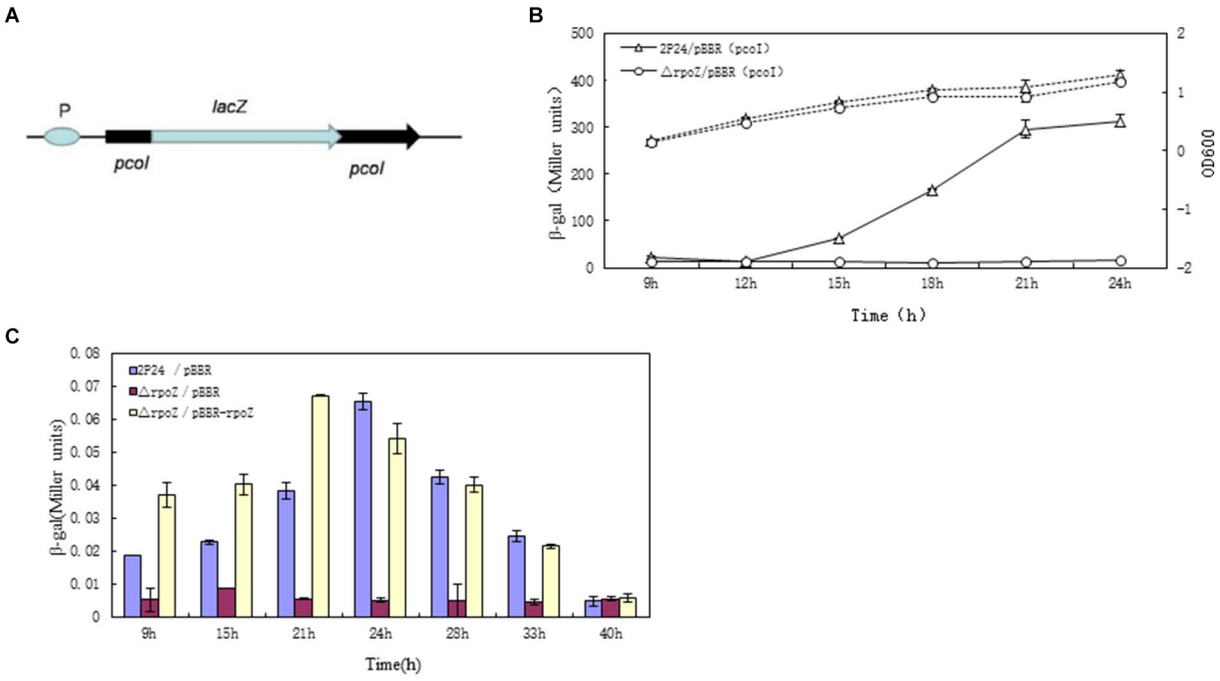


FIGURE 2
RopZ regulates *pcoI* expression in *P. fluorescens* 2P24. **(A)** The graphic presentation of constructed *E. coli* reporter plasmids. pSB-pPcoI contained the promoter region of *pcoI* fused to the promoterless *lacZ* gene. **(B)** The β-galactosidase activity of wild type 2P24/pBBR (pcoI) and ropZ mutant △rpoZ/pBBR (pcoI) was detected and shown as the solid lines with Miller units. The dotted lines were growth curve. **(C)** Detection of the signal accumulation of ropZ mutants in *A. tumefaciens* NTL4 (pZLR4). The signal molecules were extracted from wild type 2P24/pBBR, ropZ mutant △rpoZ/pBBR and complementary strain △rpoZ/pBBR-rpoZ cultures and incubated with *A. tumefaciens* NTL4 (pZLR4). Then, the β-gal activities were detected at different time points as shown in horizontal coordinate. The ordinate represents the ratio of β-galactosidase activity to OD_{600} value of each detected strain. Each value was calculated by 3 replicates. The values were from at least three independent assays. Three independent experiments were performed and the error bars were calculated standard deviations of experimental data.

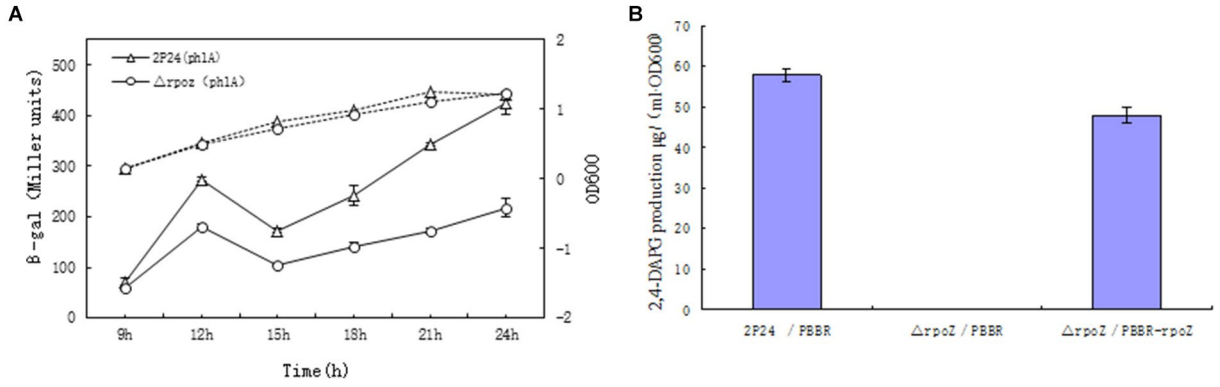


FIGURE 3
RopZ regulates *phlA* expression and 2,4-DAPG production in *P. fluorescens* 2P24. **(A)** β-galactosidase activity of wild type 2P24 and △rpoZ and ropZ mutant was detected and shown. **(B)** 2,4-DAPG production was detected in from wild type 2P24/pBBR, ropZ mutant △rpoZ/pBBR and complementary strain △rpoZ/pBBR-rpoZ.

system, which positively regulates the expression of small RNA molecules. *RpoZ* gene did not regulate the transcription of *rsmX*, *rsmY* and *rsmZ* genes. This was indicating that the effect of *rpoZ* gene on antibiotic 2,4-DAPG production was not regulated by the expression of *rsmX*, *rsmY* and *rsmZ* gene at transcription level, but was regulated by the *rsmA*, which directly repressing the translation of *phlACBD* mRNA and inducing the reduction the production of 2,4-DAPG.

3.5. *RpoZ* regulates transcription of *rpoS* gene

PhlF is an inhibitor of 2,4-DAPG synthesis. By binding to the operon of PhO, PhlF protein can inhibit the binding of its RNA synthase and the region of the *phlA* gene promoter, thus inhibiting the transcription of *phlACBD* gene (Schnider-Keel et al., 2000; Li et al., 2018). However, with the growth of bacteria, a small amount of

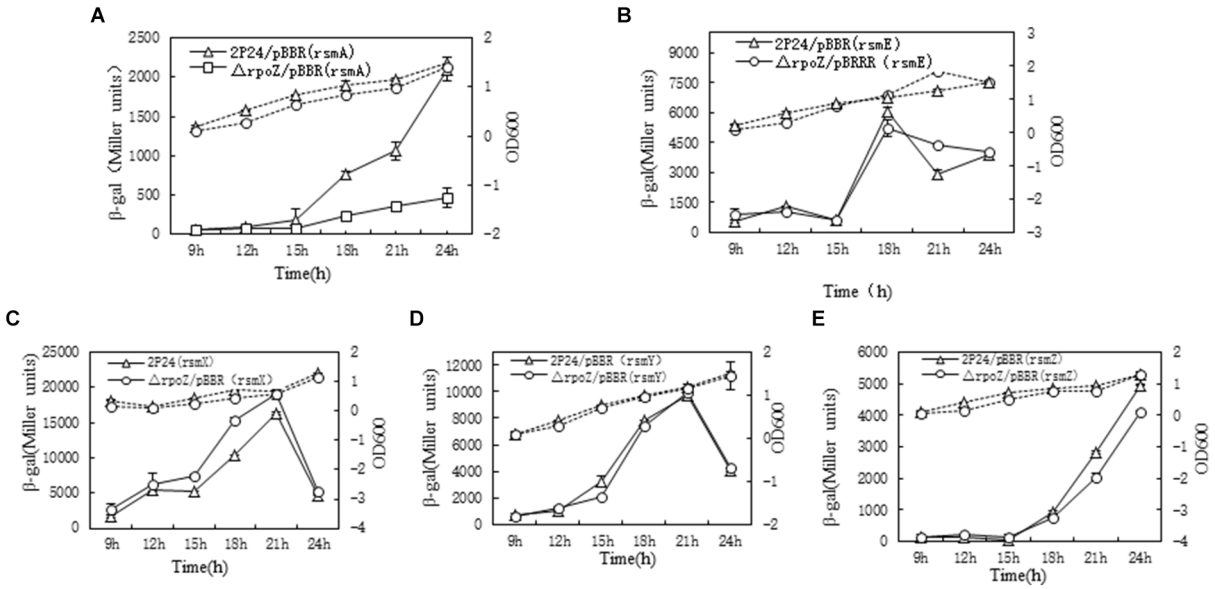


FIGURE 4
Determination of rpoZ gene influence gene related with 2,4 DAPG production. To determine ropZ regulation effect on expression of rsmA, rsmE, rsmX, rsmY and rsmZ, the β-gal activities in different strains were detected at different time points as shown in **A–E**. The β-galactosidase activity to OD_{600} value of each detected strain were shown as the solid lines with Miller units. The dotted lines were growth curve. Each value was calculated by 3 replicates. The values were from at least three independent assays. Three independent experiments were performed and the error bars were calculated.

2,4-DAPG can bind PhlF protein to counter this inhibition effect, which resulted to synthesize a large amount of 2.4-DAPG. As can be seen from Figure 5A, theβ-galactosidase activity of △rpoZ/pBBR and wild type 2P24/pBBR showed no difference, indicated that *rpoZ* gene had no effect on *phlF* gene transcription. Therefore, it was believed that *rpoZ* did not regulate the expression of *phlF* gene at the transcriptional level in strain 2P24.

RpoS is an important regulatory factor in the quiescent phase of bacteria (Hengge-Aronis, 2002), which is regulated by bacteria growth state. And *rpoS* affects the synthesis of secondary metabolites. The *rpoS* gene promoter fusion plasmid rpoS-LacZ was transferred into the wild strain 2P24 and the mutant strain respectively, and the β-galactosidase activity was detected. The results showed that the β-galactosidase activity of the wild-type strain was 1,189 Miller units, and the mutant strain was 800 Miller units. This indicated that *rpoZ* gene positively regulated the expression of *rpoS* at the transcriptional level (Figure 5B). From Figure 5B, we can also see that *rpoS* gene plays a certain role in the growth of bacteria. Before the stable period, the growth of *rpoZ*-deficient strain was slower than that of wild-type strain, but after the stable period, the growth of *rpoZ*-deficient strains was faster than that of wild-type strain.

The multidrug-resistant pump with active bacterial efflux is a system that bacteria can resist starting under adverse environment, by which harmful substances can be excluded from the cell. EmhR-emhABC pumps in *P. fluorescens* 2P24 are typical multi-drug resistant pumps. EmhR is the regulator of EmhABC pump, and the transcription level is negatively regulated by EmhABC. As shown in Figures 5C,D, there was no difference in β-galactosidase activity between mutant △rpoZ/pBBR and wild-type 2P24/pBBR, suggested that *rpoZ* did not regulate the expression of emhABC and emhR genes at the transcriptional level in strain 2P24. The growth curve of culture medium showed that *emhR* gene had no obvious influence on the bacteria growth.

3.6. Determination of biocontrol characteristic of *rpoZ* in *Pseudomonas fluorescens* 2P24

Previous studies have shown that 2,4-DAPG is the main biocontrol factor in many biocontrol bacteria (Wei and Zhang, 2005; Tian et al., 2010; Zhao et al., 2020). HPLC results showed that the wild strain could produce 2,4-DAPG, while the mutant strain △rpoZ/pBBR produced very little 2,4-DAPG (Figure 6A; Table 1). The antagonism experiment also proved that the mutant strain of *rpoZ* lost the antagonism ability against the *Rhizoctonia solani* in cotton. These results indicated that *rpoZ* had an important antagonistic effect on 2,4-DAPG pathogens. *RpoZ* gene had a significant effect on the swimming of the strain. The swimming of *rpoZ* deletion mutant was significantly slower than that of the wild type (Figure 6B; Table 2).

Biofilm is a solid structure similar to membrane formed by extracellular polysaccharides and microorganisms on the solid surface. Bacteria in the biofilm form an interactive community and are not free planktic cells (Cvitkvitch et al., 2003). It has been reported that *pcoI* positively regulated the formation of biofilm in *P. fluorescens* 2P24, and *rpoZ* positively regulated the transcription expression of *pcoI* (Wang et al., 2021). Therefore, we tested the regulatory effect of *rpoZ* gene on biofilm. We found that the formation of biofilm in *rpoZ* mutant strains decreased significantly compared with wild-type strain, and the complementary strain also basically recovered the level of wild-type strain (Figure 6C). This indicated that *rpoZ* gene had an effect on biofilm formation.

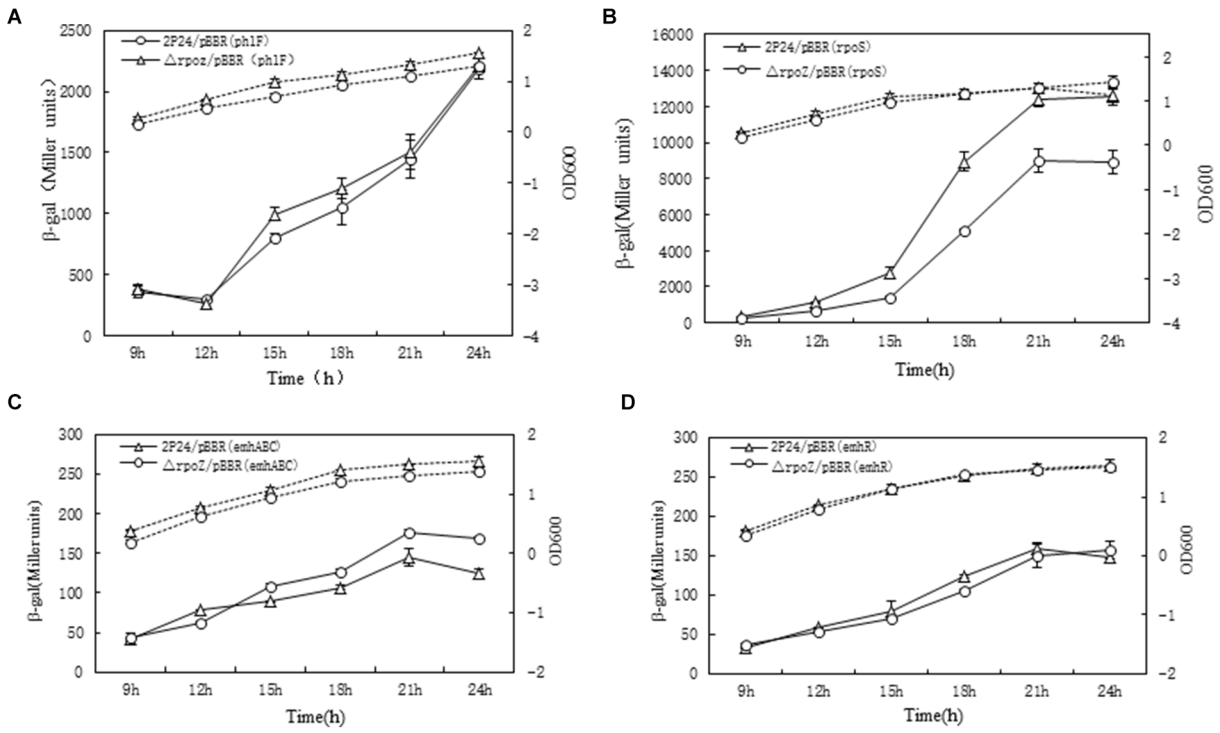


FIGURE 5
RpoZ regulates *phlF, rpoS, emhABC and emhR* expression in *P. fluorescens* 2P24. the β-gal activities in different strains were detected at different time points as shown in **(A–D)**. The β-galactosidase activity to OD_{600} value of each detected strain was shown as the solid lines with Miller units. The dotted lines were growth curve. Each value was calculated by 3 replicates. The values were from at least three independent assays. Three independent experiments were performed and the error bars were calculated.

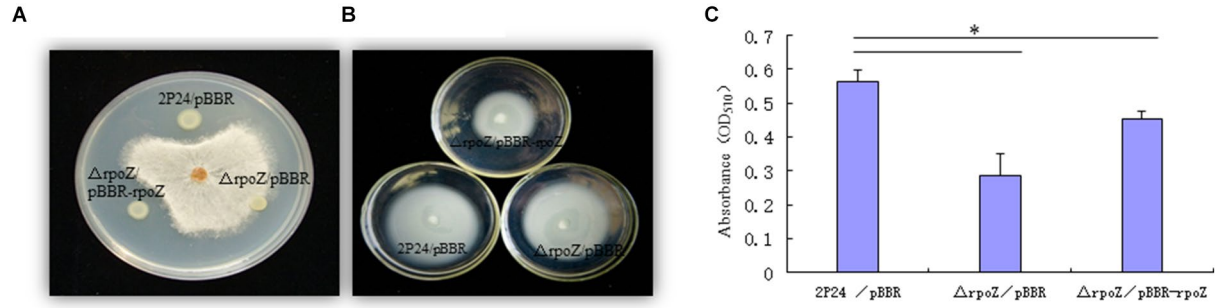


FIGURE 6
Determination of biocontrol characteristic of *rpoZ* in *P. fluorescens* 2P24. **(A)** Antagonistic ability of *rpoZ* in *P. fluorescens* 2P24 and its derivates against *Rhizoctonia solani* are shown. **(B)** The mobilities of *P. fluorescens* 2P24 and its derivates on medium are shown. **(C)** Regulation of *rpoZ* on biofilm formation. *p<0.05 by Student's t test.

4. Discussion

It was well known that the widely conserved omega subunit encoded by RpoZ is the smallest subunit of *Escherichia coli* RNA polymerase but is dispensable for bacterial growth (Sarkar et al., 2013). *RpoZ* is located upstream of *spoT* and shared a promoter with *spoT*. *SpoT* and *rpoZ* mutant strains showed slower growth. The researchers concluded that the slower growth phenotype of *rpoZ* mutant could be inhibited by *relA*, suggesting that the slower growth was not caused by the loss of *rpoZ*, but was the result of increased (P)ppGpp levels due to changes in polarity (Wendrich et al., 2000). To mimic the strict response, RNAP (with or without ω subunit) was analysed using an *in vitro* fusion transcription system. ω subunit was involved in the regulation of intracellular ppGpp. ω subunit was involved in the regulation of *relA* gene expression. When ω subunit was absent, *relA* transcriptional expression was decreased, thus ppGpp level and mRNA expression were decreased. The insertion mutation of *rpoZ* conferred a slow-growth phenotype when it was introduced into most strains (Gentry and Burgess, 1989). These results suggested that *rpoZ*

TABLE 1 Antagonistic ability of *rpoZ* in *Pseudomonas fluorescens* 2P24 and its derivates against *Rhizoctonia solani*.

	Antagonistic zone of no pathogen growth (mm)
2P24/pBBR	8.2 ± 0.2
△rpoZ/pBBR	0
△rpoZ/pBBR-rpoZ	8.0 ± 0.3

Data are averages of three repeat.

TABLE 2 The mobility of *P. fluorescens* 2P24 and its derivates.

Strains	The mobility of bacteria (mm)
2P24/pBBR	65 ± 1.0
△rpoZ/pBBR	40 ± 0.8
△rpoZ/pBBR-rpoZ	62 ± 0.7

Data are averages of three repeat.

indirectly regulates the intracellular levels of (P)ppGpp (Gentry and Burgess, 1989). The slower growth phenotype of *rpoZ* mutant wasn't attributed to the polar of *spoT* in the downstream. The researchers concluded that the slower growth phenotype of *rpoZ* mutant could be inhibited by *relA*, suggesting that the slower growth was not caused by the loss of *rpoZ*. And deletion of ω subunit does not affect the synthesis of (p)ppGpp, and any difference in phenotype observed is possibly due to the reduced binding of (p)ppGpp to RNAP in ΔrpoZ strain (Bhardwaj et al., 2018). But other scientists reported that deletion of ω does not affect the synthesis of (p)ppGpp, and any difference in phenotype observed is possibly due to the reduced binding of (p)ppGpp to RNAP in ΔrpoZ strain (Bhardwaj et al., 2018), So far, ΔrpoZ stain showed slow growth because of the influence of (p)ppGpp, rather than the polar of the *spoT* gene.

To further explore the regulation role of *rpoZ*, in this study, we identified *rpoZ* gene from *P. fluorescens* 2P24 and demonstrated it had positive regulation role in the expression of *phlA* at the transcriptional level, which affected the production of antibiotic 2,4-DAPG, and it resulted in the decreased biocontrol ability in ΔrpoZ stain (Zhou et al., 2005). It has been showed that the *rpoZ* gene was required for antibiotic production and morphological differentiation but is not essential for growth in *Streptomyces kasugaensis* (Kojima et al., 2002; Santos-Beneit et al., 2011). Deletion of the gene *rpoZ* in *Mycobacterium smegmatis* results in reducing growth rate, a change in colony morphology and fragmentation of the beta' subunit in the enzyme assembly (Mathew et al., 2005). In *rpoZ* mutant, production of actinorhodin,undecylprodigiosin and gray pigment closely associated with spores decreased, especially the expression of actinorhodin and gray pigment was completely inhibited (Santos-Beneit et al., 2011).

Previous studies have shown that PcoI/PcoR QS system in strain *P. fluorescens* 2P24 positively regulates biofilm formation and colonization of root circumference (Yan et al., 2009a,b). *RpoZ* gene positive regulation role in the expression *pcoI* gene at the transcriptional level, which affected the production of the signalling molecule AHL, and *RpoZ* gene positively regulated QS system. Therefore, we proposed that *rpoZ* could also affect colonization ability of root circumference and biofilm formation through QS system, thus affecting biocontrol ability. It reported that ΔrpoZ strain of *E. coli* showed defective biofilm formation only in minimal media and this indicated that ω subunit plays an important role in biofilm formation under stress conditions (Weiss et al., 2017). ΔrpoZ strain in *M. smegmatis* and *S. aureus* is known to be defective in biofilm formation (Mathew et al., 2006; Weiss et al., 2017).

The absence of *RpoZ* leads to a different set of genes being transcribed as seen in *E. coli*. It had an effect on the expression of *rsmA*, but it had no effect on the expression and transcription of untranslated other small RNA (Weiss et al., 2017). *RsmA* and *rsmE* are carbon storage regulatory factors, which can bind to the ribosome binding site RBS of mRNA transcribed by secondary metabolites HCN, 2,4-DAPG and Plt synthesis genes, thus preventing the initiation of translation and regulating the synthesis of these secondary metabolites.

5. Conclusion

Using Tn5 mutagenesis, we obtained *rpoZ* mutant in *P. fluorescens* 2P24. This facilitated us to study the *rpoZ* function. Our data showed that *rpoZ* gene was an important regulator of antibiotic 2,4-DAPG and QS system, which is important for investigating the mechanism of biocontrol activity in *P. fluorescens* 2P24.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

MZ, HZ, and BD wrote the manuscript. LZ, XW, and HZ designed the experiments. YW and XW performed the experiments. BD, DW, NL, and QZ analyzed the sequencing data. BD and DW advised on the English language editing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (grant no. 30860166), Science and Technology Xingmeng project of Inner Mongolia in China (grant no. KJXM-EEDS-2020008), and China Agriculture Research System of MOF and MARA (grant no. CARS-07-C-3).

Acknowledgments

We thank Gail P. Ferguson, Martin Schuster, Zhaoqing Luo and Rui Zhou for plasmids and reagents preparation. We thank Jianying Yue and Zhiying Wang for critical discussions on the project design and 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.

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.2023.1160913/full#supplementary-material

References

Bangera, M. G., and Thomashow, L. S. (1999). Identification and characterization of a gene cluster for synthesis of the polyketide antibiotic 2,4-diacetylphloroglucinol from *Pseudomonas fluorescens* Q2–87. *J. Bacteriol.* 181, 3155–3163. doi: 10.1128/JB.181.10.3155-3163.1999

Bertani, I., and Venturi, V. (2004). Regulation of the N-acyl homoserine lactone-dependent quorum-sensing system in rhizosphere *Pseudomonas putida* WCS358 and cross-talk with the stationaryphase RpoS sigma factor and the global regulator GacA. *Appl. Environ. Microbiol.* 70, 5493–5502. doi: 10.1128/AEM.70.9.5493-5502.2004

Bhardwaj, N., Syal, K., and Chatterji, D. (2018). The role of ω-subunit of *Escherichia coli* RNA polymerase in stress response. *Genes Cells* 23, 357–369. doi: 10.1111/gtc.12577

Burgess, R. R. (1969). Separation and characterization of the subunits of ribonucleic acid polymerase. *J. Biol. Chem.* 244, 6168–6176. doi: 10.1016/S0021-9258(18)63521-5

Chai, Y., Zhu, J., and Winans, S. C. (2001). And TrlR, a defective TraR-like protein of *Agrobacterium tumefaciens*, blocks TraR function in vitro by forming inactive TraR: Tra R dimer. *Mol. Microbiol.* 40, 414–421. doi: 10.1046/j.1365-2958.2001.02385.x

Chilton, M. D., Currier, T. C., Farrand, S. K., Bendich, A. J., Gordon, M. P., and Nester, E. W. (1974). *Agrobacterium tumefaciens* DNA and PS8 bacteriophage DNA not detected in crown gall tumors. *Proc. Natl. Acad. Sci. U. S. A.* 71, 3672–3676. doi: 10.1073/pnas.71.9.3672

Cvitkvitch, D. G., Li, Y. H., and Ellen, R. P. (2003). Quorum sensing and bioflim formation in streptococcal infections. *Microbiology* 112, 1626–1632. doi: 10.1172/JCI20430

Darst, S. A., Edwards, A. M., Kubalek, E. W., and Kornberg, R. D. (1991). Three-dimensional structure of yeast RNA polymerase II at 16 a resolution. *Cells* 66, 121–128. doi: 10.1016/0092-8674(91)90144-n

Darst, S. A., Kubalek, E. W., and Kornberg, R. D. (1989). Three-dimensional structure of *Escherichia coli* RNA polymerase holoenzyme determined by electron crystallography. *Nature* 340, 730–732. doi: 10.1038/340730a0

Dove, S. L., and Hochschild, A. (1998). Conversion of the omega subunit of *Escherichia coli* RNA polymerase into a transcriptional activator or an activation target. *Genes Dev* 12, 745–754. doi: 10.1101/gad.12.5.745

Fenton, A. M., Stphens, P. M., and Crowley, J. (1992). Exploitation of gene (s) involved in 2,4-diacetylphloroglucinol biosynthesis to confer a new biocontrol capability to a *Pseudomonas* strain. *Appl. Environ. Microbiol.* 58, 3873–3878. doi: 10.1002/bit.260401020

Gentry, D. R., and Burgess, R. (1989). Rpo Z, encoding the omega Subuni t of *Escherichia coli* RNA polymerase, is in the same operon as spot. *J. Bacteriol.* 171, 1271–1277. doi: 10.1128/jb.171.3.1271-1277.1989

Ghosh, P., Ishihama, A., and Chatterji, D. (2001). *Escherichia coli* RNA polymerase subunit omega and its N-terminal domain bind full-length β′to facilitate incorporation into the $\alpha_2\beta$ sub-assembly. *FEBS J.* 268, 4621–4627. doi: 10.1046/j.1432-1327.2001.02381.x

Gonzalez, J. E., and Keshavan, N. D. (2006). Messing with bacterial quorum sensing. *Microbiol. Mol. Biol. Rev.* 70, 859–875. doi: 10.1128/MMBR.00002-06

Hengge-Aronis, R. (2002). Signal transduction and regulatory mechanisms involve in control of the σs (RpoS) subnit of RNA polymerase. *Microbiol Mol Biol R.* 66:373. doi: 10.1128/MMBR.66.3.373-395.2002

Juhas, M., Wiehlmann, L., Huber, B., Jordan, D., Lauber, J., Salunkhe, P., et al. (2004). Global regulation of quorum sensing and virulence by Vqs R in *Pseudomonas aeruginosa*. *Microbiology* 150, 831–841. doi: 10.1099/mic.0.26906-0

Kievit, T. R., and Iglewski, B. H. (2000). Bacterial quorum sensing in pathogenic relationships. *Infect. Immun.* 68, 4839–4849. doi: 10.1128/IAI.68.9.4839-4849.2000

King, E. O., Ward, M. K., and Raney, D. E. (1954). Two simple media for the demonstration of pyocyanin and fluorescein. *J. Lab. Clin. Med.* 44, 301–307.

Kojima, I., Kasuga, K., Kobayashi, M., Fukasawa, A., Mizuno, S., Arisawa, A., et al. (2002). The rpo Z gene, encoding the RNA polymerase omega subunit, is required for antibiotic production and morphological differentiation in *Streptomyces kasugaensis*. *J. Bacteriol.* 184, 6417–6423. doi: 10.1128/JB.185.1.386.2003

Li, X., Gu, G. Q., Chen, W., Gao, L. J., Wu, X. H., and Zhang, L. Q. (2018). The outer membrane protein Opr F and the sigma factor sig X regulate antibiotic production in *Pseudomonas fluorescens* 2P24. *Microbiol Res.* 206, 159–167. doi: 10.1016/j.micres.2017.10.006

Lonetto, M., Gribskov, M., and Gross, C. A. (1992). The 70 family: sequence conservation and evolutionary relationships. *J. Bacteriol.* 174, 3843–3849. doi: 10.1128/jb.174.12.3843-3849.1992

Mathew, R., and Chatterji, D. (2006). The evolving story of the omega subunit of bacterial RNA polymerase. *Trends Microbiol.* 14, 450–455. doi: 10.1016/j.tim.2006.08.002

Mathew, R., Chatterji, D., Balachandar, R., and Chatterji, D. (2006). Deletion of the rpoZ gene, encoding the ω subunit of RNA polymerase, results in pleiotropic surface-related phenotypes in *Mycobacterium smegmatis*. *Microbiology* 152, 1741–1750. doi: 10.1099/mic.0.28879-0

Mathew, R., Ramakanth, M., and Chatterji, D. (2005). Deletion of the gene rpo Z, encoding the omega subunit of RNA polymerase, in *Mycobacterium smegmatis* results in fragmentation of the beta' subunit in the enzyme assembly. *J. Bacteriol.* 187, 6565–6570. doi: 10.1128/JB.187.18.6565-6570.2005

Miller, J. H. (1972). *Experiments in molecular genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Pessi, G., Williams, F., Hindle, Z., Heurlier, K., Holden, M. T., Camara, M., et al. (2001). The global posttranscriptional regulator Rsm a modulates production of virulence determinants and N-acylhomoserine lactones in *Pseudomonas aeruginosa*. *J. Bacteriol.* 183, 6676–6683. doi: 10.1128/JB.183.22.6676-6683.2001

Rampioni, G., Polticelli, F., Bertani, I., Righetti, K., Venturi, V., Zennaro, E., et al. (2007). The Pseudomonas quorum-sensing regulator Rsa L belongs to the tetrahelical superclass of H-T-H proteins. *J. Bacteriol.* 189, 1922–1930. doi: 10.1128/JB.01552-06

Rashid, M. H., and Kornberg, A. (2000). Inorganic polyphosphate is needed for swimming, swarming, and twitching motilities of *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U. S. A.* 97, 4885–4890. doi: 10.1073/pnas.060030097

Reimmann, C., Beyeler, M., Latifi, A., Winteler, H., Foglino, M., Lazdunski, A., et al. (1997). The global activator GacA of *Pseudomonas aeruginosa* PAO1 positively controls the production of the autoinducer N-butyryl-homoserine lactone and the formation of the virulence factors pyocyanin, cyanide, and lipase. *Mol. Microbiol.* 24, 309–319. doi: 10.1046/j.1365-2958.1997.3291701.x

Sakuragi, Y., and Kolter, R. (2007). Quorum-sensing regulation of the biofilm matrix genes (pel) of *Pseudomonas aeruginosa*. *J. Bacteriol.* 189, 5383–5386. doi: 10.1128/JB.00137-07

Santos-Beneit, F., Barriuso-Iglesias, M., Fernández-Martínez, L. T., Martínez-Castro, M., Sola-Landa, A., Rodríguez-García, A., et al. (2011). The RNA polymerase omega factor Rpo Z is regulated by pho P and has an important role in antibiotic biosynthesis and morphological differentiation in *Streptomyces coelicolor*. *Appl. Environ. Microbiol.* 77, 7586–7594. doi: 10.1128/AEM.00465-11

Sarkar, P., Sardesai, A. A., Murakami, K. S., and Chatterji, D. (2013). Inactivation of the bacterial RNA polymerase due to acquisition of secondary structure by the ω subunit. *J. Biol. Chem.* 288, 25076–25087. doi: 10.1074/jbc.M113.468520

Schnider-Keel, U., Seematter, A., Maurhofer, M., Blumer, C., Duffy, B., Gigot-Bonnefoy, C., et al. (2000). Autoinduction of 2,4-diacetylphloroglucinol biosynthesis in the biocontrol agent *Pseudomonas fluorescens* CHA0 and repression by the bacterial metabolites salicylate and pyoluteorin. *J. Bacteriol.* 182, 1215–1225. doi: 10.1128/JB.182.5.1215-1225.2000

Schultz, P., Célia, H., Riva, M., Sentenac, A., and Oudet, P. (1993). Three-dimensional model of yeast RNA polymerase I determined by electron microscopy of two-dimensional crystals. *EMBO J.* 12, 2601–2607. doi: 10.1002/j.1460-2075.1993.tb05920.x

Shanahan, P., O'Sullivan, D. J., Simpson, P., Glennon, J. D., and O'Gara, F. (1992). Isolation of 2,4-diacetylphloroglucinol from a fluorescent pseudomonad and

investigation of physiological parameters influencing its production. *Appl. Environ. Microbiol.* 58, 353–358. doi: 10.1016/S0065-2164(08)70256-9

Tian, T., Wu, X. G., Duan, H. M., and Zhang, L. Q. (2010). The resistance-nodulation- division efflux pump Emh ABC influences the production of 2,4-diacetylphloroglucinol in *Pseudomonas fluorescens* 2P24. *Microbiology* 156, 39–48. doi: 10.1099/mic.0.031161-0

Vincent, M. N., Harrison, L. A., Brackin, J. M., Kovacevich, P. A., Mukerji, P., Weller, D. M., et al. (1991). Genetic analysis of the antifungal activity of a soilborne *Pseudomonas aureofaciens* strain. *Appl. Environ. Microbiol.* 57, 2928–2934. doi: 10.1128/aem.57.10.2928-2934.1991

Wang, J., Luo, Y., Gu, Y., and Wei, H. L. (2021). Characterization of the SPI-1 type III secretion system in *Pseudomonas fluorescens* 2P24. *Front. Microbiol.* 12:749037. doi: 10.3389/fmicb.2021.749037

Waters, C. M., Lu, W., Rabinowitz, J. D., and Bassler, B. L. (2008). Quorum sensing controls biofilm formation in Vibrio cholera through modulation of cyclic di-GMP levels and repression of vps T. *J. Bacteriol.* 190, 2527–2536. doi: 10.1128/JB.01756-07

Wei, H. L., Wang, Y., Zhang, L. Q., and Tang, W. H. (2004). Identification and characterization of biocontrol bacterial strain 2P24 and CPF-10. *Acta Phytopathol Sin.* 34, 80–85. doi: 10.13926/j.cnki.apps.2004.01.014

Wei, H. L., and Zhang, L. Q. (2005). Cloning and functional characterization of the *gacS* gene of the biocontrol strain *Pseuodomonas fluorescens* 2P24. *Acta Microbiol Sin.* 45, 368–372. doi: 10.13343/j.cnki.wsxb.2005.03.011

Wei, H. L., and Zhang, L. Q. (2006). Quorum-sensing system influences root colonization and biological control ability in *Pseudomonas fluorescens* 2P24. *Antonie Van Leeuwenhoek* 89, 267–280. doi: 10.1007/s10482-005-9028-8

Weiss, A., Moore, B. D., Tremblay, M. H., Chaput, D., Kremer, A., and Shaw, L. N. (2017). The omega subunit governs RNA polymerase stability and transcriptional specificity in *Staphylococcus aureus*. *J. Bacteriol.* 199:e00459-16. doi: 10.1128/JB.00459-16

Wendrich, T. M., Beckering, C. L., and Marahiel, M. (2000). Charecter ization of the Rel a/Spol T gene from *Bacillus Stearothermophilus*. *FEMS Mirobiologv Letter* 190, 195–201. doi: 10.1016/s0378-1097(00)00335-9

Wu, X., Liu, J., Zhang, W., and Zhang, L. (2012). Multiple-level regulation of 2,4-diacetylphloroglucinol production by the sigma regulator Psr a in *Pseudomonas fluorescens* 2P24. *PLoS One* 7:e50149. doi: 10.1371/journal.pone.0050149

Yan, Q., Gao, W., Wu, X. G., and Zhang, L. Q. (2009a). Regulation of the Pco I/Pco R quorum-sensing system in *Pseudomonas fluorescens* 2P24 by the pho P/pho Q two-component system. *Microbiology* 155, 124–133. doi: 10.1099/mic.0.020750-0

Yan, Q., Wu, X. G., Wei, H. L., Wang, H. M., and Zhang, L. Q. (2009b). Differential control of the Pco I/Pco R quorum-sensing system in *Pseudomonas fluorescens* 2P24 by sigma factor RpoS and the Gac S/GacA two-component regulatory system. *Microbiol. Res.* 164, 18–26. doi: 10.1016/j.micres.2008.02.001

Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K., and Darst, S. A. (1999). Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 a resolution. *Cells* 98, 811–824. doi: 10.1016/s0092-8674(00)81515-9

Zhang, Y., Zhang, B., Wu, H., Wu, X., Yan, Q., and Zhang, L. Q. (2020b). Pleiotropic effects of Rsm a and Rsm E proteins in *Pseudomonas fluorescens* 2P24. *BMC Microbiol.* 20, 20:191. doi: 10.1186/s12866-020-01880-x

Zhang, Y., Zhang, B., Wu, X., and Zhang, L. Q. (2020a). Characterization the role of GacA-dependent small RNAs and Rsm a family proteins on 2,4-diacetylphloroglucinol production in *Pseudomonas fluorescens* 2P24. *Microbiol Res. Mar* 233:126391. doi: 10.1016/j.micres.2019.126391

Zhao, M. M., Lyu, N., Wang, D., Wu, X. G., Zhao, Y. Z., Zhang, L. Q., et al. (2020). Phl G mediates the conversion of DAPG to MAPG in *Pseudomonas fluorescens* 2P24. *Sci. Rep.* 10, 10:4296. doi: 10.1038/s41598-020-60555-9

Zhou, H. Y., Wei, H. L., Liu, X. L., Wang, Y., Zhang, L. Q., and Tang, W. H. (2005). Improving biocontrol activity of *Pseudomonas fluorescens* through chromosomal integration of 2,4-diacetylphloroglucinol biosynthesis genes. *Chin. Sci. Bull.* 50, 775–781. doi: 10.1007/BF03183678

frontiers | Frontiers in Microbiology

TYPE Mini Review
PUBLISHED 07 June 2023
DOI 10.3389/fmicb.2023.1168246

OPEN ACCESS

EDITED BY
Feng Gao,
Tianjin University, China

REVIEWED BY
Masataka Tsuda,
Tohoku University, Japan
Rafael Rivilla,
Autonomous University of Madrid, Spain
Weihong Zhong,
Zhejiang University of Technology, China

*CORRESPONDENCE
Hikaru Suenaga
✉ suenaga-hikaru@aist.go.jp

RECEIVED 17 February 2023
ACCEPTED 16 May 2023
PUBLISHED 07 June 2023

CITATION
Fujihara H, Hirose J and Suenaga H (2023) Evolution of genetic architecture and gene regulation in biphenyl/PCB-degrading bacteria. *Front. Microbiol.* 14:1168246. doi: 10.3389/fmicb.2023.1168246

Evolution of genetic architecture and gene regulation in biphenyl/PCB-degrading bacteria

Hidehiko Fujihara[1], Jun Hirose[2] and Hikaru Suenaga[3]*

[1]Department of Food and Fermentation Sciences, Faculty of Food and Nutrition Sciences, Beppu University, Beppu, Japan, [2]Department of Applied Chemistry, Faculty of Engineering, University of Miyazaki, Miyazaki, Japan, [3]Cellular and Molecular Biotechnology Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan

A variety of bacteria in the environment can utilize xenobiotic compounds as a source of carbon and energy. The bacterial strains degrading xenobiotics are suitable models to investigate the adaptation and evolutionary processes of bacteria because they appear to have emerged relatively soon after the release of these compounds into the natural environment. Analyses of bacterial genome sequences indicate that horizontal gene transfer (HGT) is the most important contributor to the bacterial evolution of genetic architecture. Further, host bacteria that can use energy effectively by controlling the expression of organized gene clusters involved in xenobiotic degradation will have a survival advantage in harsh xenobiotic-rich environments. In this review, we summarize the current understanding of evolutionary mechanisms operative in bacteria, with a focus on biphenyl/PCB-degrading bacteria. We then discuss metagenomic approaches that are useful for such investigation.

KEYWORDS

xenobiotic compounds, degrading bacteria, mobile genetic elements, gene regulation and expression, adaptive evolution

Introduction

Xenobiotic compounds are man-made chemicals that are present at unnaturally high concentrations in the natural environment. A variety of bacteria in the environment can utilize various xenobiotic compounds as a source of carbon and energy (Van der Meer et al., 1992). Phylogenetically unrelated bacterial strains often share similar metabolic pathways and enzyme systems for the degradation of xenobiotic compounds. It is believed that bacteria have acquired the ability to degrade even xenobiotic compounds they have never encountered (Nagata et al., 2015; Jeffries et al., 2018; Miglani et al., 2022). The bacterial strains degrading xenobiotics are suitable models to investigate the adaptation and evolutionary processes of bacteria because they appear to have emerged relatively soon after the release of these compounds into the natural environment.

Polychlorinated biphenyls (PCBs) are xenobiotic compounds in which the aromatic biphenyl carbon skeleton contains between one and 10 chlorine atoms. The high chemical stability, superhydrophobicity, and toxicity of PCBs make them some of the most serious and persistent environmental pollutants (Abraham et al., 2002; Furukawa and Fujihara, 2008). It is therefore somewhat surprising that many microbes that are capable of degrading PCBs have been identified. A number of biphenyl-utilizing bacteria with the ability to degrade PCBs have been isolated and characterized (Pieper and Seeger, 2008). Lignin is a complex substance with a phenylpropane structure; it contains various biphenyl molecules and is widely distributed

throughout the earth. Biphenyl-degrading bacteria are thought to be responsible for the final stage of lignin degradation (Iram et al., 2021). Since biphenyl dioxygenase, the first enzyme in the biphenyl catabolic pathway, hydroxylates plant-derived flavonoids (Zubrova et al., 2021) and its homologous enzyme oxidizes dehydroabietic acid (Witzig et al., 2007), a biphenyl-degrading pathway might be involved in the degradation of plant secondary metabolites other than lignin. Thus, the ancestry of biphenyl-utilizing bacteria and their catabolic genes is quite ancient, and the genes may be distributed across a wide range of bacteria. Further, they would have the potential to adapt to different aromatic compounds, including PCBs. Therefore, the biphenyl/PCB-degradation system in bacteria appear to be a suitable model for the study of microbial adaptive evolution.

Recently, many bacterial genomes and metagenomes derived from environments contaminated with xenobiotic compounds have been analyzed at an accelerated pace (Garrido-Sanz et al., 2018; Hirose et al., 2019; Miglani et al., 2022). Much evidence has been found to support the idea that different biphenyl/PCB-degrading bacteria have evolved in the environment through different processes. In this review manuscript, we summarize the diversity, recruitment, and expression of degradation genes for biphenyl/PCB, shedding light on the sophistication of degradation gene systems and the adaptive evolution of these host bacteria.

Bacterial mobile genetic elements

Mobilization of the catabolic genes in bacteria can be accomplished through a variety of mobile genetic elements (MGEs), including plasmid, transposon, and integrative and conjugative elements (ICEs). These genes are modified and rearranged in different ways in host bacterial cells. Degradation gene clusters for biphenyl/PCB (*bph*) are often located on MGEs, and can be transferred between bacterial cells, conferring degradation capacity to non-degrading bacteria (Van der Meer et al., 1992; Tsuda et al., 1999; Springael and Top, 2004; Satola et al., 2013).

Plasmids

Plasmids are mobile genetic elements that facilitate rapid adaptation and evolution by conjugative transfer between bacterial cells in the environment (Smillie et al., 2010; Aminov, 2011; Rodríguez-Beltrán et al., 2020). Catabolic plasmids contain the complete set of genes encoding the enzymes for the degradation of a xenobiotic compound. They are relatively large (more than 50 kb), due to the presence of numerous insertion sequences (ISs) and transposons. The important characteristic of catabolic plasmids is incompatibility. That is, plasmids are classified into incompatibility (Inc) groups based on their replication and partitioning systems; two plasmids of the same group cannot replicate in the same cell and are considered incompatible (Top and Springael, 2003; Popowska and Krawczyk-Balska, 2013; Shintani et al., 2015). Genes for xenobiotic degradation are often found on broad-host-range IncP-1 plasmids, such as pSS60 (Burlage et al., 1990), pBRC60 (Burlage et al., 1990) and pJP4 (Newby et al., 2000).

A variety of PCB-degrading phenotypes have also been attributed to catabolic plasmids (Furukawa and Fujihara, 2008). The *bph* genes of thermophilic *Geobacillus* sp. strain JF8 are located on a plasmid pBt40 (Mukerjee-dhar et al., 2005). *Rhodococcus* sp. RHA1 harbors large linear plasmids, including pRHL1 (1,100 kb), pRHL2 (450 kb), and pRHL3 (330 kb), and its *bph* genes are mainly located on the pRHL1 (Shimizu et al., 2001). Many other *bph* gene clusters have been identified on the linear mega-plasmids of different *Rhodococcus* species (Taguchi et al., 2004; Garrido-Sanz et al., 2020). Both the order and sequence of the *bph* genes have been shown to differ among rhodococci, and there is evidence of recombination around *bph* gene clusters, such as insertion of transpose (Taguchi et al., 2007). These findings suggest that these *bph* gene clusters evolved separately and were spread in rhodococci by horizontal transfer. *Cupriavidus* sp. SK-4, a PCB-degrading strain reported to utilize di-*ortho*-substituted biphenyl, was found to harbor a single plasmid: pSK4. Experimental results showed that pSK4 could be mobilized into *Pseudomonas putida* MB1335 and the PCB-degrading enzymes could be expressed in this strain (Ilori et al., 2015). Bacterial plasmids are important vehicles for horizontal gene transfer (Redondo-Salvo et al., 2020), and can therefore play a key role in the evolution of catabolic pathways and their hosts.

Transposons

Transposons are defined segments of DNA that are able to move from one genetic location to another target location in the absence of any nucleotide sequence homology. Most bacterial transposable elements, including ISs and transposons, can be traditionally classified into three classes: class I, class II, and conjugative transposons (Tan, 1999; Tsuda et al., 1999). However, a revised classification system has been proposed wherein conjugative transposons, genomic islands, and integrative plasmids would be collectively called ICEs (Burrus and Waldor, 2004; Johnson and Grossman, 2015).

Bacterial class I elements include the simple ISs, which carry only the genetic determinants for transposition (usually transposase) and composite transposons, in which various genetic traits unrelated to transposition are flanked by two copies of very similar ISs in direct or inverted orientation (Ross et al., 2021). Bacterial class II transposons generally carry the genes for their transposition (transposase, and resolvase) and one or more phenotypic traits between their terminal inverted repeats (Ross et al., 2021). It has been reported that several bacterial class II transposons play a crucial role in the widespread distribution of various catabolic gene clusters, such as the cluster of genes encoding toluene, naphthalene, and carbazole degradation pathways (Tsuda et al., 1999; Sota et al., 2006; Nojiri, 2012). Only a few bacterial transposons carrying catabolic genes for biphenyl/PCB have been experimentally proven to be mobile. An example of a functional catabolic transposon is Tn*5280*, which was identified on plasmid pP51 of chlorobenzene-utilizing *Pseudomonas* sp. strain P51 (Tan, 1999; Top et al., 2002). Tn*5280* transposes randomly at different chromosomal sites in *P. putida* KT2442 (van der Meer et al., 1991). This strongly suggests that some of the PCB degradation genes in strain P51 originated from a toluene or benzene degradation pathway, probably by HGT.

Integrative and conjugative elements

Integrative and conjugative elements (ICEs) are bacterial mobile genetic elements that are excised from the chromosome by site-specific recombination between the attachment ends (*attL* and *attR*) mediated by an ICE-specific integrase (Figure 1; Delavat et al., 2017). The excised ICE molecule undergoes single-strand cleavage at the

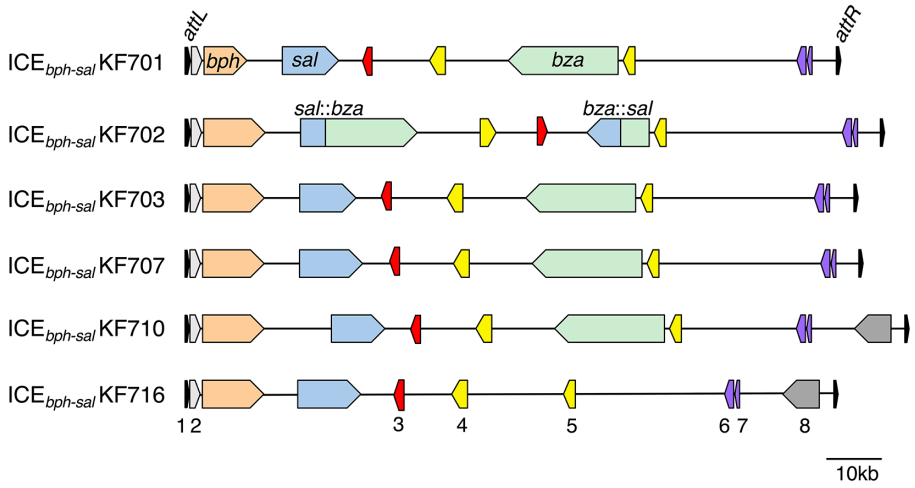


FIGURE 1
Organization of the ICE$_{bph\text{-}sal}$ in KF strains. ICE$_{bph\text{-}sal}$KF701, ICE$_{bph\text{-}sal}$KF702, ICE$_{bph\text{-}sal}$KF703, ICE$_{bph\text{-}sal}$KF707, and ICE$_{bph\text{-}sal}$KF710 carry the *int* gene, *bph* genes, *sal* genes, and *bza* genes. ICE$_{bph\text{-}sal}$KF716 carries the *int* gene, *bph*, and *sal* genes, but not the *bza* genes. The *bza* genes and *sal* genes in ICE$_{bph\text{-}sal}$KF702 are formed in a fusion gene. 1, tRNA-Gly (CCC) genes (partial) (black); 2, *int* genes (grey); 3, *traI* genes (red); 4, VirB4 components of the type IV secretory pathway (yellow); 5, a VirD4 component of the type IV secretory pathway (yellow); 6, *parB* genes (purple); 7, *parA* genes (purple); and 8, YbhFSR ABC-type transporter (dark grey). The *attL* (18bp) site, including the 18bp of the 3′ end of the tRNA-Gly gene, and *attR* (18bp) site are indicated.

origin of transfer (*oriT*). TraI (relaxase) binds to the *oriT* of the circular intermediate of the ICE, cleaves one strand and binds to the 5′ end, which is then recognized by the VirD4 complex and transported through the mating pair formation system (MPF) to a recipient cell. VirD4 and VirB4 are large (> 70 kDa) ATPase proteins, and VirB4 is a component of MPF (Christie et al., 2014). The double-stranded DNA is reconstituted and site-specifically recombines with the recipient's *attB* attachment site on the chromosome to become re-integrated (Guglielmini et al., 2011). The integration sites in the genome of the recipient cell, such as the structural genes for tRNAs, have been reported for many conjugative transposons (Wozniak and Waldor, 2010; Delavat et al., 2017).

Although ICEs often carry cargo genes involved in pathogenicity, antibiotic resistance and heavy metal resistance to endow hosts with phenotypes beneficial for niche adaptation (Bellanger et al., 2014), they are also known to carry cargo genes for the catabolism of xenobiotic compounds (Hirose, 2023). ICE*clc* is a well-known ICE in xenobiotic biodegradation and carries cargo genes involved in the metabolism of chlorocatechol (*clc* genes) and aminophenol. The element was originally identified in the 3-chlorobenzoate-degrader *Pseudomonas knackmussii* B13 (Gaillard et al., 2006), and was almost identical to the ICEs inserted into the chromosome of *Paraburkholderia xenovorans* (formerly *Burkholderia xenovorans*) LB400 (ICE*clc*-LB400) (Chain et al., 2006) and *P. aeruginosa* JB2 (ICE*clc*-JB2) (Obi et al., 2018). *P. xenovorans* LB400 is also a well-characterized PCB-degrading bacteria, in which chlorobenzoate is formed during the degradation of PCB (Mondello, 1989). This bacterial strain provide insight into the roles of ICEs in the evolution of catabolic pathways for the biodegradation of chlorinated aromatic compounds. On degradative ICEs, catabolic gene products (e.g., *bph*, *nah*, and *sal*) that are related to the same substrate always share nearly 100% identity, whereas the sequences of the gene products of the transmission module exhibit variations (Mohapatra et al., 2022; Hirose, 2023). Although it has been suggested that there is a mechanism governing catabolic gene insertion and exchange on ICEs, there is little evidence to support this idea.

It is known that some biphenyl/PCB catabolic genes, *bph*, are horizontally transferred via ICEs. Several of the degradative ICEs carrying the *bph* gene belong to either of two groups, the Tn*4371* family or ICE*clc* family, whose respective members share a common core region. The transfer module is required for the conjugal transfer from donor to recipient forming the type IV secretion system (T4SS) (Bellanger et al., 2014; Delavat et al., 2017). Many components required for conjugal transfer constitute a "core region" that is conserved among ICE family members. Tn*4371* is the first ICE carrying *bph* genes and was found in the chromosome of *Cupriavidus oxalacticus* A5 (Springael et al., 1993). The chromosome of *Acidovorax* sp. strain KKS102 contains ICE$_{KKS102}$*4677* (Ohtsubo et al., 2012), which belongs to the Tn*4371* family (Toussaint et al., 2003). ICE$_{KKS102}$*4677* is known to be transferred by conjugation to a wide range of bacteria across the genera via a circular intermediate. The *bph* genes of *Cupriavidus basilensis* KF708 and *Commamonas testosteroni* KF712 are also located on an ICE. The genes ICE$_{bph}$KF708 and ICE$_{bph}$KF712, which carry the *bph* genes of KF708 (Suenaga et al., 2015) and KF712 (Hirose et al., 2015a), are almost identical to ICE$_{KKS102}$*4677* and Tn*4371*, respectively. A new ICE*clc* family carrying *bph* genes and salicylic acid catabolic genes, *sal*, was found in the PCB-degrading strain *Pseudomonas stutzeri* KF716 (Hirose et al., 2015b). The ICE$_{bph\text{-}sal}$KF716 contains common core regions that show homology with those of ICE*clc* from *P. knackmussii* B13 (Gaillard et al., 2006) and ICE$_{XTD}$ from *Azoarcus* sp. CIB (Zamarro et al., 2016). A comparison of the genetic loci revealed that several putative ICEs from *P. putida* B6-2 (Tang et al., 2011), *P. alcaliphila* JAB1 (Ridl et al., 2018), *P. stutzeri* AN10 (Brunet-Galmés et al., 2012), and *P. stutzeri* 2A20 (Heinaru et al., 2016) had core regions highly conserved with those of ICE$_{bph\text{-}sal}$KF716, along with a variable region encoding the catabolic genes for phenol, naphthalene and biphenyl. ICE$_{bph\text{-}sal}$KF716 was reported to have been transferred from *P. stutzeri* KF716 to *P. aeruginosa* PAO1

via a circular extrachromosomal intermediate form (Hirose et al., 2021). These reports demonstrate that ICE subfamily members that share core regions highly conserved with those of ICE*bph-sal*KF716 are widely distributed among aromatic-degrading bacteria.

Evidence of the evolution of PCB-degradation gene clusters in polluted sites

In the past few years, several novel studies have provided insights into the diversity and evolution of biphenyl/PCB catabolic genes in the process of adaptation to environmental niches (Hirose et al., 2019, 2021). Furukawa et al. determined the whole genomes of 10 biphenyl/PCB-degrading bacterial strains (KF strains) isolated from a biphenyl-contaminated soil sample (Furukawa et al., 1989). Genome analyses revealed that all 10 strains had the *bph* genes, while seven strains also had *sal* genes (Figure 1). A series of ICEs named ICE$_{bph\text{-}sals}$ that were larger (more than 110 kb) than many other ICEs (Bellanger et al., 2014) contained highly conserved *bph* genes and *sal* genes (Hirose et al., 2019). Most of these ICE$_{bph\text{-}sals}$ possessed benzoate catabolic genes encoding the extradiol cleavage (*bza*) pathway. The fusion gene cluster of *sal:bza* and *bza:sal* was found in ICE$_{bph\text{-}sal}$KF702 (Fujihara et al., 2015), and was likely generated by homologous recombination between the *sal* and the *bza* genes.

ICE$_{bph\text{-}sal}$ of *P. putida* KF715 existed both in an extrachromosomal circular form (referred to as ICE$_{bph\text{-}sal}$ [circular] or pKF715A; hereinafter called pKF715A) and an integrated form in the chromosome (referred to as ICE$_{bph\text{-}sal}$KF715 [integrated]) in stationary phase culture (Suenaga et al., 2017). The ICE$_{bph\text{-}sal}$ KF715 was transferred at high frequency into *P. putida* AC30 and *P. putida* KT2440, and it was stably maintained in a circular form, pKF715A. The pKF715A in these transconjugant strains was further transferred into *P. putida* F39/D, and it existed in an integrated form in the chromosome. The structural features of *bph* and its flanking regions between KF701 and KF715 were almost identical, indicating that the *bph-sal* clusters were horizontally transferred to one another at this site between KF701 and KF715. Various mobile genes encoding transposases and a retron encoding retron-type reverse transcriptases have been shown to be inserted in the *bph-sal* clusters of ICE*bph-sal*s. It has been hypothesized that these inserted sequences contribute to the exchange of the *bph* gene with the upper *nah* operon, the counterpart of the *bph* gene in the *nah-sal* cluster which is involved in naphthalene catabolism (Hirose et al., 2021; Mohapatra et al., 2022). These studies demonstrate that the *bph* gene clusters have integrative functions, are transferred among soil bacteria by various MGEs, including plasmids, transposons, and ICEs, and are diversified through modification.

Regulation of catabolic operons

The regulatory systems of aromatic compound-degradation genes are crucial evidence of the evolution of xenobiotic compound-degradation genes such as biphenyl, naphthalene, and phenol. In general, the expression of these degradation genes is controlled by one or more regulatory proteins (Diaz and Prieto, 2000). The regulatory protein binds to the effector, the substrate itself, or intermediates, thereby affecting the transcription of the operons. The many PCB catabolic operons in Gram-negative bacteria are regulated by regulatory proteins belonging to the GntR family (Merlin et al., 1997; Watanabe et al., 2000; Ohtsubo et al., 2001). In general, the regulators belonging to the GntR family acts as repressors (Tropel and van der Meer, 2004). The regulators bind to the operator region in the absence of substrates (biphenyl/PCBs) and inhibit the transcription of their operon (Figure 2). One such strain, KKS102, possesses a *bph* operon consisting of *bphSEFGA1A2A3BCA4R*. This operon is downregulated by a BphS belonging to the GntR family in the absence of HOPD (2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoic acid), an intermediate of biphenyl catabolism (Ohtsubo et al., 2001). The *bph* genes in Tn*4371* isolated from *C. oxalacticus* A5 consisting of *bphEFGA1A2A3BCD* are also downregulated by BphS belonging to the GntR family (Merlin et al., 1997). These degrading microorganisms are thought to reduce energy consumption by suppressing the transcription of degradation genes in the absence of substrates.

On the other hand, in *Pseudomonas furukawaii* KF707 (formerly *P. pseudoalcaligenes* KF707) (Kimura et al., 2018), a *bph* gene cluster consisting of *bphR1A1A2A3A4BCX0X1X2X3D* is upregulated by BphR1 (Watanabe et al., 2000). This regulator protein also belongs to the GntR family and acts as an activator having an effect opposite that of KKS102 (Ohtsubo et al., 2001). The details of this regulator protein will be described later. In addition, *bph* genes in *P. xenovorans* LB400 are also upregulated by ORF0, which belongs to the GntR family (Denef et al., 2004). This phenomenon is likely caused by the positional relationship of the promoter and operator of *bph* genes (Figure 2). In strain KKS102, the promoter and operator sequence of *bphS* exists between the *bphS* and *bphEGFABDR* genes, and thus the binding of BphS on the operator inhibits the binding of RNA polymerase on the promoter sequence of the *bphEGFABDR* genes (Ohtsubo et al., 2001). On the other hand, in strain KF707, the promoter and operator regions of the *bphR1* and *bphA* genes do not overlap, and thus the binding of BphR1 on the *bphR1* operator does not inhibit the binding of RNA polymerase. The binding of a complex of HOPD-BphR1 and HOPD-BphR2 on the operator could induce the bending of DNA molecules and enhance the binding of RNA polymerase on the promoter region of *bph* genes (Figure 2). Therefore, the expression of the *bph* operon is greatly enhanced in the presence of biphenyl in strain KF707 (Watanabe et al., 2000, 2003). Although the regulatory system for biphenyl/PCB degradation differs between strains KKS102 and KF707, in both cases the regulatory gene plays an important role in the efficient use of energy for growth and in the survival of the host in polluted environments.

The KF707 strain has another putative regulator, BphR2, that belongs to the LysR family and acts as an activator of the *bph* upper operon (*bphR1* and *bphABC*) (Figure 2). The *bphR2* gene is located far downstream (6.6 kb) of the *bph* genes and just upstream of the salicylate catabolic (*sal*) genes (Fujihara et al., 2006). BphR2 binds to the operator regions of *bphR1* and *bphA1* and activates the transcription of the *bph* upper operon. Interestingly, BphR2 also activates the transcription of *sal* genes in the presence of hydroxy muconate semi-aldehyde, the intermediate of salicylate. This regulatory behavior is the same as that of NahR, the regulator of the naphthalene catabolic (*nah*) operon. Actually, the *bphR2* and *sal* genes of strain KF707 have high similarity to the *sal* genes of the *nah* gene cluster containing *nahR* on *P. stutzeri* AN10 (Bosch et al., 1999). This indicates that the *bph-sal* gene cluster of strain KF707 might have

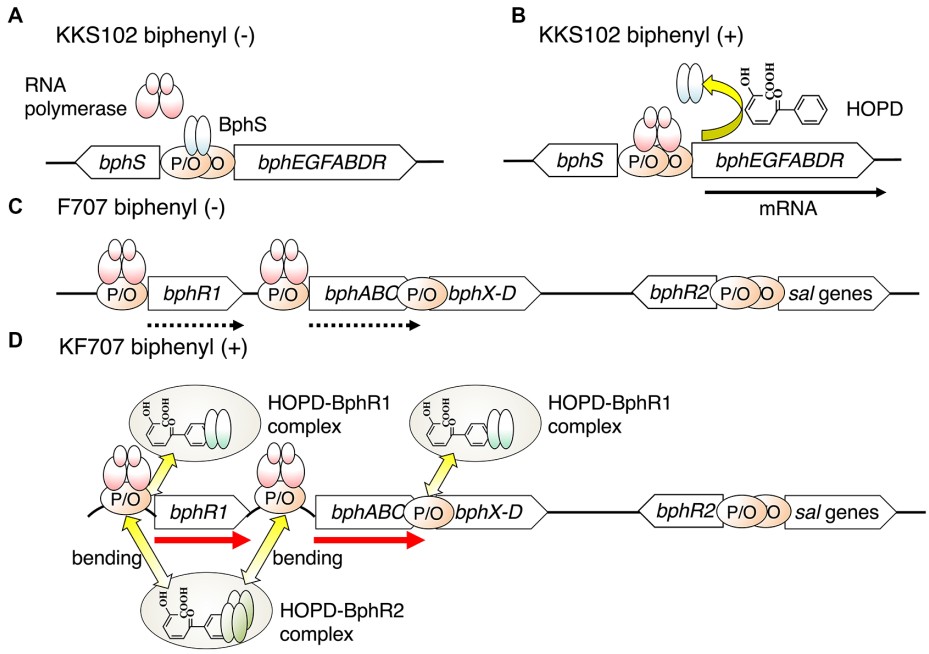


FIGURE 2
A transcriptional model of the *bph* gens in KKS102 in the absence **(A)** and in the presence **(B)** of biphenyl, and in KF707 in the absence **(C)** and in the presence of biphenyl **(D)**. The black arrow shows induced transcription, the dotted arrow constitutive transcription, and the red arrow upregulated transcription. P, promoter; O, operator.

evolved through module exchange between the *nah* gene and *bph* gene in the *nah-sal* cluster.

Several investigations into the regulation of biphenyl/PCB-degrading Gram-positive bacteria have also been reported. In strain *R. jostii*, RHA1 possesses diverse biphenyl/PCB-degradation genes that encode multiple isozymes for each metabolism step and are distributed among multiple clusters (Takeda et al., 2004). The transcriptions of these genes are induced by dual regulatory systems. In the presence of biphenyl, a BphS1T1 two-component system induced five biphenyl-degrading gene clusters. And another two-component system, BphS2T2, also induced the biphenyl-degrading genes. However, in these cases the effector molecule was not biphenyl but other aromatic compounds such as ethylbenzene, benzene and so on (Takeda et al., 2010). And the transcriptions of *bpd* genes in another biphenyl/PCB-degrading gram-positive bacteria, *Rhodococcus* sp. strain M5, are regulated by a BpdST two-component system that is induced by the biphenyl (Labbé et al., 1997).

Control of the gene expression

In order to survive under adverse conditions, bacteria transitorily induce the expression of particular genes to deal with environmental stresses. Tandem repeats (TRs) of short nucleotide sequences are often found in the intergenic regions in bacterial genomes (van Belkum et al., 1998). Several studies have reported that the rearrangement of TRs is involved in the gene expression of bacteria at either the transcriptional or translational level (Gemayel et al., 2010; Zhou et al., 2014).

Recently, a novel TR sequence, T(G/A)ACATG(A/C)T, and polymorphisms consisting of a number of repeats located in the region upstream of catechol 2,3-dioxygenase (C23O)-encoding genes were identified by metagenomic analysis (Suenaga et al., 2022). The level of protein expression of C23O dramatically increased as the number of TRs increased, reaching a maximum value with three and four repeats. Experimental results indicated that this nonanucleotide TR would affect the translational efficiency of the gene expression system. A metagenomic sample was collected from activated sludge used to treat industrial wastewater that contained mono- and polycyclic aromatic chemicals (Suenaga et al., 2009). C23O is a key enzyme in the degradation of aromatic compounds because catecholic compounds are the common intermediates in the degradation pathways (Figure 3; Hirose et al., 1994). In this harsh, aromatic-rich environment, the increase in C23O activity realized by adjusting the number of TRs is probably providing the host microbes a survival advantage. In fact, a metagenomic analysis demonstrated that, among the polymorphisms consisting of a number of repeats, tandem repeats of three and four, which indicate higher C23O enzyme activity, dominate in this environment (Suenaga et al., 2022).

Future perspective

In this review, based on the genome sequence of isolated xenobiotic compound- degrading bacteria, the current understanding of the evolutionary mechanisms occurring in degrading gene systems is presented. DNAs of particular interest, such as MGE, within the genomes of these bacterial strains have been a focus of recent studies

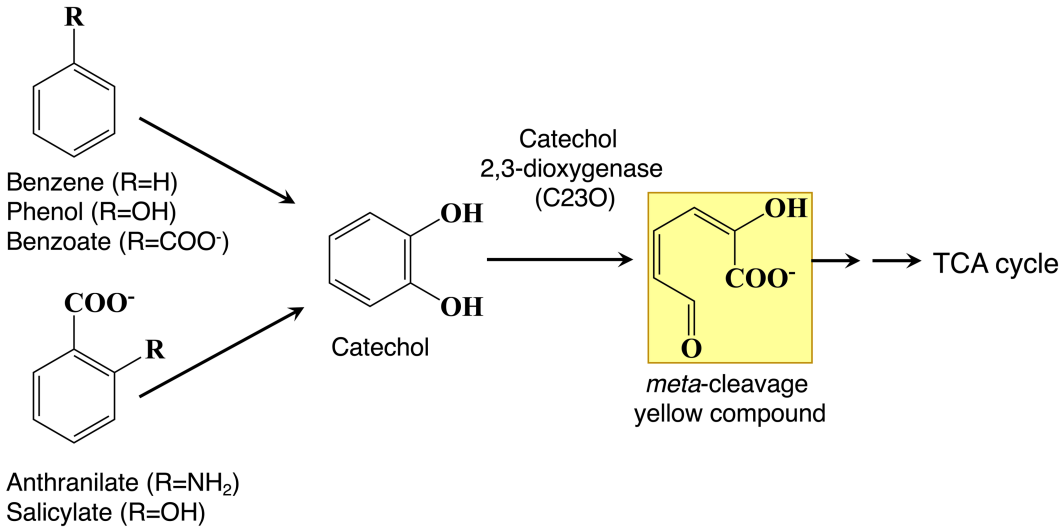


FIGURE 3
Bacterial degradation pathways of aromatic compounds. Aromatics are converted to catechol as a central intermediate. Catechol is cleaved by catechol 2,3-dioxygenase (C23O) to produce *meta*-cleavage yellow compounds, ultimately leading to TCA cycle intermediates.

and have been carefully analyzed. The selective collection of DNA molecules from the environment also seems to be efficient for the study of these evolutionary mechanisms. For instance, plasmids are a relatively undiscovered region of the genome and frequently contain genes that are essential for the survival of the host, such as genes involved in biodegradation and antibiotic resistance. Therefore, a metaplasmid or metamobilome study targeting total plasmid DNA in the environment is a more effective approach for understanding the content and composition of genes in microbial communities for key ecological processes (Suenaga, 2012; Alanin et al., 2021). The comprehensive acquisition of degradative genes by a metaplasmid approach at a single site will contribute to our understanding of their on-site adaptive evolution.

Author contributions

HS drafted the original manuscript. All authors critically reviewed and revised the manuscript draft and approved the final version for submission.

Funding

This work was performed as part of a project supported by JSPS KAKENHI and the Institution for Fermentation, Osaka (IFO).

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.

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.

References

Abraham, W. R., Nogales, B., Golyshin, P. N., Pieper, D. H., and Timmis, K. N. (2002). Polychlorinated biphenyl-degrading microbial communities in soils and sediments. *Curr. Opin. Microbiol.* 5, 246–253. doi: 10.1016/S1369-5274(02)00323-5

Alanin, K. W. S., Jørgensen, T. S., Browne, P. D., Petersen, B., Riber, L., Kot, W., et al. (2021). An improved direct metamobilome approach increases the detection of larger-sized circular elements across kingdoms. *Plasmid* 115:102576. doi: 10.1016/j.plasmid.2021.102576

Aminov, R. I. (2011). Horizontal gene exchange in environmental microbiota. *Front. Microbiol.* 2:158. doi: 10.3389/fmicb.2011.00158

Bellanger, X., Payot, S., Leblond-Bourget, N., and Guédon, G. (2014). Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS Microbiol. Rev.* 38, 720–760. doi: 10.1111/1574-6976.12058

Bosch, R., García-Valdés, E., and Moore, E. R. (1999). Genetic characterization and evolutionary implications of a chromosomally encoded naphthalene-degradation upper pathway from *Pseudomonas stutzeri* AN10. *Gene* 236, 149–157. doi: 10.1016/S0378-1119(99)00241-3

Brunet-Galmés, I., Busquets, A., Peña, A., Gomila, M., Nogales, B., García-Valdés, E., et al. (2012). Complete genome sequence of the naphthalene-degrading bacterium *Pseudomonas stutzeri* AN10 (CCUG 29243). *J. Bacteriol.* 194, 6642–6643. doi: 10.1128/JB.01753-12

Burlage, R. S., Bemis, L. A., Layton, A. C., Sayler, G. S., and Larimer, F. (1990). Comparative genetic organization of incompatibility group P degradative plasmids. *J. Bacteriol.* 172, 6818–6825. doi: 10.1128/jb.172.12.6818-6825.1990

Burrus, V., and Waldor, M. K. (2004). Shaping bacterial genomes with integrative and conjugative elements. *Res. Microbiol.* 155, 376–386. doi: 10.1016/j.resmic.2004.01.012

Chain, P. S., Denef, V. J., Konstantinidis, K. T., Vergez, L. M., Agullo, L., Reyes, V. L., et al. (2006). *Burkholderia xenovorans* LB400 harbors a multi-replicon, 9.73-Mbp

genome shaped for versatility. *Proc. Natl. Acad. Sci. U. S. A.* 103, 15280–15287. doi: 10.1073/pnas.0606924103

Christie, P. J., Whitaker, N., and González-Rivera, C. (2014). Mechanism and structure of the bacterial type IV secretion systems. *Biochim. Biophys. Acta* 1843, 1578–1591. doi: 10.1016/j.bbamcr.2013.12.019

Delavat, F., Miyazaki, R., Carraro, N., Pradervand, N., and van der Meer, J. R. (2017). The hidden life of integrative and conjugative elements. *FEMS Microbiol. Rev.* 41, 512–537. doi: 10.1093/femsre/fux008

Denef, V. J., Park, J., Tsoi, T. V., Rouillard, J. M., Zhang, H., Wibbenmeyer, J. A., et al. (2004). Biphenyl and benzoate metabolism in a genomic context: outlining genome-wide metabolic networks in *Burkholderia xenovorans* LB400. *Appl. Environ. Microbiol.* 70, 4961–4970. doi: 10.1128/aem.70.8.4961-4970.2004

Diaz, E., and Prieto, M. A. A. (2000). Bacterial promoters triggering biodegradation of aromatic pollutants. *Curr. Opin. Biotechnol.* 11, 467–475. doi: 10.1016/S0958-1669(00)00126-9

Fujihara, H., Yamazoe, A., Hosoyama, A., Suenaga, H., Kimura, N., Hirose, J., et al. (2015). Draft genome sequence of *Pseudomonas aeruginosa* KF702 (NBRC 110665), a polychlorinated biphenyl-degrading bacterium isolated from biphenyl-contaminated soil. *Genome Announc.* 3, e00517–e00515. doi: 10.1128/genomeA.00517-15

Fujihara, H., Yoshida, H., Matsunaga, T., Goto, M., and Furukawa, K. (2006). Cross-regulation of biphenyl- and salicylate-catabolic genes by two regulatory systems in *Pseudomonas pseudoalcaligenes* KF707. *J. Bacteriol.* 188, 4690–4697. doi: 10.1128/JB.00329-06

Furukawa, K., and Fujihara, H. (2008). Microbial degradation of polychlorinated biphenyls: biochemical and molecular features. *J. Biosci. Bioeng.* 105, 433–449. doi: 10.1263/jbb.105.433

Furukawa, K., Hayase, N., Taira, K., and Tomizuka, N. (1989). Molecular relationship of chromosomal genes encoding biphenyl/poychlorinated biphenyl catabolism: some soil bacteria possess a highly conserved bph operon. *J. Bacteriol.* 171, 5467–5472. doi: 10.1128/jb.171.10.5467-5472.1989

Gaillard, M., Vallaeys, T., Vorhölter, F. J., Minoia, M., Werlen, C., Sentchilo, V., et al. (2006). The CLC element of Pseudomonas sp. strain B13, a genomic island with various catabolic properties. *J. Bacteriol.* 188, 1999–2013. doi: 10.1128/jb.188.5.1999-2013.2006

Garrido-Sanz, D., Manzano, J., Martín, M., Redondo-Nieto, M., and Rivilla, R. (2018). Metagenomic analysis of a biphenyl-degrading soil bacterial consortium reveals the metabolic roles of specific populations. *Front. Microbiol.* 9:232. doi: 10.3389/fmicb.2018.00232

Garrido-Sanz, D., Sansegundo-Lobato, P., Redondo-Nieto, M., Suman, J., Cajthaml, T., Blanco-Romero, E., et al. (2020). Analysis of the biodegradative and adaptive potential of the novel polychlorinated biphenyl degrader Rhodococcus sp. WAY2 revealed by its complete genome sequence. *Microbial Gemomics* 6:363. doi: 10.1099/mgen.0.000363

Gemayel, R., Vinces, M. D., Legendre, M., and Verstrepen, K. J. (2010). Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annu. Rev. Genet.* 44, 445–477. doi: 10.1146/annurev-genet-072610-155046

Guglielmini, J., Quintais, L., Garcillán-Barcia, M. P., de la Cruz, F., and Rocha, E. P. C. (2011). The repertoire of ICE in prokaryotes underscores the Unity, diversity, and ubiquity of conjugation. *PLoS Genet.* 7:e1002222. doi: 10.1371/journal.pgen.1002222

Heinaru, E., Naanuri, E., Grünbach, M., Jõesaar, M., and Heinaru, A. (2016). Functional redundancy in phenol and toluene degradation in *Pseudomonas stutzeri* strains isolated from the Baltic Sea. *Gene* 589, 90–98. doi: 10.1016/j.gene.2016.05.022

Hirose, J. (2023). Diversity and evolution of integrative and conjugative elements involved in bacterial aromatic compound degradation and their utility in environmental remediation. *Microorganisms* 11:438. doi: 10.3390/microorganisms11020438

Hirose, J., Fujihara, H., Watanabe, T., Kimura, N., Suenaga, H., Futagami, T., et al. (2019). Biphenyl/PCB degrading bph genes of ten bacterial strains isolated from biphenyl-contaminated soil in Kitakyushu, Japan: comparative and dynamic features as integrative conjugative elements (ICEs). *Genes (Basel)* 10:404. doi: 10.3390/genes10050404

Hirose, J., Kimura, N., Suyama, A., Kobayashi, A., Hayashida, S., and Furukawa, K. (1994). Functional and structural relationship of various extradiol aromatic ring-cleavage dioxygenases of Pseudomonas origin. *FEMS Microbiol. Lett.* 118, 273–277. doi: 10.1111/j.1574-6968.1994.tb06840.x

Hirose, J., Watanabe, T., Futagami, T., Fujihara, H., Kimura, N., Suenaga, H., et al. (2021). A new ICEclc subfamily integrative and conjugative element responsible for horizontal transfer of biphenyl and salicylic acid catabolic pathway in the PCB-degrading strain *Pseudomonas stutzeri* KF716. *Microorganisms* 9:2462. doi: 10.3390/microorganisms9122462

Hirose, J., Yamazoe, A., Hosoyama, A., Kimura, N., Suenaga, H., Watanabe, T., et al. (2015a). Draft genome sequence of the polychlorinated biphenyl-degrading bacterium *Comamonas testosteroni* KF712 (NBRC 110673). *Genome Announc.* 3, e01214–e01215. doi: 10.1128/genomeA.01214-15

Hirose, J., Yamazoe, A., Hosoyama, A., Kimura, N., Suenaga, H., Watanabe, T., et al. (2015b). Draft genome sequence of the polychlorinated biphenyl-degrading bacterium *Pseudomonas stutzeri* KF716 (NBRC 110668). *Genome Announc.* 3:e01215. doi: 10.1128/genomeA.01215-15

Ilori, M. O., Picardal, F. W., Aramayo, R., Adebusoye, S. A., Obayori, O. S., and Benedik, M. J. (2015). Catabolic plasmid specifying polychlorinated biphenyl degradation in Cupriavidus sp. strain SK-4: mobilization and expression in a pseudomonad. *J. Basic Microbiol.* 55, 338–345. doi: 10.1002/jobm.201200807

Iram, A., Berenjian, A., and Demirci, A. (2021). A review on the utilization of lignin as a fermentation substrate to produce lignin-modifying enzymes and other value-added products. *Molecules* 26:2960. doi: 10.3390/molecules26102960

Jeffries, T. C., Rayu, S., Nielsen, U. N., Lai, K., Ijaz, A., Nazaries, L., et al. (2018). Metagenomic functional potential predicts degradation rates of a model organophosphorus xenobiotic in pesticide contaminated soils. *Front. Microbiol.* 9:147. doi: 10.3389/fmicb.2018.00147

Johnson, C. M., and Grossman, A. D. (2015). Integrative and conjugative elements (ICEs): what they do and how they work. *Annu. Rev. Genet.* 49, 577–601. doi: 10.1146/annurev-genet-112414-055018

Kimura, N., Watanabe, T., Suenaga, H., Fujihara, H., Futagami, T., Goto, M., et al. (2018). Pseudomonas furukawaii sp. nov., a polychlorinated biphenyl-degrading bacterium isolated from biphenyl-contaminated soil in Japan. *Int. J. Syst. Evol. Microbiol.* 68, 1429–1435. doi: 10.1099/ijsem.0.002670

Labbé, D., Garnon, J., and Lau, P. C. (1997). Characterization of the genes encoding a receptor-like histidine kinase and a cognate response regulator from a biphenyl/polychlorobiphenyl-degrading bacterium, Rhodococcus sp. strain M5. *J. Bacteriol.* 179, 2772–2776. doi: 10.1128/jb.179.8.2772-2776.1997

Merlin, J. W. L. C., Springael, D., Mergeay, M., and Toussaint, A. (1997). Organisation of the bph gene cluster of transposon Tn4371, encoding enzymes for the degradation of biphenyl and 4-chlorobiphenyl compounds. *Mol. Gen. Genet. MGG* 253, 499–506. doi: 10.1007/s004380050349

Miglani, R., Parveen, N., Kumar, A., Ansari, M. A., Khanna, S., Rawat, G., et al. (2022). Degradation of xenobiotic pollutants: an environmentally sustainable approach. *Meta* 12:818. doi: 10.3390/metabo12090818

Mohapatra, B., Malhotra, H., and Phale, P. S. (2022). Life within a contaminated niche: comparative genomic analyses of an integrative conjugative element ICEnahCSV86 and two genomic islands from Pseudomonas bharatica CSV86T suggest probable role in colonization and adaptation. *Front. Microbiol.* 13:928848. doi: 10.3389/fmicb.2022.928848

Mondello, F. J. (1989). Cloning and expression in *Escherichia coli* of Pseudomonas strain LB400 genes encoding polychlorinated biphenyl degradation. *J. Bacteriol.* 171, 1725–1732. doi: 10.1128/jb.171.3.1725-1732.1989

Mukerjee-dhar, G., Shimura, M., Miyazawa, D., Kimbara, K., and Hatta, T. (2005). Bph genes of the thermophilic PCB degrader, Bacillus sp. JF8: characterization of the divergent ring-hydroxylating dioxygenase and hydrolase genes upstream of the Mn-dependent BphC. *Microbiology* 151, 4139–4151. doi: 10.1099/mic.0.28437-0

Nagata, Y., Ohtsubo, Y., and Tsuda, M. (2015). Properties and biotechnological applications of natural and engineered haloalkane dehalogenases. *Appl. Microbiol. Biotechnol.* 99, 9865–9881. doi: 10.1007/s00253-015-6954-x

Newby, D. T., Josephson, K. L., and Pepper, I. L. (2000). Detection and characterization of plasmid pJP4 transfer to indigenous soil bacteria. *Appl. Environ. Microbiol.* 66, 290–296. doi: 10.1128/aem.66.1.290-296.2000

Nojiri, H. (2012). Structural and molecular genetic analyses of the bacterial carbazole degradation system. *Biosci. Biotechnol. Biochem.* 76, 1–18. doi: 10.1271/bbb.110620

Obi, C. C., Vayla, S., de Gannes, V., Berres, M. E., Walker, J., Pavelec, D., et al. (2018). The integrative conjugative element clc (ICEclc) of *Pseudomonas aeruginosa* JB2. *Front. Microbiol.* 9:1532. doi: 10.3389/fmicb.2018.01532

Ohtsubo, Y., Delawary, M., Kimbara, K., Takagi, M., Ohta, A., and Nagata, Y. (2001). BphS, a key transcriptional regulator of bph genes involved in polychlorinated biphenyl/biphenyl degradation in Pseudomonas sp. KKS102. *J. Biol. Chem.* 276, 36146–36154. doi: 10.1074/jbc.M100302200

Ohtsubo, Y., Ishibashi, Y., Naganawa, H., Hirokawa, S., Atobe, S., Nagata, Y., et al. (2012). Conjugal transfer of polychlorinated biphenyl/biphenyl degradation genes in acidovorax sp. strain KKS102, which are located on an integrative and conjugative element. *J. Bacteriol.* 194, 4237–4248. doi: 10.1128/JB.00352-12

Pieper, D. H., and Seeger, M. (2008). Bacterial metabolism of polychlorinated biphenyls. *J. Mol. Microbiol. Biotechnol.* 15, 121–138. doi: 10.1159/000121325

Popowska, M., and Krawczyk-Balska, A. (2013). Broad-host-range IncP-1 plasmids and their resistance potential. *Front. Microbiol.* 4:44. doi: 10.3389/fmicb.2013.00044

Redondo-Salvo, S., Fernández-López, R., Ruiz, R., Vielva, L., de Toro, M., Rocha, E. P. C., et al. (2020). Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. *Nat. Commun.* 11:3602. doi: 10.1038/s41467-020-17278-2

Ridl, J., Suman, J., Fraraccio, S., Hradilova, M., Strejcek, M., Cajthaml, T., et al. (2018). Complete genome sequence of *Pseudomonas alcaliphila* JAB1 (=DSM 26533), a versatile degrader of organic pollutants. *Stand. Genomic Sci.* 13:3. doi: 10.1186/s40793-017-0306-7

Rodríguez-Beltrán, J., Sørum, V., Toll-Riera, M., de la Vega, C., Peña-Miller, R., and San Millán, Á. (2020). Genetic dominance governs the evolution and spread of mobile genetic elements in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 117, 15755–15762. doi: 10.1073/pnas.2001240117

Ross, K., Varani, A. M., Snesrud, E., Huang, H., Alvarenga, D. O., Zhang, J., et al. (2021). TnCentral: a prokaryotic transposable element database and web portal for transposon analysis. *mBio* 12, e02060–e02021. doi: 10.1128/mBio.02060-21

Satola, B., Wübbeler, J. H., and Steinbüchel, A. (2013). Metabolic characteristics of the species *Variovorax paradoxus*. *Appl. Microbiol. Biotechnol.* 97, 541–560. doi: 10.1007/s00253-012-4585-z

Shimizu, S., Kobayashi, H., Masai, E., and Fukuda, M. (2001). Characterization of the 450-kb linear plasmid in a polychlorinated biphenyl degrader, *Rhodococcus* sp. strain RHA1. *Appl. Environ. Microbiol.* 67, 2021–2028. doi: 10.1128/AEM.67.5.2021-2028.2001

Shintani, M., Sanchez, Z. K., and Kimbara, K. (2015). Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. *Front. Microbiol.* 6, 1–16. doi: 10.3389/fmicb.2015.00242

Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P., and de la Cruz, F. (2010). Mobility of plasmids. *Microbiol. Mol. Biol. Rev.* 74, 434–452. doi: 10.1128/mmbr.00020-10

Sota, M., Yano, H., Ono, A., Miyazaki, R., Ishii, H., Genka, H., et al. (2006). Genomic and functional analysis of the IncP-9 naphthalene-catabolic plasmid NAH7 and its transposon Tn4655 suggests catabolic gene spread by a tyrosine recombinase. *J. Bacteriol.* 188, 4057–4067. doi: 10.1128/JB.00185-06

Springael, D., Kreps, S., and Mergeay, M. (1993). Identification of a catabolic transposon, Tn4371, carrying biphenyl and 4-chlorobiphenyl degradation genes in *Alcaligenes eutrophus* A5. *J. Bacteriol.* 175, 1674–1681. doi: 10.1128/jb.175.6.1674-1681

Springael, D., and Top, E. M. (2004). Horizontal gene transfer and microbial adaptation to xenobiotics: new types of mobile genetic elements and lessons from ecological studies. *Trends Microbiol.* 12, 53–58. doi: 10.1016/j.tim.2003.12.010

Suenaga, H. (2012). Targeted metagenomics: a high-resolution metagenomics approach for specific gene clusters in complex microbial communities. *Environ. Microbiol.* 14, 13–22. doi: 10.1111/j.1462-2920.2011.02438.x

Suenaga, H., Fujihara, H., Kimura, N., Hirose, J., Watanabe, T., Futagami, T., et al. (2017). Insights into the genomic plasticity of *Pseudomonas putida* KF715, a strain with unique biphenyl-utilizing activity and genome instability properties. *Environ. Microbiol. Rep.* 9, 589–598. doi: 10.1111/1758-2229.12561

Suenaga, H., Koyama, Y., Miyakoshi, M., Miyazaki, R., Yano, H., Sota, M., et al. (2009). Novel organization of aromatic degradation pathway genes in a microbial community as revealed by metagenomic analysis. *ISME J.* 3, 1335–1348. doi: 10.1038/ismej.2009.76

Suenaga, H., Matsuzawa, T., and Sahara, T. (2022). Discovery by metagenomics of a functional tandem repeat sequence that controls gene expression in bacteria. *FEMS Microbiol. Ecol.* 98:37. doi: 10.1093/femsec/fiac037

Suenaga, H., Yamazoe, A., Hosoyama, A., Kimura, N., Hirose, J., Watanabe, T., et al. (2015). Draft genome sequence of the polychlorinated biphenyl-degrading bacterium *Cupriavidus basilensis* KF708 (NBRC 110671) isolated from biphenyl-contaminated soil. *Genome Announc.* 3, e00143–e00115. doi: 10.1128/genomeA.00143-15

Taguchi, K., Motoyama, M., Iida, T., and Kudo, T. (2007). Polychlorinated biphenyl/biphenyl degrading gene clusters in Rhodococcus sp. K37, HA99, and TA431 are different from well-known bph gene clusters of rhodococci. *Biosci. Biotechnol. Biochem.* 71, 1136–1144. doi: 10.1271/bbb.60551

Taguchi, K., Motoyama, M., and Kudo, T. (2004). Multiplicity of 2,3-dihydroxybiphenyl dioxygenase genes in the gram-positive polychlorinated biphenyl degrading bacterium *Rhodococcus rhodochrous* K37. *Biosci. Biotechnol. Biochem.* 68, 787–795. doi: 10.1271/bbb.68.787

Takeda, H., Shimodaira, J., Yukawa, K., Hara, N., Kasai, D., Miyauchi, K., et al. (2010). Dual two-component regulatory systems are involved in aromatic compound degradation in a polychlorinated-biphenyl degrader, *Rhodococcus jostii* RHA1. *J. Bacteriol.* 192, 4741–4751. doi: 10.1128/JB.00429-10

Takeda, H., Yamada, A., Miyauchi, K., Masai, E., and Fukuda, M. (2004). Characterization of transcriptional regulatory genes for biphenyl degradation in Rhodococcus sp. strain RHA1. *J. Bacteriol.* 186, 2134–2146. doi: 10.1128/JB.186.7.2134-2146.2004

Tan, H. M. (1999). Bacterial catabolic transposons. *Appl. Microbiol. Biotechnol.* 51, 1–12. doi: 10.1007/s002530051356

Tang, H., Yu, H., Li, Q., Wang, X., Gai, Z., Yin, G., et al. (2011). Genome sequence of *Pseudomonas putida* strain B6-2, a Superdegrader of polycyclic aromatic hydrocarbons and dioxin-like compounds. *J. Bacteriol.* 193, 6789–6790. doi: 10.1128/JB.06201-11

Top, E. M., and Springael, D. (2003). The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. *Curr. Opin. Biotechnol.* 14, 262–269. doi: 10.1016/s0958-1669(03)00066-1

Top, E. M., Springael, D., and Boon, N. (2002). Catabolic mobile genetic elements and their potential use in bioaugmentation of polluted soils and waters. *FEMS Microbiol. Ecol.* 42, 199–208. doi: 10.1111/j.1574-6941.2002.tb01009.x

Toussaint, A., Merlin, C., Benotmane, M. A., Monchy, S., Benotmane, M. A., Leplae, R., et al. (2003). The biphenyl- and 4-chlorobiphenyl-catabolic transposon Tn4371, a member of a new family of genomic islands related to IncP and Ti plasmids. *Appl. Environ. Microbiol.* 69, 4837–4845. doi: 10.1128/AEM.69.8.4837-4845.2003

Tropel, D., and van der Meer, J. R. (2004). Bacterial transcriptional regulators for degradation pathways of aromatic compounds. *Microbiol. Mol. Biol. Rev.* 68, 474–500. doi: 10.1128/MMBR.68.3.474-500.2004

Tsuda, M., Tan, H. M., Nishi, A., and Furukawa, K. (1999). Mobile catabolic genes in bacteria. *J. Biosci. Bioeng.* 87, 401–410. doi: 10.1016/S1389-1723(99)80086-3

van Belkum, A., Scherer, S., van Alphen, L., and Verbrugh, H. (1998). Short-sequence DNA repeats in prokaryotic genomes. *Microbiol. Mol. Biol. Rev.* 62, 275–293. doi: 10.1128/MMBR.62.2.275-293.1998

Van der Meer, J. R., Alexander, J. B., Zehnder, A. J. B., and De Vos, W. M. (1991). Identification of a novel composite transposable element, Tn5280, carrying chlorobenzene dioxygenase genes of Pseudomonas sp. strain P51. *J. Bacterol.* 173, 7077–7083. doi: 10.1128/jb.173.22.7077-7083.1991

Van der Meer, J. R., De Vos, W. M., Harayama, S., and Zehnder, A. J. B. (1992). Molecular mechanisms of genetic adaptation to xenobiotic compounds. *Microbiol. Rev.* 56, 677–694. doi: 10.1128/mmbr.56.4.677-694.1992

Watanabe, T., Fujihara, H., and Furukawa, K. (2003). Characterization of the second LysR-type regulator in the biphenyl-catabolic gene cluster of *Pseudomonas pseudoalcaligenes* KF707. *J. Bacteriol.* 185, 3575–3582. doi: 10.1128/JB.185.12.3575-3582.2003

Watanabe, T., Inoue, R., Kimura, N., and Furukawa, K. (2000). Versatile transcription of biphenyl catabolic bphOperon in *Pseudomonas pseudoalcaligenes* KF707*. *J. Biol. Chem.* 275, 31016–31023. doi: 10.1074/jbc.M003023200

Witzig, R., Aly, H. A., Strompl, C., Wray, V., Junca, H., and Pieper, D. H. (2007). Molecular detection and diversity of novel diterpenoid dioxygenase DitA1 genes from proteobacterial strains and soil samples. *Environ. Microbiol.* 9, 1202–1218. doi: 10.1111/j.1462-2920.2007.01242.x

Wozniak, R. A. F., and Waldor, M. K. (2010). Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat. Rev. Microbiol.* 8, 552–563. doi: 10.1038/nrmicro2382

Zamarro, M. T., Martín-Moldes, Z., and Díaz, E. (2016). The ICEXTD of Azoarcus sp. CIB, an integrative and conjugative element with aerobic and anaerobic catabolic properties. *Environ. Microbiol.* 18, 5018–5031. doi: 10.1111/1462-2920.13465

Zhou, K., Aertsen, A., and Michiels, C. W. (2014). The role of variable DNA tandem repeats in bacterial adaptation. *FEMS Microbiol. Rev.* 38, 119–141. doi: 10.1111/1574-6976.12036

Zubrova, A., Michalikova, K., Semerad, J., Strejcek, M., Cajthaml, T., Suman, J., et al. (2021). Biphenyl 2,3-dioxygenase in *Pseudomonas alcaliphila* JAB1 is both induced by phenolics and monoterpenes and involved in their transformation. *Front. Microbiol.* 12:657311. doi: 10.3389/fmicb.2021.657311

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 29 June 2023
DOI 10.3389/fmicb.2023.1216714

OPEN ACCESS

EDITED BY
Feng Gao,
Tianjin University, China

REVIEWED BY
Paul Greenfield,
Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia
Stefano Ghignone,
National Research Council (CNR), Italy

*CORRESPONDENCE
Chang-Hong Liu
✉ chliu@nju.edu.cn

RECEIVED 04 May 2023
ACCEPTED 09 June 2023
PUBLISHED 29 June 2023

CITATION
Jiang J-P, Liu X, Liao Y-F, Shan J, Zhu Y-P and Liu C-H (2023) Genomic insights into *Aspergillus sydowii* 29R-4-F02: unraveling adaptive mechanisms in subseafloor coal-bearing sediment environments.
Front. Microbiol. 14:1216714.
doi: 10.3389/fmicb.2023.1216714

Genomic insights into *Aspergillus sydowii* 29R-4-F02: unraveling adaptive mechanisms in subseafloor coal-bearing sediment environments

Jun-Peng Jiang[1], Xuan Liu[1], Yi-Fan Liao[1], Jun Shan[2], Yu-Ping Zhu[1] and Chang-Hong Liu[1]*

[1]State Key Laboratory of Pharmaceutical Biotechnology, Nanjing University, Nanjing, China, [2]State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing, China

Introduction: *Aspergillus sydowii* is an important filamentous fungus that inhabits diverse environments. However, investigations on the biology and genetics of *A. sydowii* in subseafloor sediments remain limited.

Methods: Here, we performed *de novo* sequencing and assembly of the *A. sydowii* 29R-4-F02 genome, an isolate obtained from approximately 2.4 km deep, 20-million-year-old coal-bearing sediments beneath the seafloor by employing the Nanopore sequencing platform.

Results and Discussion: The generated genome was 37.19Mb with GC content of 50.05%. The final assembly consisted of 11 contigs with N_{50} of 4.6Mb, encoding 12,488 putative genes. Notably, the subseafloor strain 29R-4-F02 showed a higher number of carbohydrate-active enzymes (CAZymes) and distinct genes related to vesicular fusion and autophagy compared to the terrestrial strain CBS593.65. Furthermore, 257 positively selected genes, including those involved in DNA repair and CAZymes were identified in subseafloor strain 29R-4-F02. These findings suggest that *A. sydowii* possesses a unique genetic repertoire enabling its survival in the extreme subseafloor environments over tens of millions of years.

KEYWORDS

***Aspergillus sydowii*, subseafloor sediments, fungi, CAZymes, adaptive evolution**

Introduction

Ocean Drilling Program revealed a new ecosystem composed of microorganisms in subseafloor sediments (D'Hondt et al., 2004; Schrenk et al., 2010; Lomstein et al., 2012), which vertically extends for ~2.5 km below the seafloor (Selbmann et al., 2013; Inagaki et al., 2015). This deep subseafloor biosphere harbors organisms from three major domains of life: Archaea (Orsi et al., 2013), Bacteria (Schippers et al., 2005), and Eukarya (Borgonie et al., 2015). Fungi, as the dominant component of eukaryotes, also widely exist in the subseafloor sediments (Rédou et al., 2015; Nagano et al., 2016; Liu et al., 2017). Most subseafloor fungi were affiliated with the phyla Ascomycota, Basidiomycota and Chytridiomycota (Nagano and Nagahama, 2012). Phylogenetic analysis indicated that subseafloor fungi were closely related to the isolates from terrestrial, freshwater and marine environments (Liu et al., 2017; Quemener et al., 2020). Various culture-dependent and culture-independent investigations show that fungi are active and well-adapted to subseafloor environments (Rédou et al., 2015; Liu et al., 2017). Some fungi, such as *Aquamyces* sp.

and *Orbilia* sp., can live in a saprophytic mode and maintain their life in anaerobic habitats by degrading organic substances (Ortega-Arbulu et al., 2019). Our previous studies have shown that *Schizophyllum commune* isolated from ~2 km coal-bearing sediment below seafloor has a variety of anaerobic survival strategies, including improving the metabolic efficiency of ethanol and amino acids, increasing the number of mitochondria in cells, and forming autophagosomes, but it cannot complete sexual propagation (Zain Ul Arifeen et al., 2021a,b). Genome analysis of *S. commune* revealed significant expansion of FunK1 protein kinase, the NmrA family, and transposon, but significant reduction of nucleotide diversity, substitution rates, and homologous recombination, reflecting the unique genetic evolution of fungi to maintain their long-term life in subseafloor environments (Liu et al., 2022). However, due to the limited number of pure culture strains of subseafloor fungi, the genomes of only a few fungi have been sequenced, and there is still a lack of understanding of the genetic characteristics of the fungi in deep biosphere.

Aspergillus is one of the best-studied filamentous fungi, mainly because some species of this genus are widely used in the fields of medicine (e.g., *A. fumigatus*, *A. terreus*), food spoilage (e.g., *A. flavus*, *A. parasiticus*, and *A. hancockii*), and industry (e.g., *A. niger*, *A. aculeatus*, and *A. oryzae*; de Vries et al., 2017; Pitt et al., 2017). *A. sydowii*, in particular, is distributed throughout the world and can survive as saprophytes (Raper and Fennell, 1965; Geiser et al., 1998; Klich, 2002). Our previous study found that this fungus is also one of the dominant fungi in ~2 km sediments below the seafloor (Liu et al., 2017). Some strains of *A. sydowii* isolated from marine ecosystems can cause an epizootic infection of sea fan corals (Kim and Harvell, 2004; Hayashi et al., 2016), while others can produce a series of secondary metabolites with novel structure and bioactivities (Wang et al., 2019; Niu et al., 2020), and degrade the refractory substances such as xylan (Brandt et al., 2020), lignocellulose (Cong et al., 2017), and waste engine oil (Kumar et al., 2021). Given the wide distribution and important application value of *A. sydowii*, five strains of this species have been sequenced, assembled and annotated (de Vries et al., 2017; Brandt et al., 2021; Kumar et al., 2021; Jing et al., 2022), of which strain CBS593.65 had been fully studied. Of the sequenced strains, AS31, AS42, Fsh102, CBS593.65, and BOBA1 were isolated from terrestrial land and deep sea, respectively. However, the genome of any *A. sydowii* strains derived from subseafloor sediments is not yet available.

In order to acquire insight into the survival and environmental adaptation mechanism of subseafloor fungi, we performed the *de novo* whole genome assembly and comprehensive analysis of *A. sydowii* 29R-4-F02 buried in 2,405 m coal-bearing sediment below the seafloor. A large number of genes were identified to be involved in the secondary metabolism and carbohydrate metabolism. The carbohydrate-active enzymes (CAZymes) repertoire of *A. sydowii* 29R-4-F02 was identified and compared with that of other fungi. Moreover, genes or gene families were investigated through comparative genomes to reveal the special genetic evolution process of the strain.

Materials and methods

Fungal strain and culture conditions

A. sydowii 29R-4-F02 (CCTCC AF 2022080) strain was isolated from ~2.4 km coal-bearing sediments below the seafloor at drilling Site C0020 (41°10′35″N, 142°12′01″E) off the Shimokita Peninsula, Japan (Liu et al., 2017) and maintained aerobically on potato dextrose agar (PDA, 200 g/L potato, 20 g/L glucose, and 15 g/L agar) slants at 4°C. Details of the habitat and culture conditions of *A. sydowii* 29R-4-F02 have been described previously (Liu et al., 2017). For DNA isolation, mycelium of *A. sydowii* 29R-4-F02 was grown in 250 ml conical flask containing 100 ml PD (200 g/L potato and 20 g/L glucose) and incubated at 30°C and 200 rpm for 2–3 days.

Genome sequencing and assembly

DNA extraction was carried out by HP fungal DNA mini kit (Omega Bio-Tek, GA, USA) according to the manufacturer's instructions. The DNA concentration (>50 ng/μl), quality and integrity were determined by using a Qubit Flurometer (Invitrogen, USA) and a NanoDrop Spectrophotometer (Thermo Scientific, USA). Genome sequencing was performed using the Nanopore sequencing platform at the Beijing Novogene Bioinformatics Technology Co., Ltd. After filtering out low quality reads (less than 500 bp) by Nanoplot v 1.41.0 software, a total of 441,200 bases reads was retained (De Coster et al., 2018). The clean reads were mapped with each other to find the overlap between sequences by Minimap2 v 2.26 software (Li, 2018), and then assembled by Miniasm software to constructed scaffolds (Li, 2016). Finally, the Racon v 1.5.0 software was used to constructed consensus sequence and correct the assembly results (Vaser et al., 2017). BUSCO v 5.4.7 was used to assess the completeness of gene regions using the dataset fungi_odb9 (Waterhouse et al., 2017). Quast v 5.1.0 software was further used to evaluate the assembly quality of *A. sydowii* 29R-4-F02 genome with the default parameters (Gurevich et al., 2013). Reference-based assembly of the *A. sydowii* 29R-4-F02 mitochondrial genome was conducted using the Rebaler v 0.2.0 software (Baeza and García-De León, 2022). We executed Rebaler using *A. nidulans* mitochondrial genome as reference genome. Also, we ran Rebaler with the option "circular = true" indicating that the reference genome was circular so that Rebaler "rotated" contigs between polishing rounds to ensure improved accuracy of the final assembled mitochondrial genome.

Genome annotation

After obtaining the whole-genome data of *A. sydowii* 29R-4-F02, we used two strategies to annotated genes: (i) Augustus v 2.7 was used for *de novo* prediction (Stanke et al., 2008); (ii) Homologous species was predicted in Genewise v 2.4.1 using the reference gene models of *A. sydowii* CBS593.65 (Birney et al., 2004). Gene models from these two approaches were combined using the EVidenceModeler v 2.1.0 and updated by PASA v2.4.1 (Haas et al., 2008). Dispersed repeat sequences (DRs) and tandem repeat sequences (TRs) were identified by RepeatMasker v open-4.0.5 (Saha et al., 2008) and TandemRepeat v 4.07b Finder (TRF), respectively (Benson, 1999). The tRNA and rRNA were predicted by tRNAscan-SE v 1.3.1 (Chan et al., 2021) and rRNAmmer v 1.2 (Lagesen et al., 2007). The sRNA, snRNA and miRNA were first performed Rfam database v 31.0 (Gardner et al., 2009) comparison annotation, and then use the cmsearch program (v 1.1rc4; default parameter) to determine the final sRNA, snRNA and miRNA.

Gene functions were annotated based on Gene Ontology (GO, v 20171011), Kyoto Encyclopedia of Genes and Genomes (KEGG, v

20181107), Eukaryotic Clusters of Orthologous Groups (KOG, v 201711), Non-redundant protein database (Nr, v 201703), PFAM databases (v 31.0), SwissProt (v 20180716), and Carbohydrate-Active enZYmes Database (v 20181107; Wang et al., 2021; Liu et al., 2022). The secreted proteins were predicted by the SingalP 4.1 and TMHMM 2.0c (Petersen et al., 2011) signal database. Meanwhile, fungal secondary metabolite pathways were predicted using the online tool AntiSMASH v 4.0.2 (Medema et al., 2011). The cytochrome P450 gene family were predicted using Cytochrome P450 Database (Fischer et al., 2007).

Phylogenetic and synteny analyses

A total of 21 fungal species (i.e., 22 strains) of *Aspergillus* with whole genome sequence, including *A. sydowii* (two strains 29R-4-F02 and CBS593.65), *A. aculeatus*, *A. brasiliensis*, *A. carbonarius*, *A. clavatus*, *A. fischeri*, *A. flavus*, *A. fumigatus*, *A. glaucus*, *A. homomorphus*, *A. ibericus*, *A. luchuensis*, *A. nidulans*, *A. niger*, *A. oryzae*, *A. phoenicis*, *A. terreus*, *A. tubingensis*, *A. versicolor*, *A. wentii*, and *A. zonatus* were used to perform gene family analysis. Except *A. sydowii* 29R-4-F02, protein sequences of these species were downloaded from NCBI and JGI databases, and analyzed by OrthoFinder with default parameters to find orthologous genes families (Emms and Kelly, 2015).

The single copy orthologous genes of the 22 analyzed strains were aligned using MAFFT and create a super-gene for each species (Katoh and Standley, 2013). After Gblocks alignment optimization (Castresana, 2000), the conserved blocks of super-genes were used for phylogenetic tree construction using RAxML (Stamatakis, 2014). The final phylogenetic tree was visualized and edited in iTOL9 (Letunic and Bork, 2016).

In addition, synteny analysis of strains 29R-4-F02 and CBS593.65 was performed using MUMmer sequence alignment package (Delcher et al., 2002). Conserved syntenic blocks between the two strains were identified using the MCScanX package with default parameters based on a minimum requirement of five collinear orthologous genes (Wang et al., 2012). Gene Ontology (GO) enrichment was analyzed by Cytoscape software.

Identification and classification of CAZymes families

For comparative analysis of the CAZymes, HmmScan in the HMMER v 3.1b2 package (Eddy, 1998) was used to search protein sequences of *A. sydowii* 29R-4-F02 against the family-specific HMM profiles of CAZymes from dbCAN database v 9 (Yin et al., 2012) to detect putative CAZymes. The identified CAZymes were then classified into glycoside hydrolases (GHs), glycosyltransferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), auxiliary activities (AAs), and carbohydrate-binding modules (CBMs) base on the CAZy database (Lombard et al., 2014).

Positively selected genes

The positive selected genes (PSGs) between *A. sydowii* 29R-4-F02 and *A. sydowii* CBS593.65 were detected using the KaKs_Calculator v 2.0 software with the default parameters (Zhang et al., 2006).

Results

Genome sequencing and assembly

The genome of *A. sydowii* 29R-4-F02 was sequenced using Nanopore sequencing platform. After filtering out low-quality reads, over 3.11 Gb reads (~83.62×) were assembled into 11 contigs with a contig N50 of 4.6 Mb, and the GC content of 50.05% (Figure 1; Table 1), resulting in a total assembly size of 37.19 Mb that was consistent with the estimated genome size of 29.00–39.00 Mb and GC content of 48.00–53.00% of different *Aspergillus* isolates (de Vries et al., 2017; Kumar et al., 2021). The integrity of the gene region was evaluated using BUSCO, which showed high quality of sequence assembly, with 92.8% of single copies intact and only 3.7% missing. Quast analysis showed that the genomic quality of strain 29R-4-F02 was superior to that of strains Fsh102 and BOBA1 (Supplementary Figure 1), and comparable to that of strain CBS593.65 (de Vries et al., 2017). 86.10% of genes of strain 29R-4-F02 were collinear with those of strain CBS593.65 (Supplementary Figure 2). In addition, we assembled the mitochondrial genome of strain 29R-4-F02, which has a length of 26,185 bp and contains 13 genes, including 7 NADH–ubiquinone oxidoreductase genes, 3 cytochrome c oxidase genes, 1 cytochrome b gene, 1 ribosomal protein gene, and 1 ATP synthase gene, with a GC content of 23.67%. The size and number of genes in the mitochondrial genome of strain 29R-4-F02 are similar to those of other *Aspergillus* species, with genome sizes ranging from 27,817 to 77,649 bp and gene numbers ranging from 14 to 21 (Zhao et al., 2012; Takahashi et al., 2021; Jing et al., 2022).

Genome and functional annotation

A total of 12,488 protein-coding genes were predicted with an average sequence length of 1,567 bp and a cumulative length of 19.6 Mb, accounting for 52.62% of the total genome sequence. The total length of repeat sequences was about 0.76 Mb, accounting for 2.06% of the genomic size. Among the repetitive elements, tandem and interspersed repeats account for 0.74 and 1.31%, respectively. Moreover, 160 tRNA genes, 35 sRNA genes, 15 snRNA genes, and 48 rRNA genes were predicted in the genome (Supplementary Table 1). Additionally, functional annotation of the predicted genes was performed based on the public database of Nr, SwissProt, KEGG, KOG, GO, and Pfam, protein databases (Supplementary Table 2).

Based on the GO database, 68.51% of protein-coding genes (8,555) were annotated into three categories: biological process, cellular components, and molecular function (Figure 2; Supplementary Table 2). For biological processes, genes involved in "metabolic process" (4,783) are the most abundant, followed by "cellular process" (4,637), "localization" (1,606), "establishment of localization" (1,570), "biological regulation" (1,150), and "regulation of biological process" (1,150). Among the genes involved in cellular components, 3,368, 3,368, 1,453, and 712 were associated with "cell," "cell parts," "organelle," and "macromolecular complex," respectively. Furthermore, for genes related to molecular function, 4,585, 4,487, 769, and 531 were involved in "binding," "catalytic activity," "transporter activity," and "nucleic acid binding transcription factor activity," respectively. These results revealed that the subseafloor *A. sydowii* 29R-4-F02 had abundant genes related to metabolic

10.3389/fmicb.2023.1216714

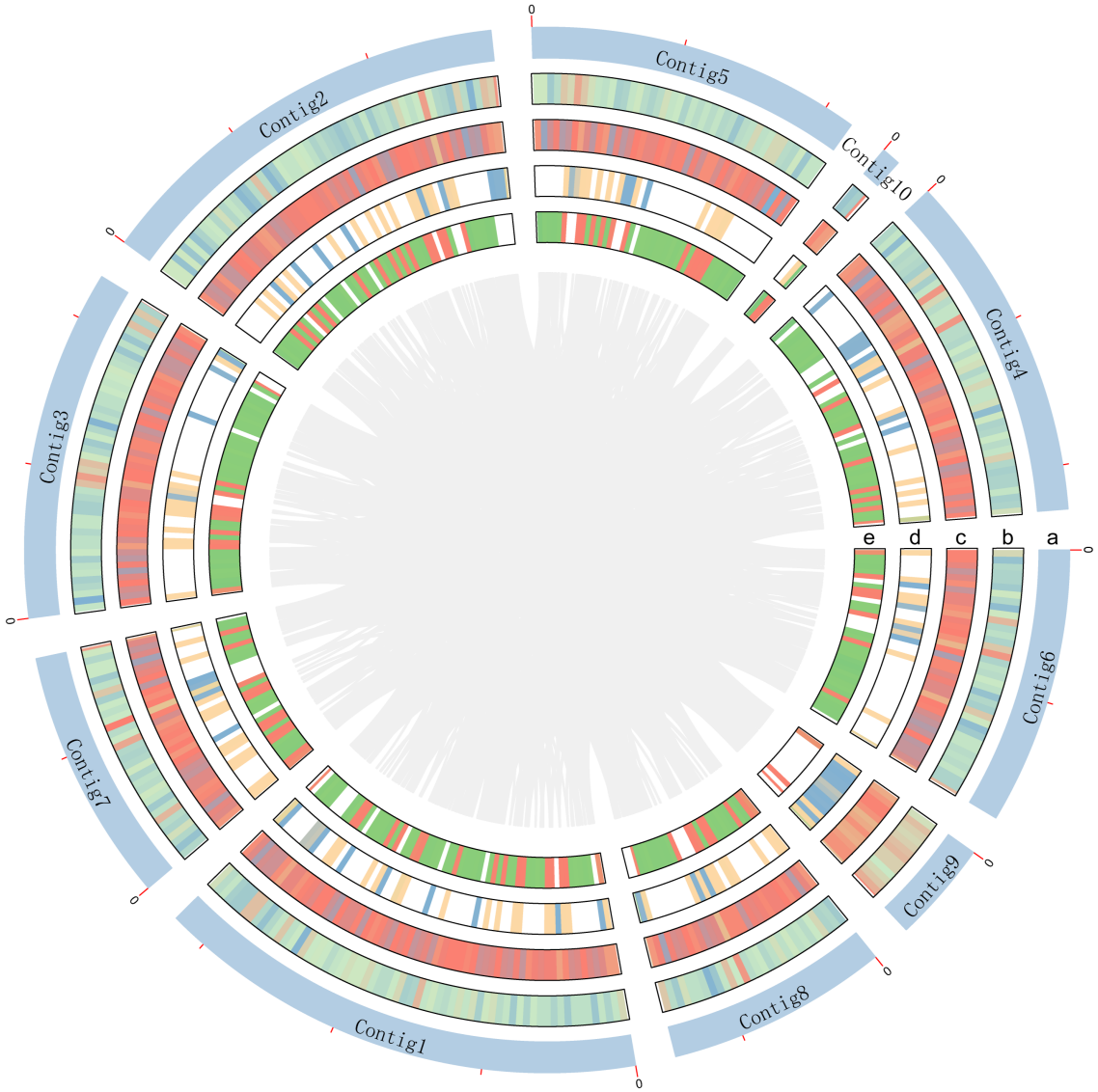


FIGURE 1
Schematic representation of genomic characteristics of *A. sydowii* 29R-4-F02. Circle a represents contigs, but excluding contig11 without annotation gene. Circle b and c are CDS on the positive and negative chains. Circle d and e are gene density on the positive and negative chains.

activities. Additionally, 2,237 protein-coding genes (17.91%) were annotated based on the KOG database (Figure 3; Supplementary Table 2). The majority of the genes were associated with "general function prediction only" (287), followed by "Energy production and conversion," "post-translational modification, protein turnover, chaperones," "Amino acid transport and metabolism," and "Translation, ribosomal structure and biogenesis." These findings indicated that *A. sydowii* 29R-4-F02 has a variety of protein metabolism and energy metabolism processes, which may help the fungus to better absorb and utilize nutrients from the subseafloor environment and maintain long-term survival. To further investigate the gene functions in *A. sydowii* 29R-4-F02, 10,893 (87.22%) protein-coding genes were assigned to their orthologs in the KEGG database. The KEGG function classification was shown in Figure 4 and Supplementary Table 2, including Cellular Processes (460), Environmental Information Processing (230), Genetic Information Processing (649), Human Diseases (578), Metabolism (1,756), and Organismal Systems (428). Consistent with the KOG annotation, metabolism and biosynthesis categories in KEGG were significantly annotated.

A total of 52 gene clusters involved in the secondary metabolism, 231 cytochrome P450 genes, and 920 secreted proteins were identified in *A. sydowii* 29R-4-F02 genome (Figure 5; Supplementary Table 3). Among these P450 genes (Figure 5A), 13 were classified as P450, CYP52 class, 9 as E−class P450, CYP2D class, 9 as E−class P450, CYP3A class, 1 as E−class P450, CYP2A class, 93 as E−class P450, group I class, 27 as E−class P450, group IV class, 15 as Pisatin demethylase−like class, 1 as B−class P450 class, 7 as Cytochrome P450 class, and 56 as undetermined class. The E-class P450, group I class encoded by the most genes are involved in the oxidation–reduction reactions, while 56 undetermined class were predicted to participate in various secondary metabolic processes, such as

TABLE 1 Genome assembly and annotation summary of *A. sydowii* strains.

Assembly feature	29R-4-F02	CBS593.65	Fsh102	BOBA1
Genome size (Mb)	37,191,710	34,381,026	35,409,723	38,795,664
Coverage (X)	83.6×	95.1×	88.65×	100.0×
Max Length (bp)	6,486,838	5,537,812	860,766	353,923
Number of Scaffold	11	97	474	2,582
Scaffold N50	4,575,881	2,288,531	260,502	31,598
Scaffold L50	4	5	40	330
GC content (%)	50.05	49.98	50.5	52.18
Gene number	12,488	13,717	10,761	18,932
Reference	In this study	de Vries et al. (2017)	Brandt et al. (2021)	Kumar et al. (2021)

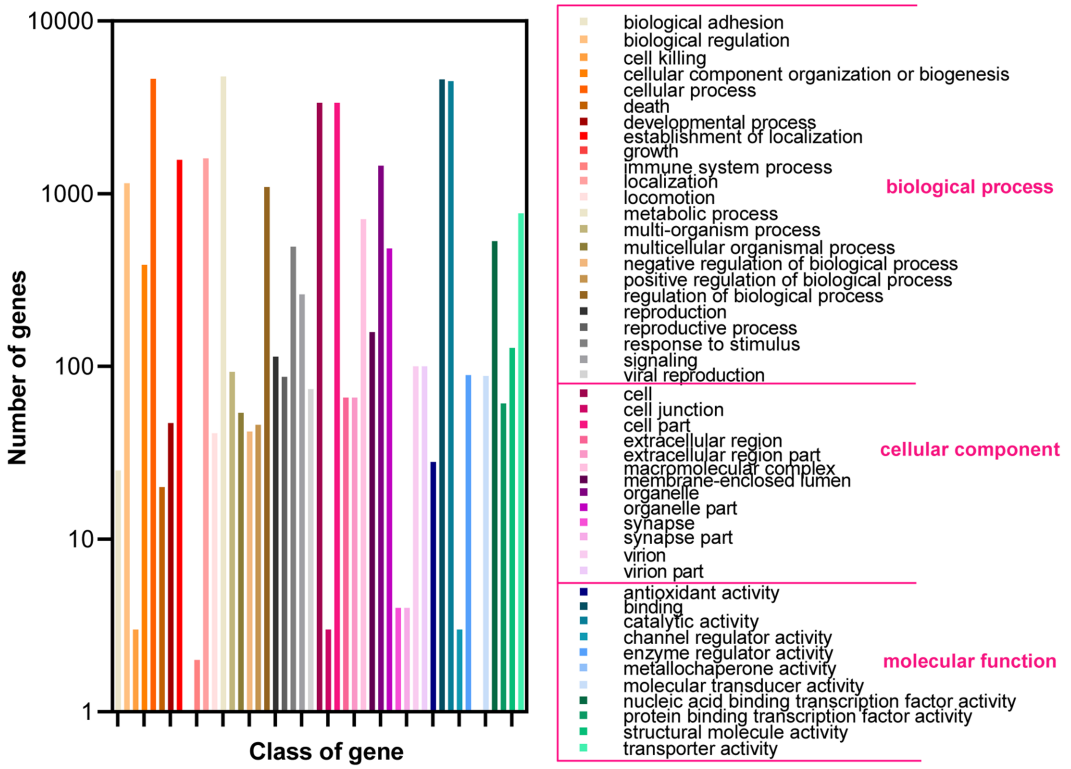


FIGURE 2
Functional annotation of *A. sydowii* 29R-4-F02 in Gene Ontology (GO) database.

xenobiotic biodegradation and carbohydrate metabolism. Similarly, among the secondary metabolite gene clusters (Figure 5B), 12 were classified as T1pks, 10 as Nrps, 8 as Terpenes, 6 as Indole, 4 as T1pks-Nrps, 2 as T1pks-Terpene, 1 as Indole-T1pks, and 9 others. These clusters contain whole or partial genes from known clusters such as Indole-T1pks and T1pks-Terpene cluster from *Penicillium digitatum* (CBS130527; Wang et al., 2021), T1pks and Terpene cluster from *Calcarisporium* sp. (KF525; Kumar et al., 2018), Nrps cluster from *Aspergillus hancockii* (FRR3425; Pitt et al., 2017), T1pks-Nrps cluster from *Aspergillus westerdijkiae* (CBS112803; Han et al., 2016), and T1pks-Nrps cluster from *Pestalotiopsis* sp. (KF079; Kumar et al., 2018), suggesting that *A. sydowii* 29R-4-F02 is capable of synthesizing these metabolites. These genes play a crucial role in fungi adapting to extreme environments and maintaining population competitive advantages (Choi et al., 2010; Cresnar and Petric, 2011; Keller, 2019; Yurchenko et al., 2021).

The CAZyme family

CAZymes play important roles in degrading cellulose, hemicellulose, pectin, and lignin polysaccharides and are responsible for the acquisition of nutrients for fungi (de Vries et al., 2017). In *A. sydowii* 29R-4-F02 genome, 670 CAZyme coding genes were identified, including 337 encoding genes for GHs, 78 encoding genes for AAs, 126 encoding genes for GTs, 39 encoding genes for CEs, 67 encoding genes

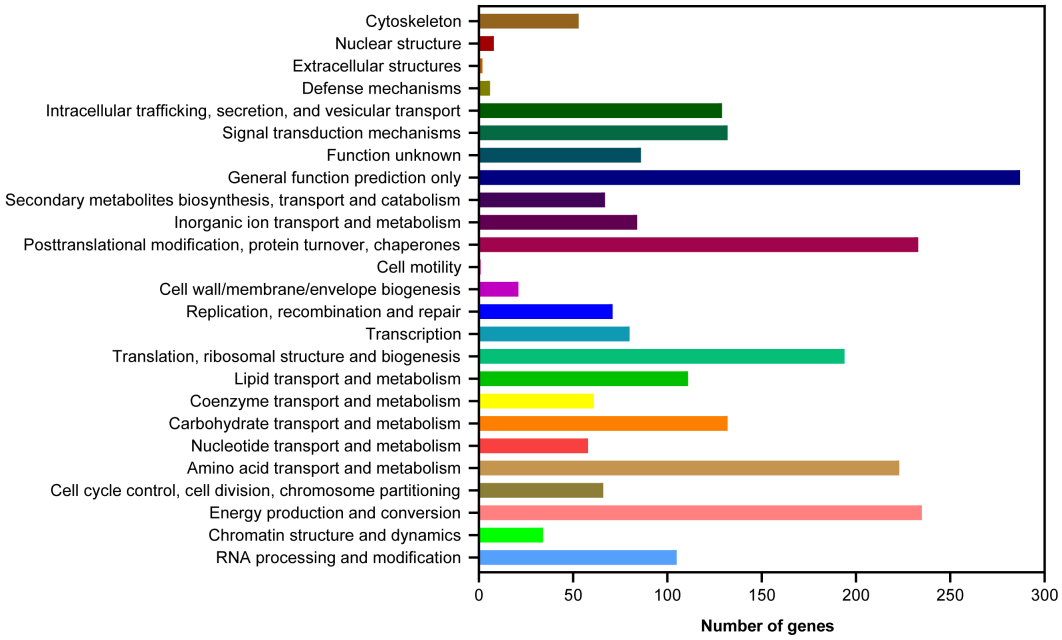


FIGURE 3
Clusters of Orthologous Groups of proteins (KOG) function classification of proteins in *A. sydowii* 29R-4-F02.

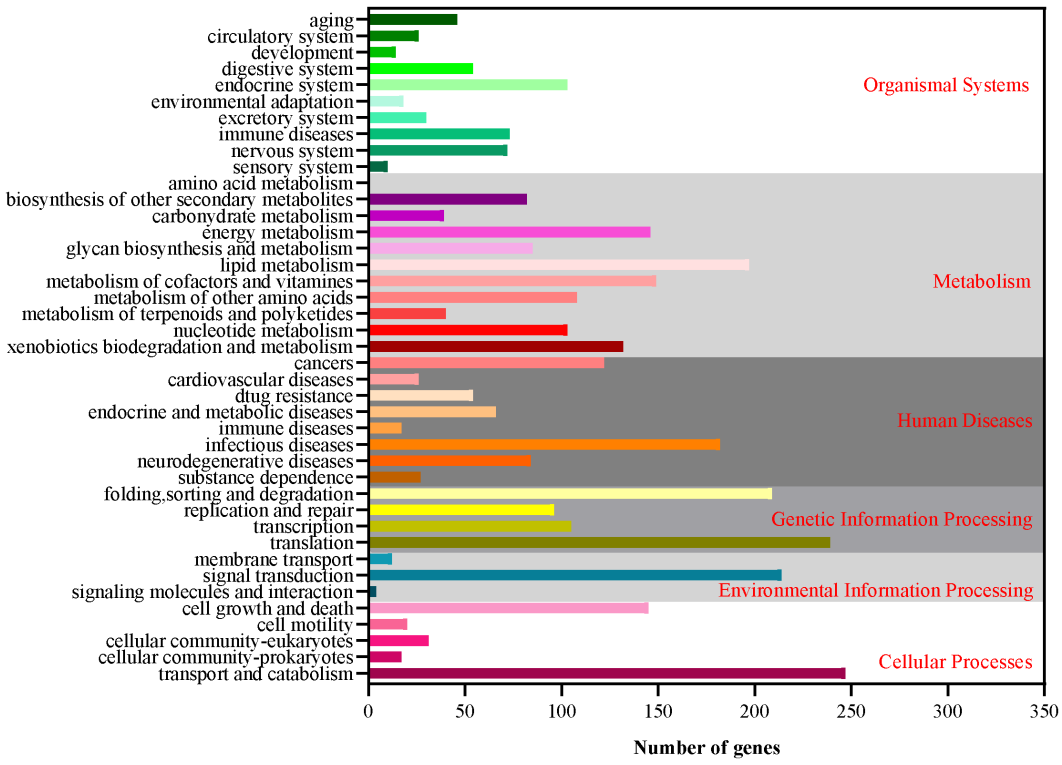


FIGURE 4
The Kyoto Encyclopedia of Genes and Genomes (KEGG) function annotation of *A. sydowii* 29R-4-F02.

for CBMs, and 23 encoding genes for PLs (Supplementary Table 4), which was much more than the genes in other species of *Aspergillus* (Table 2). Strain 29R-4-F02 also contains a large number of CAZyme coding genes related to the degradation of plant cell wall polysaccharides, including (hemi-) cellulases degrading enzymes (GH10, GH11, GH5, GH6, GH7, GH45, and GH115), pectin degrading enzymes (GH28,

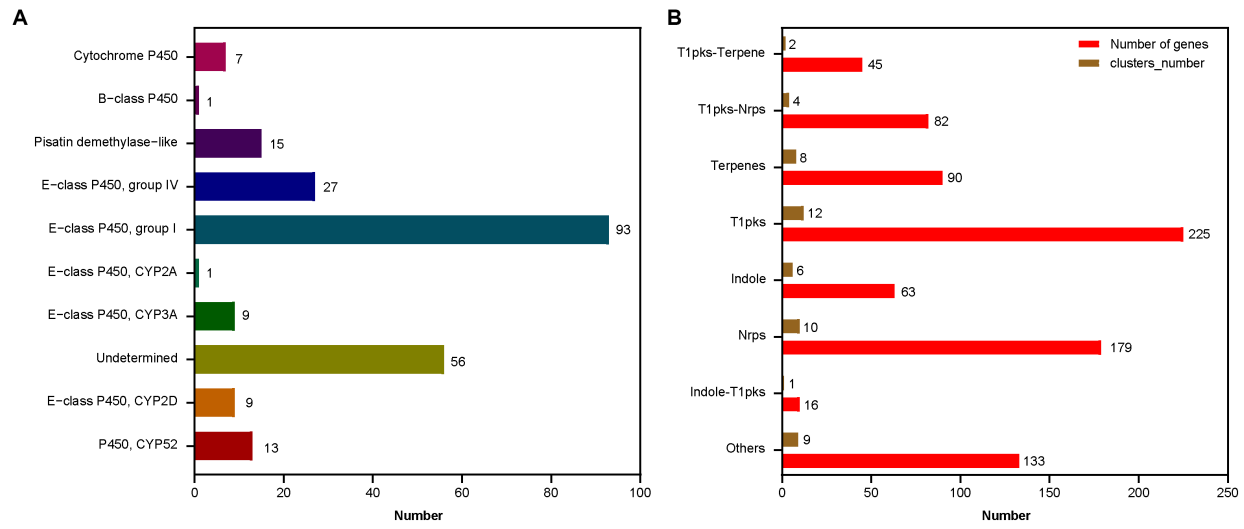


FIGURE 5
The function classification of the P450 genes **(A)** and the secondary metabolic gene clusters **(B)** in *A. sydowii* 29R-4-F02 genome. The horizontal axis represents the number of genes or gene clusters.

GH43, GH51, GH53, GH78, GH88, GH93, PL1, PL3, PL4, and PL9), xylan (GH10, GH11, GH62, GH67, CE1, and CE15), laccase (AA1), cellobiose dehydrogenase (AA3_1), vanillyl alcohol oxidase (AA7), lytic polysaccharide monooxygenases (AA9), and rhamnogalacturonan lyases (PL11) coding genes. Some CAZyme coding genes related to the degradation of non-plant polysaccharides (GH79 and GH88 families) were identified as well in the strain. These data indicate that *A. sydowii* 29R-4-F02 has a strong ability to degrade organic matter to facilitate its growth energy in lignite-containing subseafloor sediments.

Gene family evolution and phylogenetic analysis

OrthoFinder analysis showed that 9,678 gene families were identified from 22 sequenced fungi. Among them, 3,559 gene families were shared by 22 fungi and 364 gene families were unique to *A. sydowii* 29R-4-F02 (Figure 6). These 364 gene families are composed of species-specific genes (8) and uncluster genes (356). Of these, the species-specific genes (A00461, A03330, A06713, and A12243) with functional annotation were found to be primarily associated with vesicle fusion and autophagy. 185 uncluster genes (Supplementary Figure 3), as determined through gene functional annotation, were found to be mainly involved in biological processes, cell components and molecular functions according to the results of GO enrichment analysis (Supplementary Figure 4; Supplementary Table 5). These data reveal that *A. sydowii* 29R-4-F02 has the ability to adapt to complex subseafloor environments.

A phylogenetic tree was constructed based on 386 single copy orthologous genes identified from 22 fungi. The inference shows that *A. sydowii* 29R-4-F02 and *A. sydowii* CBS593.65 were clustered into the same clade (Figure 7). Compared with strain CBS593.65, strain 29R-4-F02 had 1,168 unique genes, including 12 genes belonging to 2 species-specific families (OG0000029 and OG0000322) and 1,156 uncluster genes (Supplementary Figures 5, 6; Supplementary Table 6). Among them, OG0000029 and OG0000322 were involved in ZEB2-regulated ABC transporter and glycopeptide α-N-acetylgalactosaminidase, respectively. ZEB2 regulated ABC transporters have been shown to play a crucial role in secreting virulence factors or compounds in pathogenic fungi (Lee et al., 2011; Zhang et al., 2023). Glycopeptide α-N-acetylgalactosaminidase belongs to glycoside hydrolase family (GH101) and has been primarily demonstrated to hydrolyze core 1-type O-glycans from glycoproteins (Fujita et al., 2005). Although this enzyme has been reported in bacteria, its presence in fungi has not been documented yet. In addition, 36 gene families involved in energy metabolism and environmental adaptation were significantly expanded in the genome of *A. sydowii* 29R-4-F02 (Supplementary Table 7). For example, OG0000024, OG0000315, OG0000220, and OG0000464 were associated with NADP-dependent alcohol dehydrogenase, short-chain dehydrogenase, heat shock protein, and nucleotide-excision repair, respectively. Given the nucleotide-excision repair mechanism has been proven to play an important role in genome conservation among subsurface microbes (Becraft et al., 2021; Liu et al., 2022), we believe that these extended genes may provide the genetic basis necessary for the long-term survival of *A. sydowii* 29R-4-F02 in extreme subseafloor sedimentary environments.

Positive selection improves the environmental adaptation of *A. sydowii* 29R-4-F02

Positive selection, acting on the evolution of functionally important gene families, is an important driving force for microorganisms in adaptation to the complex environment. To investigate those genes related to the adaptability of *Aspergillus* to the subseafloor environment, we conducted a positive selection analysis between *A. sydowii* 29R-4-F02 and *A. sydowii* CBS593.65 using KaKs_Calculator (Ka/Ks > 1 referred to positive selection; Wang et al., 2010). The results showed that 362 genes were under positive selection, including 257 annotated genes and 105 non-annotated genes. Among the annotation genes, 10 genes were

TABLE 2 Comparison of CAZyme encoding genes in different species of *Aspergillus*.

Species	GH	GT	PL	CE	CBM	AA
A. sydowii 29R-4-F02	337	126	23	39	67	78
A. parasiticus CBS 117618	315	127	25	33	51	81
A. transmontanensis CBS 130015	309	125	26	32	55	82
A. arachidicola CBS 117612	321	132	24	32	54	84
A. novoparasiticus CBS 126849	319	133	24	38	52	90
A. sergii CBS 130017	321	128	28	33	56	82
A. flavus CBS 128202	314	123	25	28	41	69
A. sojae CBS 100932	315	127	26	34	52	85
A. parvisclerotigenus CBS 121.62	310	128	25	29	49	68
A. oryzae RIB 40 FGS	302	119	26	27	41	65
A. minisclerotigenus CBS 117635	313	124	26	31	51	74
A. caelatus CBS 763.97	250	104	20	26	38	63
A. pseadocaelatus CBS 117616	323	132	26	31	49	80
A. pseudotamarii CBS 117625	321	125	27	28	51	76
A. tamarii CBS 117626	307	131	27	31	45	72
A. pseudonomius CBS 1193.88	289	127	24	27	59	75
A. nomius IBT 12657	286	126	23	27	50	74
A. bombycis CBS 117187	296	122	22	29	52	79
A. bertholletius IBT 29228	287	124	20	26	46	69
A. alliaceus CBS 536.65	283	131	21	28	41	66
A. albertensis IBT 14317	278	135	21	27	46	67
A. coremiiformis CBS 553.77	177	109	10	18	29	36
A. leporis CBS 151.66	291	127	22	27	49	68
A. avenaceus NRRL 4517	241	119	19	27	42	62
A. steynii CBS 112812	278	109	24	33	49	64
A. campestris CBS 538.81	176	94	10	15	38	36
A. terreus NIH2624	277	104	15	28	61	57
A. nidulans CBS 126972	272	97	25	28	57	54
A. fumigatus Af293CBS126847	264	106	14	27	55	35
A. niger ATCC1015	252	121	10	20	44	62

involved in CAZymes family, and 5 genes were associated with DNA repair. These genes may help *A. sydowii* 29R-4-F02 to acquire nutrients and energy in lignite containing subseafloor sediments.

Discussion

A. sydowii was a filamentous fungus with a worldwide distribution and attracted our attention due to its abundantly metabolic profile (de Vries et al., 2017; Brandt et al., 2020; Kumar et al., 2021; Jing et al., 2022). In this study, a strain of *A. sydowii* 29R-4-F02 isolated from seafloor sediments was sequenced *de novo* using Nanopore long read sequencing technique. Collinearity and BUSCO analysis showed that strain 29R-4-F02 had high genome quality and integrity. Gene annotation indicated that the strain has rich metabolic activities, especially protein metabolism and energy metabolism, which provided an opportunity for us to explore the lifestyle and ecological adaptation mechanism of fungi in the subseafloor environment.

CAZymes are thought to play an important role in the utilization of organic matter such as cellulose, hemicellulose, pectin and lignin polysaccharides by fungi to obtain nutrients and energy for growth, and are also closely related to host preference and lifestyle adaptation of fungi (Couturier et al., 2012; Ohm et al., 2012; Sista Kameshwar and Qin, 2018). Compared with other species of *Aspergillus*, *A. sydowii* 29R-4-F02 had extremely rich CAZymes encoding genes, especially GHs, CEs, and CBMs encoding genes, suggesting that *A. sydowii* 29R-4-F02 may have stronger environmental adaptability than other species of *Aspergillus*. In addition, the genome of *A. sydowii* 29R-4-F02 also contains GH79 and GH88 gene families, indicating the ability of this fungus to utilize non-plant polysaccharides, such as animal and bacterial polysaccharides (de Vries et al., 2017; Yu et al., 2020). As more and more sedimentary fungi are obtained (Nagahama et al., 2011; Bengtson et al., 2014; Xu et al., 2014; Burgaud et al., 2015; Drake et al., 2017; Hassett et al., 2019), the analysis of the diversity and abundance of CAZymes encoding genes in these fungi is

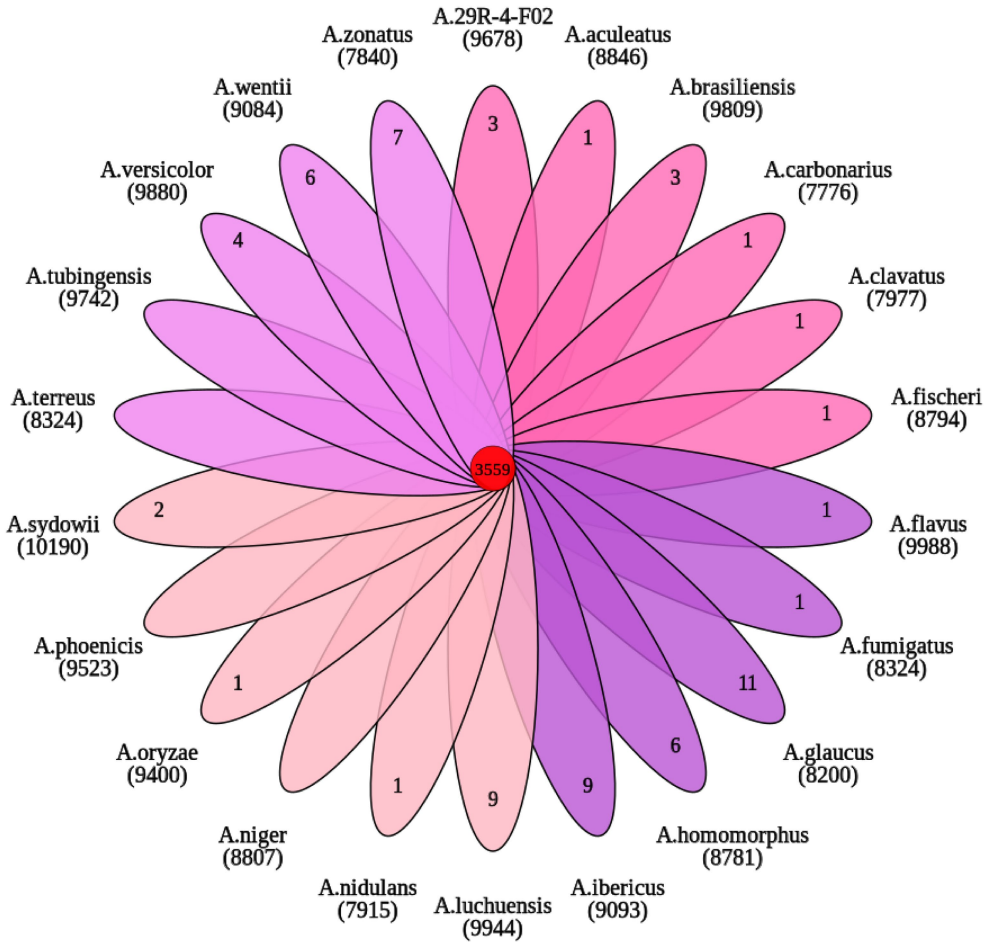


FIGURE 6
Venn diagram of orthologous gene families.

expected to provide a perspective on how fungi adapt to the extreme subseafloor environment where nutrients are scarce and difficult to utilize.

Based on orthology analysis, we found that *A. sydowii* 29R-4-F02 possesses 364 distinct gene families, among which, the OG0011277 gene family is annotated as a SNARE domain associated with vesicular fusion and autophagy in fungi. Vesicle fusion is the main form of intracellular transmembrane transport of macromolecules and particulate matter (Mohrmann et al., 2010), and autophagy is a strategy for cell renewal of cell components in response to stress or hunger (Glick et al., 2010). In addition, compared with the terrestrial strain *A. sydowii* CBS593.65, *A. sydowii* 29R-4-F02 had more unique genes, including ZEB2-regulated ABC transporters, glycopeptide α-N-acetylgalactoamidosidase, heat shock proteins and nucleotide excision repair genes. ZEB2-regulated ABC transporters are involved in the secretion of virulence factors or compounds of fungi (Coleman and Mylonakis, 2009; Lee et al., 2011), and glycopeptide α-N-acetylgalactosidase is involved in intercellular communication in higher eukaryotes (Fujita et al., 2005). Given that *A. sydowii* 29R-4-F02 was isolated from subseafloor sediment samples at *in situ* temperatures of 50–55°C, such high temperatures may cause damage to DNA, proteins and other biomolecules (Inagaki et al., 2015; Lever et al., 2015). Therefore, strain 29R-4-F02 possesses large number of nucleotides excision, DNA repair, and heat shock protein-related genes, which may be an evolutionary mechanism of heat protection in response to this environment.

Conclusion

The genome of *A. sydowii* 29R-4-F02 isolated from ~2.4 km of coal-bearing sediments below the seafloor was de vivo sequenced and assembled using a Nanopore long-read sequencing platform. Compared with other species or strains of Aspergillus, *A. sydowii* 29R-4-F02 has more abundant CAZymes encoding genes, which facilitate the fungus to obtain nutrients and energy in the subseafloor sediments. Gene expansion associated with vesicular fusion and autophagy, as well as positive selection of CAZymes and enzymes associated with DNA repair, may be one of the adaptive selection mechanisms that allowed fungi to survive 20 million years in extreme subseafloor environment. The results of this study will help us to understand the life style, evolution and survival mechanism of deep biosphere fungi.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession

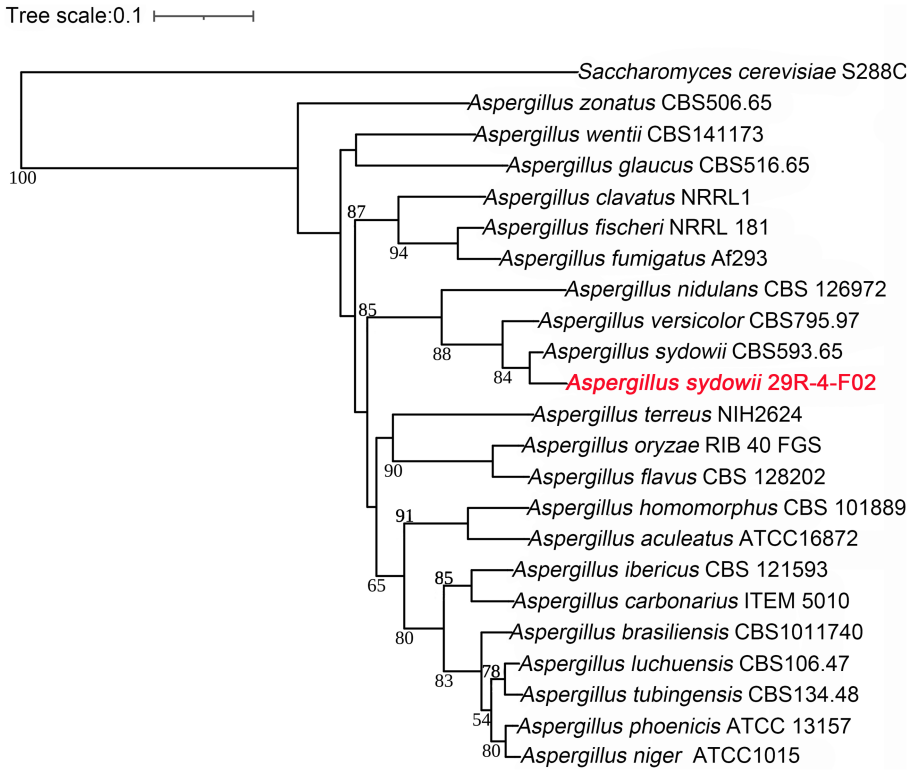


FIGURE 7
Phylogenetic tree of 21 *Aspergillus* genomes constructed based on single copy ortholog genes. The newly sequenced species in this study are represented in red font. *Saccharomyces cerevisiae* S288C as an outgroup species, the scale bar indicates evolutionary distance. Bootstrap values of <50 have been removed for clarity.

number(s) can be found below: https://www.biosino.org/elmsg/index, LMSG_G000000742.1; https://github.com/liuxuan-425lab/A.sydowii-29R-4-F02.

Author contributions

J-PJ, XL, and Y-FL performed data analysis. J-PJ wrote the first draft of this manuscript. XL and Y-PZ extracted DNA for genome sequencing. J-PJ and JS cultivated strain of *A. sydowii* 29R-4-F02. C-HL conceived the study and edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (nos. 41973073, 42273077, and 91951121) and the Science and Technology Innovation Program of Jiangsu Province (no. BK20220036).

Acknowledgments

The authors are grateful to all crews, drilling team members, lab technicians, and shipboard scientists on board the drilling vessel *Chikyu* for supporting core sampling and onboard measurements during IODP Expedition 337.

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.

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.2023.1216714/full#supplementary-material

References

Baeza, J. A., and García-De León, F. J. (2022). Are we there yet? Benchmarking low-coverage nanopore long-read sequencing for the assembling of mitochondrial genomes using the vulnerable silky shark *Carcharhinus falciformis*. *BMC Genomics* 23:320. doi: 10.1186/s12864-022-08482-z

Becraft, E. D., Vetter, M. C. Y. L., Bezuidt, O. K. I., Brown, J. M., Labont, J. M., Kauneckaite-Griguole, K., et al. (2021). Evolutionary stasis of a deep subsurface microbial lineage. *ISME J.* 15, 2830–2842. doi: 10.1038/s41396-021-00965-3

Bengtson, S., Ivarsson, M., Astolfo, A., Belivanova, V., Broman, C., Marone, F., et al. (2014). Deep-biosphere consortium of fungi and prokaryotes in Eocene subseafloor basalts. *Geobiology* 12, 489–496. doi: 10.1111/gbi.12100

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. doi: 10.1093/nar/27.2.573

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and genomewise. *Genome Res.* 14, 988–995. doi: 10.1101/gr.1865504

Borgonie, G., Linage-Alvarez, B., Ojo, A. O., Mundle, S. O., Freese, L. B., Van Rooyen, C., et al. (2015). Eukaryotic opportunists dominate the deep-subsurface biosphere in South Africa. *Nat. Commun.* 6:8952. doi: 10.1038/ncomms9952

Brandt, S. C., Brognaro, H., Ali, A., Ellinger, B., Maibach, K., Ruhl, M., et al. (2021). Insights into the genome and secretome of *Fusarium metavorans* DSM105788 by cultivation on agro-residual biomass and synthetic nutrient sources. *Biotechnol. Biofuels* 14:74. doi: 10.1186/s13068-021-01927-9

Brandt, S. C., Ellinger, B., van Nguyen, T., Harder, S., Schlüter, H., Hahnke, R. L., et al. (2020). Aspergillus sydowii: genome analysis and characterization of two heterologous expressed, non-redundant xylanases. *Front. Microbiol.* 11:2154. doi: 10.3389/fmicb.2020.573482

Burgaud, G., Hue, N. T. M., Arzur, D., Coton, M., Perrier-Cornet, J. M., Jebbar, M., et al. (2015). Effects of hydrostatic pressure on yeasts isolated from deep-sea hydrothermal vents. *Res. Microbiol.* 166, 700–709. doi: 10.1016/j.resmic.2015.07.005

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334

Chan, P. P., Lin, B. Y., Mak, A. J., and Lowe, T. M. (2021). tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* 49, 9077–9096. doi: 10.1093/nar/gkab688

Choi, J., Park, J., Kim, D., Jung, K., Kang, S., and Lee, Y. H. (2010). Fungal secretome database: integrated platform for annotation of fungal secretomes. *BMC Genomics* 11:105. doi: 10.1186/1471-2164-11-105

Coleman, J. J., and Mylonakis, E. (2009). Efflux in fungi: la piece de resistance. *PLoS Pathog.* 5:e1000486. doi: 10.1371/journal.ppat.1000486

Cong, B. L., Wang, N. F., Liu, S. H., Liu, F., Yin, X. F., and Shen, J. H. (2017). Isolation, characterization and transcriptome analysis of a novel Antarctic *Aspergillus sydowii* strain MS-19 as a potential lignocellulosic enzyme source. *BMC Microbiol.* 17:129. doi: 10.1186/s12866-017-1028-0

Couturier, M., Navarro, D., Olive, C., Chevret, D., Haon, M., Favel, A., et al. (2012). Post-genomic analyses of fungal lignocellulosic biomass degradation reveal the unexpected potential of the plant pathogen *Ustilago maydis*. *BMC Genomics* 13:57. doi: 10.1186/1471-2164-13-57

Cresnar, B., and Petric, S. (2011). Cytochrome P450 enzymes in the fungal kingdom. *Biochim. Biophys. Acta, Proteins Proteomics* 1814, 29–35. doi: 10.1016/j.bbapap.2010.06.020

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., and Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34, 2666–2669. doi: 10.1093/bioinformatics/bty149

de Vries, R. P., Riley, R., Wiebenga, A., Aguilar-Osorio, G., Amillis, S., Uchima, C. A., et al. (2017). Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus *Aspergillus*. *Genome Biol.* 18:28. doi: 10.1186/s13059-017-1151-0

Delcher, A. L., Phillippy, A., Carlton, J., and Salzberg, S. L. (2002). Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res.* 30, 2478–2483. doi: 10.1093/nar/30.11.2478

D'Hondt, S., Jorgensen, B. B., Miller, D. J., Batzke, A., Blake, R., Cragg, B. A., et al. (2004). Distributions of microbial activities in deep subseafloor sediments. *Science* 306, 2216–2221. doi: 10.1126/science.1101155

Drake, H., Ivarsson, M., Bengtson, S., Heim, C., Siljestrom, S., Whitehouse, M. J., et al. (2017). Anaerobic consortia of fungi and sulfate reducing bacteria in deep granite fractures. *Nat. Commun.* 8:55. doi: 10.1038/s41467-017-00094-6

Eddy, S. R. (1998). Profile hidden markov models. *Bioinformatics* 14, 755–763. doi: 10.1093/bioinformatics/14.9.755

Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:157. doi: 10.1186/s13059-015-0721-2

Fischer, M., Knoll, M., Sirim, D., Wagner, F., Funke, S., and Pleiss, J. (2007). The Cytochrome P450 engineering database: a navigation and prediction tool for the cytochrome P450 protein family. *Bioinformatics* 23, 2015–2017. doi: 10.1093/bioinformatics/btm268

Fujita, K., Oura, F., Nagamine, N., Katayama, T., Hiratake, J., Sakata, K., et al. (2005). Identification and molecular cloning of a novel glycoside hydrolase family of core 1 type O-glycan-specific endo-alpha-N-acetylgalactosaminidase from *Bifidobacterium longum*. *J. Biol. Chem.* 280, 37415–37422. doi: 10.1074/jbc.M506874200

Gardner, P. P., Daub, J., Tate, J. G., Nawrocki, E. P., Kolbe, D. L., Lindgreen, S., et al. (2009). Rfam: updates to the RNA families database. *Nucleic Acids Res.* 37, D136–D140. doi: 10.1093/nar/gkn766

Geiser, D. M., Taylor, J. W., Ritchie, K. B., and Smith, G. W. (1998). Cause of sea fan death in the West Indies. *Nature* 394, 137–138. doi: 10.1038/28079

Glick, D., Barth, S., and Macleod, K. F. (2010). Autophagy: cellular and molecular mechanisms. *J. Pathol.* 221, 3–12. doi: 10.1002/path.2697

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* 9:R7. doi: 10.1186/gb-2008-9-1-r7

Han, X. L., Chakrabortti, A., Zhu, J. D., Liang, Z. X., and Li, J. M. (2016). Sequencing and functional annotation of the whole genome of the filamentous fungus *Aspergillus westerdijkiae*. *BMC Genomics* 17:633. doi: 10.1186/s12864-016-2974-x

Hassett, B. T., Borrego, E. J., Vonnahme, T. R., Rämä, T., Kolomiets, M. V., and Gradinger, R. (2019). Arctic marine fungi: biomass, functional genes, and putative ecological roles. *ISME J.* 13, 1484–1496. doi: 10.1038/s41396-019-0368-1

Hayashi, A., Crombie, A., Lacey, E., Richardson, A. J., Vuong, D., Piggott, A. M., et al. (2016). *Aspergillus sydowii* marine fungal bloom in australian coastal waters, its metabolites and potential impact on symbiodinium dinoflagellates. *Mar. Drugs* 14:59. doi: 10.3390/md14030059

Inagaki, F., Hinrichs, K. U., Kubo, Y., Bowles, M. W., Heuer, V. B., Hong, W. L., et al. (2015). Exploring deep microbial life in coal-bearing sediment down to ~2.5 km below the ocean floor. *Science* 349, 420–424. doi: 10.1126/science.aaa6882

Jing, M. Y., Xu, X. H., Peng, J., Li, C., Zhang, H. C., Lian, C. L., et al. (2022). Comparative genomics of three *Aspergillus* strains reveals insights into endophytic lifestyle and endophyte-induced plant growth promotion. *J. Fungi* 8:690. doi: 10.3390/jof8070690

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Keller, N. P. (2019). Fungal secondary metabolism: regulation, function and drug discovery. *Nat. Rev. Microbiol.* 17, 167–180. doi: 10.1038/s41579-018-0121-1

Kim, K., and Harvell, C. D. (2004). The rise and fall of a six-year coral-fungal epizootic. *Am. Nat.* 164, S52–S63. doi: 10.1086/424609

Klich, M. A. (2002). *Identification of common Aspergillus species*. Centraalbureau voor Schimmelcultures, Utrecht.

Kumar, A. G., Manisha, D., Sujitha, K., Peter, D. M., Kirubagaran, R., and Dharani, G. (2021). Genome sequence analysis of deep sea *Aspergillus sydowii* BOBA1 and effect of high pressure on biodegradation of spent engine oil. *Sci. Rep.* 11:9347. doi: 10.1038/s41598-021-88525-9

Kumar, A., Sorensen, J. L., Hansen, F. T., Arvas, M., Syed, M. F., Hassan, L., et al. (2018). Genome sequencing and analyses of two marine fungi from the north sea unraveled a plethora of novel biosynthetic gene clusters. *Sci. Rep.* 8:10187. doi: 10.1038/s41598-018-28473-z

Lagesen, K., Hallin, P., Rodland, E. A., Staerfeldt, H. H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160

Lee, S., Son, H., Lee, J., Lee, Y. R., and Lee, Y. W. (2011). A putative ABC transporter gene, ZRA1, is required for zearalenone production in *Gibberella zeae*. *Curr. Genet.* 57, 343–351. doi: 10.1007/s00294-011-0352-4

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245. doi: 10.1093/nar/gkw290

Lever, M. A., Rogers, K. L., Lloyd, K. G., Overmann, J., Schink, B., Thauer, R. K., et al. (2015). Life under extreme energy limitation: a synthesis of laboratory- and field-based investigations. *FEMS Microbiol. Rev.* 39, 688–728. doi: 10.1093/femsre/fuv020

Li, H. (2016). Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* 32, 2103–2110. doi: 10.1093/bioinformatics/btw152

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191

Liu, X., Huang, X., Chu, C., Xu, H., Wang, L., Xue, Y. R., et al. (2022). Genome, genetic evolution, and environmental adaptation mechanisms of *Schizophyllum commune* in

deep subseafloor coal-bearing sediments. *iScience* 25:104417. doi: 10.1016/j.isci.2022.104417

Liu, C. H., Huang, X., Xie, T. N., Duan, N., Xue, Y. R., Zhao, T. X., et al. (2017). Exploration of cultivable fungal communities in deep coal-bearing sediments from similar to 1.3 to 2.5 km below the ocean floor. *Environ. Microbiol.* 19, 803–818. doi: 10.1111/1462-2920.13653

Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495. doi: 10.1093/nar/gkt1178

Lomstein, B. A., Langerhuus, A. T., D'Hondt, S., Jorgensen, B. B., and Spivack, A. J. (2012). Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment. *Nature* 484, 101–104. doi: 10.1038/nature10905

Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., et al. (2011). AntiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* 39, W339–W346. doi: 10.1093/nar/gkr466

Mohrmann, R., de Wit, H., Verhage, M., Neher, E., and Sorensen, J. B. (2010). Fast vesicle fusion in living cells requires at least three snare complexes. *Science* 330, 502–505. doi: 10.1126/science.1193134

Nagahama, T., Takahashi, E., Nagano, Y., Abdel-Wahab, M. A., and Miyazaki, M. (2011). Molecular evidence that deep-branching fungi are major fungal components in deep-sea methane cold-seep sediments. *Environ. Microbiol.* 13, 2359–2370. doi: 10.1111/j.1462-2920.2011.02507.x

Nagano, Y., Konishi, M., Nagahama, T., Kubota, T., Abe, F., and Hatada, Y. (2016). Retrieval of deeply buried culturable fungi in marine subsurface sediments, Suruga-Bay, Japan. *Fungal Ecol.* 20, 256–259. doi: 10.1016/j.funeco.2015.12.012

Nagano, Y., and Nagahama, T. (2012). Fungal diversity in deep-sea extreme environments. *Fungal Ecol.* 5, 463–471. doi: 10.1016/j.funeco.2012.01.004

Niu, S. W., Yang, L. H., Chen, T. T., Hong, B. H., Pei, S. X., Shao, Z. Z., et al. (2020). New monoterpenoids and polyketides from the deep-sea sediment-derived fungus *Aspergillus sydowii* MCCC 3A00324. *Mar. Drugs* 18:561. doi: 10.3390/md18110561

Ohm, R. A., Feau, N., Henrissat, B., Schoch, C. L., Horwitz, B. A., Barry, K. W., et al. (2012). Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. *PLoS Pathog.* 8:e1003037. doi: 10.1371/journal.ppat.1003037

Orsi, W. D., Edgcomb, V. P., Christman, G. D., and Biddle, J. F. (2013). Gene expression in the deep biosphere. *Nature* 499, 205–208. doi: 10.1038/nature12230

Ortega-Arbulu, A. S., Pichler, M., Vuillemin, A., and Orsi, W. D. (2019). Effects of organic matter and low oxygen on the mycobenthos in a coastal lagoon. *Environ. Microbiol.* 21, 374–388. doi: 10.1111/1462-2920.14469

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701

Pitt, J. I., Lange, L., Lacey, A. E., Vuong, D., Midgley, D. J., Greenfield, P., et al. (2017). *Aspergillus hancockii* sp. nov., a biosynthetically talented fungus endemic to southeastern Australian soils. *PLoS One* 12:e0170254. doi: 10.1371/journal.pone.0170254

Quemener, M., Mara, P., Schubotz, F., Beaudoin, D., Li, W., Pachiadaki, M., et al. (2020). Meta-omics highlights the diversity, activity and adaptations of fungi in deep oceanic crust. *Environ. Microbiol.* 22, 3950–3967. doi: 10.1111/1462-2920.15181

Raper, K. B., and Fennell, D. I. (1965). *The genus Aspergillus*. Baltimore,: Williams & Wilkins.

Rédou, V., Navarri, M., Meslet-Cladière, L., Barbier, G., and Burgaud, G. (2015). Species richness and adaptation of marine fungi from deep-subseafloor sediments. *Appl. Environ. Microbiol.* 81, 3571–3583. doi: 10.1128/aem.04064-14

Saha, S., Bridges, S., Magbanua, Z. V., and Peterson, D. G. (2008). Empirical comparison of ab initio repeat finding programs. *Nucleic Acids Res.* 36, 2284–2294. doi: 10.1093/nar/gkn064

Schippers, A., Neretin, L. N., Kallmeyer, J., Ferdelman, T. G., Cragg, B. A., Parkes, R. J., et al. (2005). Prokaryotic cells of the deep sub-seafloor biosphere identified as living bacteria. *Nature* 433, 861–864. doi: 10.1038/nature03302

Schrenk, M. O., Huber, J. A., and Edwards, K. J. (2010). Microbial provinces in the subseafloor. *Annu. Rev. Mar. Sci.* 2, 279–304. doi: 10.1146/annurev-marine-120308-081000

Selbmann, L., Egidi, E., Isola, D., Onofri, S., Zucconi, L., de Hoog, G. S., et al. (2013). Biodiversity, evolution and adaptation of fungi in extreme environments. *Plant Biosyst.* 147, 237–246. doi: 10.1080/11263504.2012.753134

Sista Kameshwar, A. K., and Qin, W. (2018). Comparative study of genome-wide plant biomass-degrading CAZymes in white rot, brown rot and soft rot fungi. *Mycology* 9, 93–105. doi: 10.1080/21501203.2017.1419296

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24, 637–644. doi: 10.1093/bioinformatics/btn013

Takahashi, H., Umemura, M., Ninomiya, A., Kusuya, Y., Shimizu, M., Urayama, S., et al. (2021). Interspecies genomic variation and transcriptional activeness of secondary metabolism-related genes in *Aspergillus* section *Fumigati*. *Front. Fungal Biol.* 2:656751. doi: 10.3389/ffunb.2021.656751

Vaser, R., Sovic, I., Nagarajan, N., and Sikic, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27, 737–746. doi: 10.1101/gr.214270.116

Wang, W. Y., Gao, M. L., Luo, Z. H., Liao, Y. Y., Zhang, B. B., Ke, W. Q., et al. (2019). Secondary metabolites isolated from the deep sea-derived fungus *Aspergillus sydowii* C1-S01-A7. *Nat. Prod. Res.* 33, 3077–3082. doi: 10.1080/14786419.2018.1519561

Wang, M. S., Ruan, R. X., and Li, H. Y. (2021). The completed genome sequence of the pathogenic ascomycete fungus *Penicillium digitatum*. *Genomics* 113, 439–446. doi: 10.1016/j.ygeno.2021.01.001

Wang, Y. P., Tang, H. B., DeBarry, J. D., Tan, X., Li, J. P., Wang, X. Y., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40:e49. doi: 10.1093/nar/gkr1293

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinform.* 8, 77–80. doi: 10.1016/S1672-0229(10)60008-3

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., et al. (2017). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35, 543–548. doi: 10.1093/molbev/msx319

Xu, W., Pang, K. L., and Luo, Z. H. (2014). High fungal diversity and abundance recovered in the deep-sea sediments of the pacific ocean. *Microb. Ecol.* 68, 688–698. doi: 10.1007/s00248-014-0448-8

Yin, Y. B., Mao, X. Z., Yang, J. C., Chen, X., Mao, F. L., and Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 40, W445–W451. doi: 10.1093/nar/gks479

Yu, F., Song, J., Liang, J. F., Wang, S. K., and Lu, J. K. (2020). Whole genome sequencing and genome annotation of the wild edible mushroom, *Russula griseocarnosa*. *Genomics* 112, 603–614. doi: 10.1016/j.ygeno.2019.04.012

Yurchenko, A. N., Girich, E. V., and Yurchenko, E. A. (2021). Metabolites of marine sediment-derived fungi: actual trends of biological activity studies. *Mar. Drugs* 19:88. doi: 10.3390/md19020088

Zain Ul Arifeen, M., Chu, C., Yang, X., Liu, J., Huang, X., Ma, Y., et al. (2021a). The anaerobic survival mechanism of *Schizophyllum commune* 20R-7-F01, isolated from deep sediment 2 km below the seafloor. *Environ. Microbiol.* 23, 1174–1185. doi: 10.1111/1462-2920.15332

Zain Ul Arifeen, M., Ma, Z. J., Wu, S., Liu, J. Z., Xue, Y. R., and Liu, C. H. (2021b). Effect of oxygen concentrations and branched-chain amino acids on the growth and development of sub-seafloor fungus, *Schizophyllum commune* 20R-7-F01. *Environ. Microbiol.* 23, 6940–6952. doi: 10.1111/1462-2920.15738

Zhang, Y., He, K., Guo, X., Jiang, J., Qian, L., Xu, J., et al. (2023). Transcriptomic profiling of *Fusarium pseudograminearum* in response to Carbendazim, Pyraclostrobin, Tebuconazole, and Phenamacril. *J. Fungi* 9:334. doi: 10.3390/jof9030334

Zhang, Z., Li, J., Zhao, X. Q., Wang, J., Wong, G. K., and Yu, J. (2006). KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinform.* 4, 259–263. doi: 10.1016/S1672-0229(07)60007-2

Zhao, G., Yao, Y., Qi, W., Wang, C., Hou, L., Zeng, B., et al. (2012). Draft genome sequence of *Aspergillus oryzae* strain 3.042. *Eukaryot. Cell* 11:1178. doi: 10.1128/ec.00160-12

frontiers | Frontiers in Microbiology

TYPE Review
PUBLISHED 30 June 2023
DOI 10.3389/fmicb.2023.1186920

OPEN ACCESS

EDITED BY
Ernesto Perez-Rueda,
Universidad Nacional Autónoma de México, Mexico

REVIEWED BY
Alexandro Rodríguez-Rojas,
University of Veterinary Medicine Vienna, Austria
Sheryl S. Justice,
The Ohio State University, United States

*CORRESPONDENCE
Jean-Philippe Rasigade
✉ jean-philippe.rasigade@univ-lyon1.fr

RECEIVED 15 March 2023
ACCEPTED 09 June 2023
PUBLISHED 30 June 2023

CITATION
Vanacker M, Lenuzza N and Rasigade J-P (2023) The fitness cost of horizontally transferred and mutational antimicrobial resistance in *Escherichia coli*. *Front. Microbiol.* 14:1186920. doi: 10.3389/fmicb.2023.1186920

The fitness cost of horizontally transferred and mutational antimicrobial resistance in *Escherichia coli*

Marie Vanacker[1], Natacha Lenuzza[1] and Jean-Philippe Rasigade[1,2]*

[1]CIRI, Centre International de Recherche en Infectiologie, Université de Lyon, Inserm U1111, Université Claude Bernard Lyon 1, CNRS, UMR5308, ENS de Lyon, Lyon, France, [2]Institut des Agents Infectieux, Hospices Civils de Lyon, Lyon, France

Antimicrobial resistance (AMR) in bacteria implies a tradeoff between the benefit of resistance under antimicrobial selection pressure and the incurred fitness cost in the absence of antimicrobials. The fitness cost of a resistance determinant is expected to depend on its genetic support, such as a chromosomal mutation or a plasmid acquisition, and on its impact on cell metabolism, such as an alteration in an essential metabolic pathway or the production of a new enzyme. To provide a global picture of the factors that influence AMR fitness cost, we conducted a systematic review and meta-analysis focused on a single species, *Escherichia coli*. By combining results from 46 high-quality studies in a multilevel meta-analysis framework, we find that the fitness cost of AMR is smaller when provided by horizontally transferable genes such as those encoding beta-lactamases, compared to mutations in core genes such as those involved in fluoroquinolone and rifampicin resistance. We observe that the accumulation of acquired AMR genes imposes a much smaller burden on the host cell than the accumulation of AMR mutations, and we provide quantitative estimates of the additional cost of a new gene or mutation. These findings highlight that gene acquisition is more efficient than the accumulation of mutations to evolve multidrug resistance, which can contribute to the observed dominance of horizontally transferred genes in the current AMR epidemic.

KEYWORDS

conjugation, plasmid, mutation, relative fitness, compensatory evolution

1. Introduction

Antimicrobial resistance (AMR) increases at an alarming rate worldwide, imposing a considerable burden to health systems and menacing the safety of modern medical procedures (Nadeem et al., 2020). AMR has developed against virtually all antibiotics in clinical use (Andersson and Levin, 1999; Levin et al., 2000; Acar and Röstel, 2001; Pope et al., 2010; Nadeem et al., 2020). AMR genes are present in most environments including livestock, sewage, or rivers (Jian et al., 2021), and many genes can access new bacterial species through horizontal transfer, enhancing their dissemination potential across ecological niches. For instance, Pu et al. (2019) have highlighted how carbapenem resistance in *Escherichia coli* provided by the bla_{NDM} AMR gene can be shared across humans, dogs, flies, and wild birds in farms. In antibiotic-free environments, however, resistant bacteria are expected to incur a

fitness cost (practically, a reduced growth rate) and to be eventually outcompeted by their susceptible, more fit counterparts. This fitness cost may result from the alteration of an enzyme by mutation, the disruption of a metabolic pathway following gene loss or inactivation, or the additional energy required for the overexpression of a gene or the expression and replication of a new gene acquired through horizontal transfer.

The biology and ecology of AMR emergence differ fundamentally depending on whether AMR results from the mutation of a chromosomal gene (hereafter, AMR mutation) or the acquisition of a transferable AMR gene (Acar and Röstel, 2001). AMR mutations typically alter essential enzymes while transferable AMR genes typically provide new enzymes that may or may not interfere with cell metabolism. From an ecological standpoint, AMR mutations and transferable AMR genes also differ because the unit of selection of the vertically inherited mutations is the bacterial cell and its offspring, while transferable genes can emancipate from their host and become themselves the unit of selection. Therefore, the impact of fitness cost on the persistence of transferable AMR genes is expected to differ from AMR mutations (Vogwill and MacLean, 2015). This point has practical importance because transferable AMR genes, including those encoding extended spectrum beta-lactamases, carbapenemases, or aminoglycoside modifying enzymes are the main drivers of the current AMR epidemic in enterobacteria, while AMR mutations are less problematic. It is still unclear, however, whether the fitness cost of AMR depends on the transfer potential of the AMR determinant. Most comparative studies of AMR fitness cost have either focused on fitness variations across AMR mechanisms, mutations, drug families or bacterial species, rather than on the difference between transferable and non-transferable resistance.

To fill this knowledge gap, we conducted a comparative meta-analysis of the fitness cost of resistance, with a focus on the transferable nature of AMR determinants. We focused on a single, well-studied species, *E. coli*, to ease interpretation of the results by avoiding biasing fitness evaluation across multiple host species. Using a multilevel meta-analysis framework, we examined whether the fitness cost of AMR, from a single determinant to an accumulation of many genes or mutations, differs when provided by horizontally transferable genes such as those encoding beta-lactamases, comparable to mutations in core genes such as those involved in fluoroquinolone and rifampicin resistance.

2. Materials and methods

2.1. Estimation of the relative fitness associated with AMR determinants

This is a systematic review and meta-analysis of the relative fitness of AMR determinants in *E. coli*. Several procedures exist to estimate relative fitness, mostly based on competition assays between a resistant strain and its susceptible, isogenic counterpart. We retained three different estimations of relative fitness, briefly described below.

In the first estimation method described by Lenski et al. (1991), the relative fitness W_r (Equation 1) is the ratio of the Malthusian parameters m (or exponential growth rate) of a resistant mutant (m_R) and a susceptible strain (m_S). The parameter m is usually estimated experimentally as the logarithm of the ratio of the final population size N^t (after t epochs) on the initial population size N^0, by solving the exponential growth equation $N^t = N^0 e^{mt}$. The ratio of Malthusian parameters can be written:

$$W_r = \frac{m_R}{m_S} \approx \frac{\ln N_R^t / N_R^0}{\ln N_S^t / N_S^0} \quad (1)$$

Remark that this equation does not explicitly take t into account, and the estimates may vary with the duration of the growth assay.

The second estimation method (Lenski, 1991) explicitly takes the duration of the competition assay into account by measuring population sizes N_R and N_S of the resistant and susceptible strains, respectively, at different time points and by regressing the logarithm of their ratio, $\ln\left(N_R / N_S\right)$, over time. The relative fitness W_s is defined as the complement of the regression slope s, $W_s = 1 - s$.

The third estimation method (Dykhuizen and Hartl, 1983; Dean and Dykhuizen, 1988; Dykhuizen, 1990) is based on the increase per time unit of the difference between the Malthusian parameters of the resistant and the susceptible strain:

$$W_t = 1 + \frac{m_R - m_S}{t} \approx 1 + \frac{1}{t}\left(\ln\left(N_R^t / N_R^0\right) - \ln\left(N_S^t / N_S^0\right)\right) \quad (2)$$

The three estimators W_r (Malthusian ratio), W_s (regression slope), and W_t (Malthusian difference per time unit) are collectively referred to as the relative fitness W. A relative fitness $W = 1$ indicates an absence of effect of resistance, $W < 1$ indicates a fitness cost, and $W > 1$ indicates a fitness advantage of the resistant strain. Although the qualitative interpretation of neutral, reduced, or increased fitness is common to all three methods of estimations, these methods differ quantitatively. W_r is a dimensionless ratio that lacks any direct interpretation and that can only be used to compare experiments of the same duration t, typically 24 h, as this duration is implicit in Eq. 1. W_s has a more direct interpretation because it reflects the relative increase of population size through time, however, it is expressed in logarithmic units and involves a complement that obfuscates its biological meaning. Finally, W_t has a similar interpretation as W_s because it represents a relative increase per unit of time, using the Malthusian parameter in place of the logarithm of population size used in W_s. It should be noted that neither W_r, W_s, or W_t are meaningful quantitative representations of the fitness because they involve either ratios of logarithms (Eq. 1), which are dimensionless, or slopes of logarithms. More interpretable quantities may be derived from, for instance, the ratio of the doubling time of the resistant and competitor strains. However, such quantities are not in common use and could not be recovered from the relative fitness data W_r, W_s, or W_t, reported in the included studies.

The biological impact of the relative fitness can be illustrated more intuitively by computing the number of generations $t = \log_2 1000 / \left(1 - W_R\right)$ after which the susceptible variant becomes 1,000x more prevalent than the resistant variant (the relation for t is easily derived from Eq. 1). A relative fitness $W_R = 0.9$ yields $t = 99.7$, which means that ~100 generations (33 h for *E. coli*, assuming a 20 min doubling time) are sufficient to virtually eliminate the resistant variant. Hence, a relative fitness of 0.9 may be considered a very strong fitness

cost. For $W_R = 0.99$, the susceptible variant becomes 1,000x more prevalent than the resistance variant after $t \approx 1{,}000$ generations, or 2 weeks for *E. coli*. Hence, a relative fitness of 0.99 may be considered a moderate fitness cost. The relation for t also illustrates that a relative fitness very close to 1, which may be difficult to estimate experimentally, can still have a substantial impact on the bacterial population over months or years of competition.

2.2. Literature search and inclusion criteria

We searched the PubMed database using terms 'fitness' and '*Escherichia coli*' or '*E. coli*' and 'antibiotic resistance' or 'antimicrobial resistance' or 'drug resistance'. Search results were limited to peer-reviewed studies in English available online by the 29th September 2022. No start date was specified.

To be included, studies had to report relative fitness findings numerically in the text, in a table or a figure. Relative fitness measurements should meet the following criteria. The relative fitness had to be measured in competition assays in an approximate proportion of 1:1 between a resistant strain (mutant) and a wild-type (susceptible or ancestral) strain only differing by the absence of resistance. Studies in which the data were reported in an unstandardized protocol without competition with a control strain were excluded. Studies comparing two resistant strains were also excluded. The competition assays had to be conducted at 37°C in a stable antibiotic free environment. Experiments involving a modification of environmental conditions were excluded. Finally, relative fitness had to be estimated using one of the three methods W_r, W_s or W_t described above. Estimations based on relative growth rates, competitive indices or uncommon methods were excluded.

Of 335 studies matching the search criteria, 127 reported relative fitness values and 46 were included in the final dataset (Figure 1), representing a total of 783 resistant strains. Details of all eligible studies and reasons for exclusion where applicable are given in the Supplementary Data S1. For each resistant strain and competition assay included in the final analysis, we collected the duration of the assay, the culture medium, the relative fitness estimation method, the mean and standard error of the relative fitness, the number of experiment replicates, the nature of the susceptible strain (either isogenic, or ancestral), the nature of its differences relative to the resistant strain (such as plasmid loss, mutation) and the number of AMR genes or mutations in the resistant strain. Detailed datasets can be found in the Supplementary Data S2, S3. Out of the 46 studies, 22 used the W_r estimator of the relative fitness, 11 used the W_s estimator, and 13 used the W_t estimator.

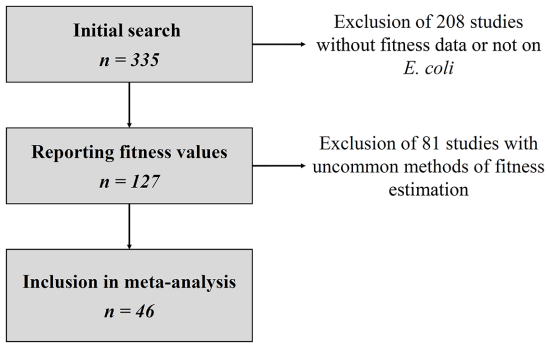


FIGURE 1
Flowchart of study selection. Details of eligible studies and reasons for exclusion can be found in the Supplementary Data S1.

2.3. Characteristics of AMR determinants

Genetic determinants of AMR in each resistant strain were classified as either acquired AMR genes, such as *bla*Oxa48, when harbored on mobile genetic elements, or AMR mutations and indels in non-transferable genes such as *rpo*B, referred to as mutated AMR genes. For each acquired AMR gene or mutation, the class, family, and mechanism of action of the targeted antibiotic was identified from the literature (McArthur et al., 2013; Poirel et al., 2018). Genes whose acquisition or mutation confer resistance to several drug families, such as *acr*B, were also identified as such (McArthur et al., 2013; Poirel et al., 2018; Wang et al., 2019; Brandis et al., 2021; Praski Alzrigat et al., 2021; Rajer and Sandegren, 2022). As acquired AMR genes may be harbored by plasmids, transposons, integrated plasmids or transposons integrated in plasmids, we did not distinguish between plasmid- and transposon-borne genes in our analyses to avoid ambiguity. When acquired AMR genes were explicitly reported as plasmid-borne, we collected the size of plasmid, the number of harbored AMR genes and the incompatibility group.

2.4. Statistical analysis

A multilevel meta-analysis of relative fitness, taking into account three levels of analysis, was conducted as described by Harrer et al. (2022):

- The first level is the level of interest for the analysis, comparing groups of resistant strains based on the relative fitness average and standard error. Several strains were thus recorded for each study.
- The second level captures intra-study variation.
- The third level captures inter-study variation and the pooling of the aggregated cluster effects leading to the overall effect.

The overall effect size $\hat{\theta}_{ij}$ is the effect of the strain i nested in the study j as described in Equation 3, where μ is the overall mean population effect, $\delta_{(2)ij}$ the intra-study heterogeneity at level 2, $\delta_{(3)j}$ the inter-study heterogeneity at level 3, and $\varepsilon_{(1)ij}$ the sampling error of strains estimated as the standard deviation of relative fitness.

$$\hat{\theta}_{ij} = \mu + \delta_{(2)ij} + \delta_{(3)j} + \varepsilon_{(1)ij} \quad (3)$$

The share of variance not attributable to sampling error in intra and inter-study heterogeneity was calculated using the I^2 statistic (Higgins and Thompson, 2002). The significance threshold was set at 5% without correction for multiple testing, in line with the exploratory nature of the analysis. To estimate the additional fitness cost associated with the accumulation of AMR mutations or acquired AMR genes, meta-regression models were constructed using the

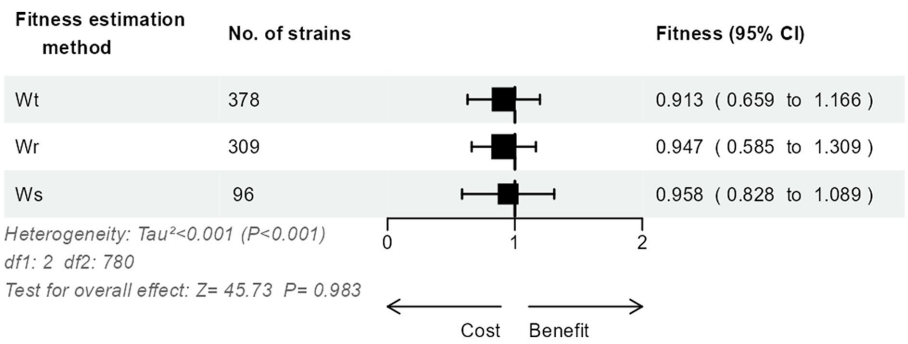


FIGURE 2
Three-level meta-analysis of the impact of the estimation method on the results of relative fitness in competition experiments of 783 *E. coli* strains from 46 studies. Shown are forest plots of the relative fitness estimated using either the Malthusian difference per time unit W_t, the Malthusian ratio W_r, or the regression slope method W_s. Square markers denote effect size, marker size is proportional to the group weight in the overall effect, and error bars denote the 95% confidence interval. No significant overall effect of the estimation method was observed.

TABLE 1 Model comparison between a three-level model (inter-study variance) and a two-level model (intra-study variance) for the meta-analysis of relative fitness in 783 *E. coli* strains from 46 studies.

	Degree of freedom	AIC	Log-likehood	Likehood ratio test (LRT)	LRT value of *p*
Three-level model	3	−971	489		
Two-level model	2	−736	370	238	<0.0001

1. The best model is the one with the lowest value of AIC. 2. The best model is the one with the highest value of log-likehood.

relative fitness as the response variable and the number of mutated or acquired AMR genes per strain, or the number of drug resistance families (e.g., beta-lactams or fluoroquinolones) as a covariate. Meta-regression analyses were conducted on strains containing either only AMR mutations or only acquired AMR genes to avoid mixing the effect of gene mutation and acquisition in the models. Where applicable, meta-regression models used a fixed intercept of 1 (neutral fitness) to account for the fact that the absence of AMR mutation, gene or resistance should yield the same fitness as the comparator strain. All analyses were performed using R version 4.1 (R Core Team, 2022) with additional packages 'metafor' (Viechtbauer, 2010), 'meta' (Balduzzi et al., 2019), 'forestploter' (Dayimu, 2022), and 'dmetar' (Harrer et al., 2019).

3. Results

3.1. Influence of experimental conditions on the estimation of relative fitness

We searched for sources of heterogeneity in the estimation of relative fitness across the 46 studies included in the analysis, totaling 783 resistant *E. coli* strains. A significant inter-study heterogeneity was found, based on I^2 test (Supplementary Figure S1). A funnel plot analysis showed a high asymmetry between the mean standardized difference and the between-study standard errors of relative fitness, suggesting the presence of publication bias. This publication bias was also confirmed by Egger's test with a value of $p < 0.001$(Supplementary Figure S2). The decomposition of variance in multi-level analysis (Supplementary Figure S3) attributed 32% of the overall variance to between-study variation (analysis level 3), 68% to within-study variation (level 2) and 0% to strain-level variation (level 1). Hence, heterogeneity was most concentrated at the within-study level. Both the culture medium and the duration of competition assays had a moderate influence on relative fitness (Supplementary Figure S4A). Most assays used Luria-Bertani medium ($n = 501$ strains out of 783, 63.4%) over 24 h ($n = 517$ strains, 66.0%). Relative fitness estimates decreased significantly with the duration of competition assays ($p = 0.035$; Supplementary Figure S4B), suggesting that longer experiments potentialize or better reveal the fitness cost.

The relative fitness of the 783 strains had been calculated using either the W_t method (Malthusian difference per time unit; $n = 378$, 48.3%), the W_r method (Malthusian ratio; $n = 309$, 39.5%), or the W_s method (regression slope; $n = 96$, 12.3%). Interestingly, we could not detect a significant impact of the estimation method on the relative fitness ($p = 0.98$; Figure 2).

Using goodness-of-fit analyses, we observed that a three-level meta-analysis model, taking inter-study variation into account, outperformed a simpler, two-level model in terms of Akaike information criterion and likelihood ratio test (Table 1). In line with this observation, we used three-level models in the subsequent analyses.

3.2. Relative fitness comparison of AMR mutations and gene acquisitions

A total of 146 unique AMR genes were studied in relative fitness experiments, of which 78 (53.4%) were acquired and 68 (46.6%) were mutated. One hundred and sixty-three unique mutations were identified among the mutated AMR genes, corresponding to an average of 2.7 mutations per gene (range, 1 to 31). The AMR determinants most commonly studied were *rpo*B and *gyr*A mutations (26.7 and 24.4% of strains, respectively; Table 2) and the most common acquired AMR gene was bla_{TEM} (12.3%). One hundred and thirty strains (16.6%) included both mutated and acquired AMR genes. Strains with only mutated AMR genes were resistant to an

TABLE 2 The mutated or acquired AMR genes present in >5% of 783 resistant *E. coli* strains from 46 studies of relative fitness.

Gene	No. of strains (%), *n*=783[1]	Genetic support	AMR family	References
*aad*A5	49 (6.3%)	Acquisition	Aminoglycosides	McArthur et al. (2013), Poirel et al. (2018), and Rajer and Sandegren (2022)
*acr*R	58 (7.4%)	Mutation	Multidrug	Praski Alzrigat et al. (2021)
aph(3″)-Ib	50 (6.4%)	Acquisition	Aminoglycosides	McArthur et al. (2013), Poirel et al. (2018), and Tang et al. (2022)
aph(6)-Id	58 (7.4%)	Acquisition	Aminoglycosides	McArthur et al. (2013), Poirel et al. (2018), and Rajer and Sandegren (2022)
*bla*CTXM-15	52 (6.6%)	Acquisition	Beta-lactams	Poirel et al. (2018) and Alonso-del Valle et al. (2021)
*bla*OXA	92 (11.7%)	Acquisition	Beta-lactams	Silva et al. (2011) and Poirel et al. (2018)
*bla*OXA-1	45 (5.7%)	Acquisition	Beta-lactams	Poirel et al. (2018), Palkovicova et al. (2022), and Rajer and Sandegren (2022)
*bla*TEM	96 (12.3%)	Acquisition	Beta-lactams	Poirel et al. (2018) and Cai et al. (2021)
*bla*TEM-1	62 (7.9%)	Acquisition	Beta-lactams	Dahlberg and Chao (2003) and Poirel et al. (2018)
*bla*TEM-1B	46 (5.9%)	Acquisition	Beta-lactams	Poirel et al. (2018), Alonso-del Valle et al. (2021), Rajer and Sandegren (2022), and Tang et al. (2022)
*dfr*A17	49 (6.3%)	Acquisition	Diaminopyrimidines	Poirel et al. (2018), Brandis et al. (2021), and Tang et al. (2022)
*gyr*A	191 (24.4%)	Mutation	Fluoroquinolones	Marcusson et al. (2009), Huseby et al. (2017), Wang et al. (2017), and Poirel et al. (2018)
*gyr*B	62 (7.9%)	Mutation	Fluoroquinolones	Wang et al. (2017) and Poirel et al. (2018)
*mar*R	108 (13.8%)	Mutation	Fluoroquinolones, beta-lactams	Marcusson et al. (2009), Huseby et al. (2017), Wang et al. (2017), Brandis et al. (2021), and Praski Alzrigat et al. (2021)
mph(A)	62 (7.9%)	Acquisition	Macrolides	Liu et al. (2021) and Rajer and Sandegren (2022)
*par*C	58 (7.4%)	Mutation	Fluoroquinolones	Marcusson et al. (2009) and Poirel et al. (2018)
*rpo*B	209 (26.7%)	Mutation	Fluoroquinolones, beta-lactams, ansamycins	Angst and Hall (2013), Durão et al. (2015), and Brandis et al. (2021)
*rps*L	118 (15.1%)	Mutation	Aminoglycosides	Angst and Hall (2013) and Durão et al. (2015)
*sul*1	84 (10.7%)	Acquisition	Sulfonamides	Enne et al. (2005) and Poirel et al. (2018)
*sul*2	68 (8.7%)	Acquisition	Sulfonamides	Enne (2004) and Poirel et al. (2018)
tet(A)	75 (9.6%)	Acquisition	Tetracyclines	Poirel et al. (2018), Alonso-del Valle et al. (2021), and Zhang et al. (2022)
tet(X)	79 (10.1%)	Acquisition	Tetracyclines	Poirel et al. (2018), Rajer and Sandegren (2022), and Tang et al. (2022)

[1]Total % higher than 100 because strains can contain multiple AMR genes.

average of 2.2 drug families, with a maximum of 5 drug families in a strain carrying AMR mutations in 11 different genes (Brandis et al., 2021). Strains with only acquired AMR genes carried resistance to 4.3 drug families, with a maximum of 8 drug families in a strain carrying plasmids each containing 10–14 AMR genes (Rajer and Sandegren, 2022). A meta-analysis model comparing the relative fitness of strains harboring resistance to each drug family did not show significant differences ($p=1$ for the overall effect; Figure 3) between groups, although a trend toward a lesser relative fitness of polymixin-resistant strains (fitness 0.85) was observed compared to other resistances with relative fitness >0.90.

The relative fitness did not differ significantly between strains with only mutated AMR genes, only acquired AMR genes, or both (Figure 4A). When strains were grouped according to the number of AMR genes, either mutated or acquired, we observed a possible trend toward a lesser relative fitness in strains accumulating more resistance, from 0.96 in strains with only one AMR gene to 0.89 in strains with five or more AMR genes (Figure 4B).

To examine the fitness impact of gene mutation or acquisition on AMR accumulation, we used meta-regression models to quantify the change of relative fitness with each additional mutated or acquired AMR gene (Figures 5A,C). Each additional mutated AMR gene significantly decreased fitness by 3.7% (95% CI, 3.1 to 4.3%) while each additional AMR gene acquisition decreased fitness by 1.1% (95% CI, 0.5 to 1.7%). Hence, the cost of each additional AMR mutation was more than 3-fold higher than the cost of each additional AMR gene acquisition. Using the same meta-regression method, we observed that the accumulation of resistance to several antimicrobial families was more costly when AMR resulted from gene mutations (Figures 5B,D). Each additional drug family in the resistance spectrum significantly decreased the relative fitness by 2.1% (95% CI 0.9 to 3.2%) in strains with only mutated AMR genes. In strains with only acquired AMR genes, however, the decrease of relative fitness per additional drug family was only 1.1% and not significantly different from zero (95% CI, −0.003 to 2.3%).

3.3. Fitness decreases with the size but not the number of AMR plasmids

Plasmids require energy for their own maintenance and replication. We examined whether the relative fitness was influenced by plasmid

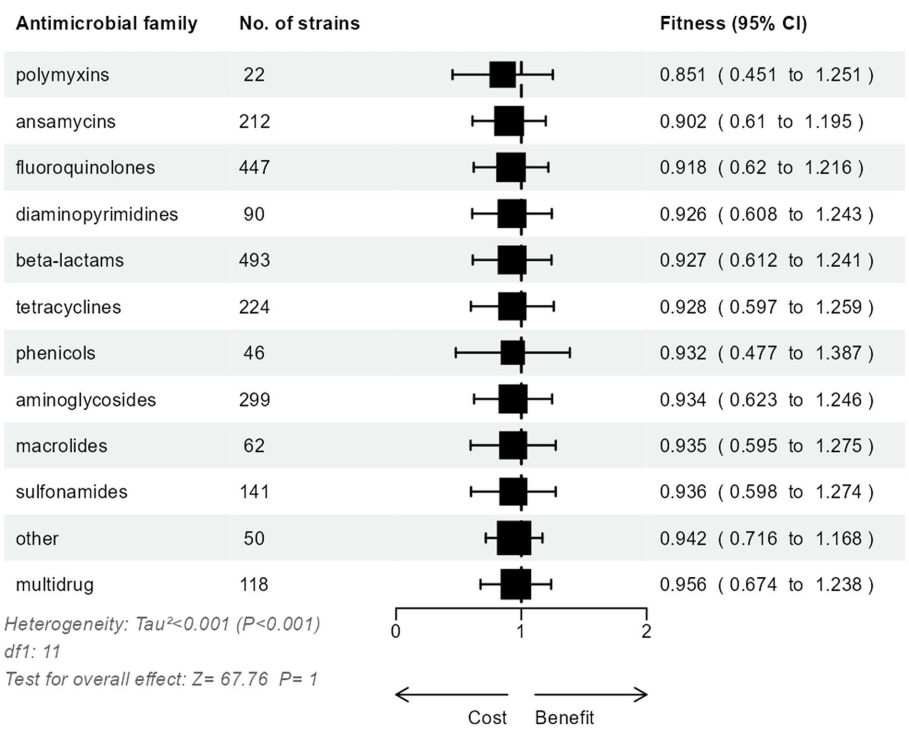


FIGURE 3
Three-level meta-analysis of the relative fitness of *E. coli* strains (n=783) resistant to various antimicrobial drug families. Square markers denote effect size, marker size is proportional to the group weight in the overall effect, and error bars denote the 95% confidence interval. No significant overall effect of the resistance to specific drug families was observed.

characteristics such as their size, number per strain, and incompatibility group (Inc). Using the meta-regression approach on the subset of strains with exactly one AMR plasmid (n=266), we observed that plasmid size, that averaged to 83 kbp, significantly correlated with a decrease of fitness cost of 0.06% per kbp (95% CI, 0.03–0.09%; Figure 6A). In strains with one or more plasmids, however, the number of plasmids had no significant effect on relative fitness (Figure 6B). The conjugative nature of the plasmid, compared to non-conjugative plasmids and to plasmids engineered *in vitro* as vectors of AMR genes, had no significant impact on relative fitness (Supplementary Figure S5A). Nine major incompatibility groups of plasmids were represented in our dataset. The relative fitness did not significantly differ between groups (Supplementary Figure S5B), although possible trends were observed. The most frequent incompatibility group was IncF (n=88 strains), with an average relative fitness of 0.928. Strains with IncA/C and IncX plasmids appeared more costly than the other Inc. groups (0.747 and 0.872, respectively), while IncP plasmids seemed to have a lesser impact on fitness (Supplementary Figure S5B).

4. Discussion

In this meta-analysis of 783 drug-resistant *E. coli* strains from 46 studies, we observed that the accumulation of AMR resulting from mutations of chromosomal genes entails a 3-fold stronger fitness cost than the accumulation of transferable AMR genes. This phenomenon may contribute to the observed dominance of transferable AMR genes in the current multidrug resistance epidemic in enterobacteria.

The most problematic lineages of multidrug resistant enterobacteria have evolved through the acquisition of horizontally transferred, mainly plasmid-borne, AMR genes rather than the accumulation of AMR mutations (Acar and Röstel, 2001). Previous work has repeatedly identified that mutational AMR is more costly than AMR gene acquisition (Vogwill and MacLean, 2015), however, the impact of the accumulation of AMR determinants through both mechanisms had never been compared. Here, we find that even though the accumulation of acquired AMR genes entails a significant cost, this cost is much reduced compared to the accumulation of AMR mutations (Figure 5). Thus, the evolutionary pathway to multidrug resistance may be strongly facilitated by the availability of transferable AMR genes in the environment, at least in species with HGT capabilities. In contrast with gene acquisition, mutational resistance is available to virtually all species in all environments. This may explain why mutational resistance prevails in species without HGT capabilities such as *Mycobacterium tuberculosis* (Merker et al., 2018), or in confined environments with a limited supply of mobile genetic elements, such as chronic lung infections with *Pseudomonas aeruginosa* (López-Causapé et al., 2018).

Interestingly, the number of AMR genes as well as the size of AMR plasmids were the dominant factors influencing the relative fitness in our analyses. Other plasmid characteristics such as their conjugative or mobilizable nature, or their incompatibility group, had a comparatively negligible impact (Supplementary Figure S5). This suggests that the AMR genes themselves and their number, rather than their plasmid vehicle, are the main source of fitness cost. This is important as the number of AMR genes present in each plasmid may vary considerably

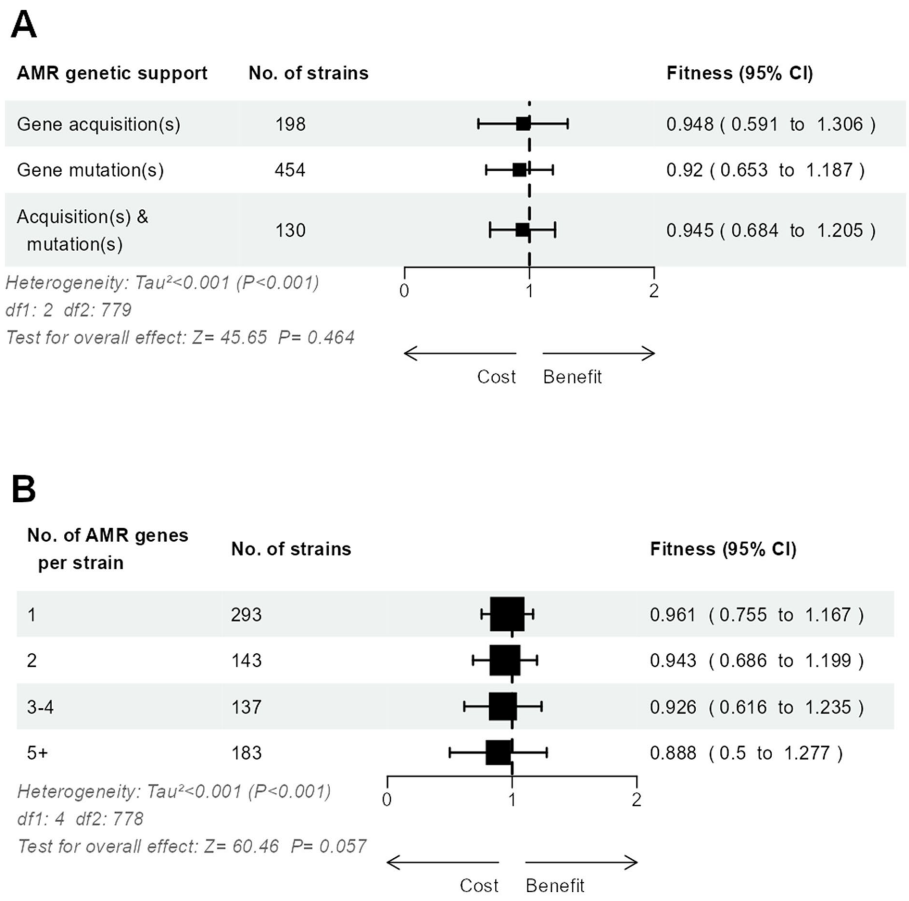


FIGURE 4
Three-level meta-analysis of the relative fitness of *E. coli* strains (n=783) with mutated and/or acquired AMR genes **(A)**, and according to the number of mutated and acquired AMR genes per strain **(B)**. Shown are forest plots in which square markers denote effect size, marker size is proportional to the group weight in the overall effect, and error bars denote the 95% confidence interval.

across mobilizable, non-mobilizable or conjugative plasmids (in average 2.7, 3.5 and 4.7 AMR genes per plasmid, respectively; Che et al. (2021)). As previously discussed by Vogwill and McLean in their multi-species meta-analysis of fitness cost (Vogwill and MacLean, 2015), AMR genes may be more recent in evolution than other plasmid genes and this shorter adaptation time may contribute to a comparatively higher cost of AMR genes. This hypothesis is supported by recent data suggesting that the fitness cost of plasmid genes results from conflicts with other genes that can be quickly alleviated by fitness-cost compensatory mutations (Yano et al., 2016; Hall et al., 2021). Of note, several of our findings are in contrast with those of the Vogwill and McLean study. They found a significant difference in the relative fitness of mutational and acquired AMR independent of AMR accumulation, while in our study the difference was only revealed by AMR accumulation (Figures 4, 5). In addition, the significant impact of plasmid size on relative fitness that we observed in isolates carrying a single AMR plasmid (Figure 6A) was not observed in their analysis. We speculate that, beyond the influence of new data accumulated since the Vogwill and McLean study, in which *E. coli* accounted for less than 30% of the data (Supplementary Table S1), our focusing on a single species in a multi-level meta-analysis framework may explain these differences.

Our study did not address the possible reasons for the lesser fitness cost of transferable AMR compared to mutational AMR, yet, recent data suggest that compensatory evolution may play a key role in this lesser fitness cost. Indeed, the fitness cost of AMR is a transient property that may decline if a sustained selection pressure enables compensatory evolution to modulate fitness cost. Compensatory evolution, by which additional genomic alterations reduce the fitness cost of AMR but not its resistance level, may be involved in both mutation- and HGT-driven resistance (Durão et al., 2018; Yang et al., 2020; Hall et al., 2021; Patel and Matange, 2021). We speculate that compensatory evolution, in combination with the additional plasticity provided by plasmid-borne AMR, may contribute to preserve both the fitness and the resistance level provided by transferable AMR genes. Indeed, in antibiotic-free environments, the reversion to a more fit, susceptible phenotype is not only driven by the outgrowth of resistant variants by their susceptible ancestors (if they survived), but also by the *de novo* emergence of susceptible variants in the resistant population. Plasmid loss provides an efficient means for this emergence, while the reversion of mutational resistance follows a more complex pathway in which additional mutations contribute to reducing both the fitness cost and the resistance level conveyed by the initial AMR mutation (Dunai et al., 2019).

We acknowledge several limitations to our study. The fitness cost of AMR is difficult to measure in controlled conditions and the experimental procedures are not standardized, which can introduce noise and bias in meta-analysis approaches. For instance, a

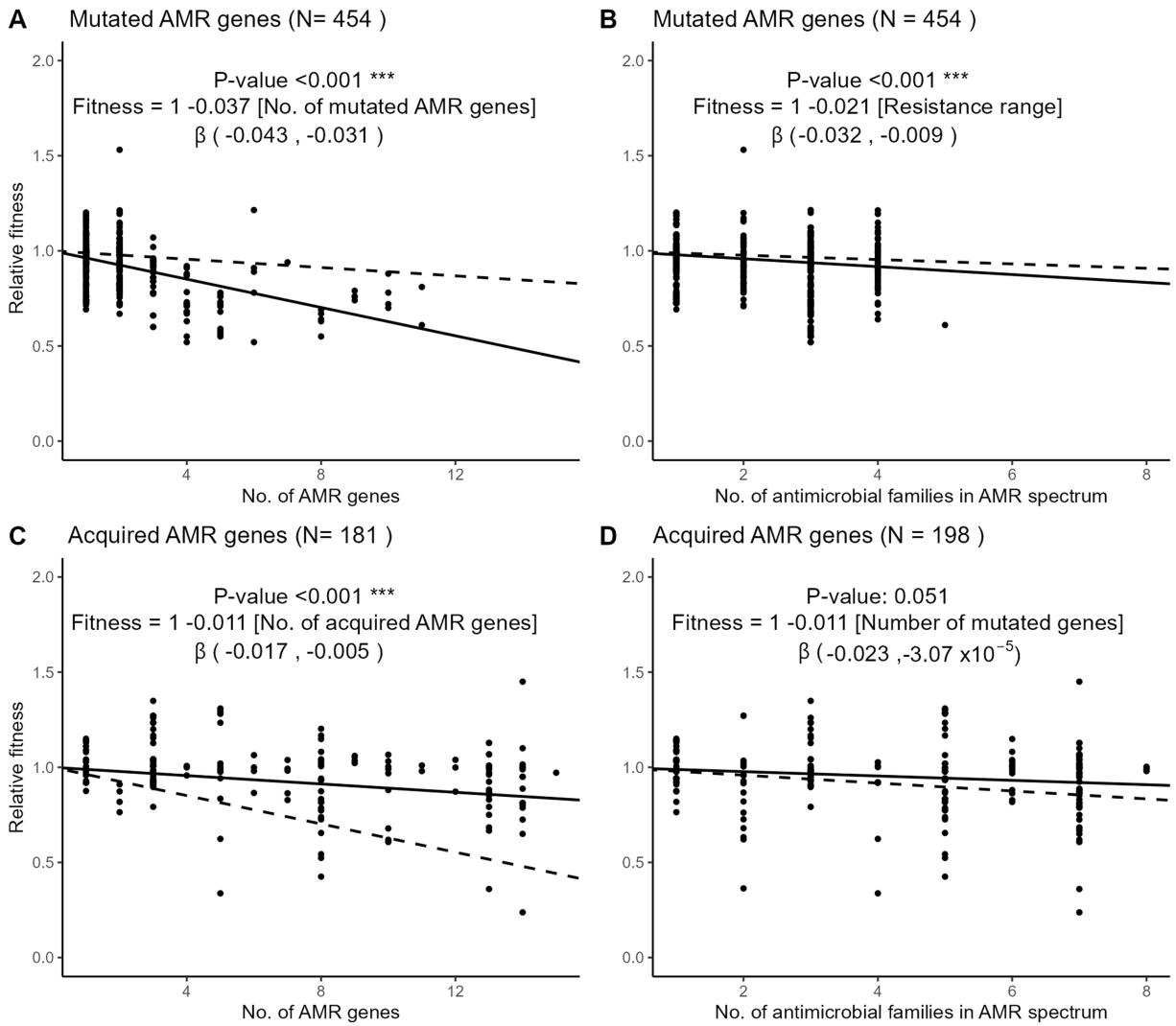


FIGURE 5
Three-level meta-regressions of the relative fitness of *E. coli* strains with either mutated **(A,B)** or acquired **(C,D)** AMR genes as a function of the number of AMR genes **(A,C)** or resistant drug families **(B,D)** per strain. Shown are meta-regression plots where each point denotes a strain and the solid line denotes the regression slope with a fixed intercept of 1. Dashed lines denote the regression slope in the comparator group, that is, the dashed line in panel **A** (mutated AMR genes) represents the regression slope of panel **C** (acquired AMR genes). The regression coefficient estimates the decrease of relative fitness with each additional AMR gene or resistant drug family. The 95% confidence interval of the β regression coefficient is shown in brackets. Antimicrobial drug families are listed in Figure 3. Sample sizes differ in panels **C,D** because of the exclusion of strains with missing information on either the no. of AMR genes or the no. of drug family resistances.

meta-analysis by Melnyk et al. (2015) found that the relative fitness assessed by the W_r method (ratio of Malthusian parameters) exceeded by a factor of 1.7 those assessed by the W_t method (Dykhuizen and Hartl, 1983; Dean and Dykhuizen, 1988; Dykhuizen, 1990). Yet, in our study this bias did not translate into a measurable difference of fitness estimations across studies, suggesting that other sources of variation outweighed the influence of the estimation method for biologically relevant values of relative fitness. Indeed, variations of culture media or competition assay duration were common in our dataset, but the observed influence on the estimation of fitness cost was at most moderate (Supplementary Figure S4), enabling for the joint analysis of those experiments in a multi-level analysis framework. Based on previous findings (Melnyk et al., 2015), a major source of variation in relative fitness is the bacterial species itself. Here, we avoided this source of noise by focusing on a single species, although we cannot exclude an unseen influence of the strains' genetic backgrounds, which were not available for analysis. In addition, our results are focused on the *E. coli* species and further research is needed to determine whether our conclusions hold in other HGT-capable species. Virtually all acquired AMR genes in our analysis were plasmid-borne and our results may not hold for other HGT vehicles such as bacteriophages. Bacteriophages have been suspected to contribute to AMR HGT in several species including *E. coli* (Billard-Pomares et al., 2014), however their relative contribution to AMR is still unclear, considering that transduction rates are several orders of magnitude lower than conjugation rates (Nazarian et al., 2018). Finally, data remain scarce regarding the fitness impact of AMR *in vivo*, and it is still unclear whether conclusions from *in vitro* experiments can be reproduced in animal models (Roux et al., 2015; Wheatley et al., 2021).

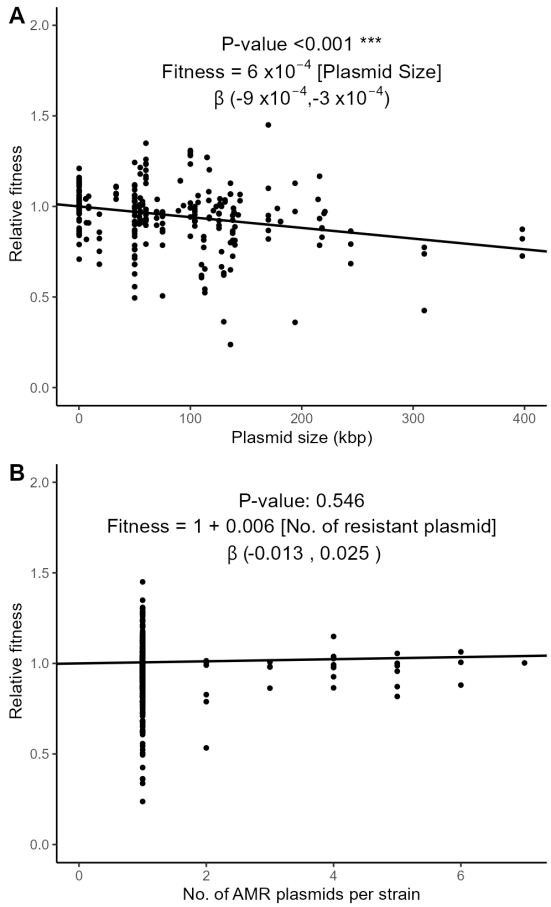


FIGURE 6
Three-level meta-regression of the relative fitness of *E. coli* strains with plasmid-borne AMR genes based on the size **(A)** and number **(B)** of AMR plasmids. The effect of plasmid size **(A)** was analyzed in the subset of strains harboring exactly one AMR plasmid (*n*=266). The effect of the number of plasmids per strain **(B)** was analyzed in the strains with at least one AMR plasmid (*n*=325), with a fixed intercept of 1. The regression coefficient estimates the decrease of relative fitness with each additional kbp in plasmid length **(A)** or each additional AMR plasmid **(B)**. The 95% confidence intervals of the β regression coefficient are shown in brackets.

To conclude, our results highlight that gene acquisition is more efficient than the accumulation of mutations to evolve multidrug resistance in *E. coli*. Although it is still unclear whether this finding may be generalized to most bacterial species, the lesser cost of horizontal transfer compared to mutational AMR stresses the need to monitor and control the diffusion of AMR plasmids as closely as the diffusion of resistant bacteria in the environment.

Author contributions

MV contributed to the conception, design of the study, organization of the database and statistical analysis and wrote the first draft of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was founded by the French National Research Agency, part of the AAPG 2020 programme (ANR-20-CE35-0012 to J-PR).

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.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1186920/full#supplementary-material

References

Acar, J., and Röstel, B. (2001). Antimicrobial resistance: an overview: -EN- -FR- -ES. *Rev. Sci. Tech. OIE* 20, 797–810. doi: 10.20506/rst.20.3.1309

Alonso-del Valle, A., León-Sampedro, R., Rodríguez-Beltrán, J., DelaFuente, J., HernándezGarcía, M., Ruiz-Garbajosa, P., et al. (2021). Variability of plasmid fitness effects contributes to plasmid persistence in bacterial communities. *Nat Commun* 12:2653. doi: 10.1038/s41467-021-22849-y

Andersson, D. I., and Levin, B. R. (1999). The biological cost of antibiotic resistance. *Curr. Opin. Microbiol.* 2, 489–493. doi: 10.1016/S1369-5274(99)00005-3

Angst, D. C., and Hall, A. R. (2013). The cost of antibiotic resistance depends on evolutionary history in Escherichia coli. *BMC Evol Biol* 13:163. doi: 10.1186/1471-2148-13-163

Balduzzi, S., Rücker, G., and Schwarzer, G. (2019). How to perform a meta-analysis with R: a practical tutorial. *Evid. Based Ment. Health* 22, 153–160. doi: 10.1136/ebmental-2019-300117

Billard-Pomares, T., Fouteau, S., Jacquet, M. E., Roche, D., Barbe, V., Castellanos, M., et al. (2014). Characterization of a P1-like bacteriophage carrying an SHV-2 extended-Spectrum β-lactamase from an *Escherichia coli* strain. *Antimicrob. Agents Chemother.* 58, 6550–6557. doi: 10.1128/AAC.03183-14

Brandis, G., Granström, S., Leber, A. T., Bartke, K., Garoff, L., Cao, S., et al. (2021). Mutant RNA polymerase can reduce susceptibility to antibiotics via ppGpp-independent induction of a stringent-like response. *J. Antimicrob. Chemother.* 76, 606–615. doi: 10.1093/jac/dkaa469

Cai, W., Tang, F., Jiang, L., Li, R., Wang, Z., and Liu, Y. (2021). Histone-like nucleoid structuring protein modulates the fitness of tet(X4)-bearing IncX1 plasmids in gram-negative bacteria. *Front. Microbiol.* 12:763288. doi: 10.3389/fmicb.2021.763288

Che, Y., Yang, Y., Xu, X., Břinda, K., Polz, M. F., Hanage, W. P., et al. (2021). Conjugative plasmids interact with insertion sequences to shape the horizontal transfer of antimicrobial resistance genes. *Proc. Natl. Acad. Sci. U. S. A.* 118:e2008731118. doi: 10.1073/pnas.2008731118

Dahlberg, C., and Chao, L. (2003). Amelioration of the cost of conjugative plasmid carriage in Escherichia coli K12. *Genetics* 165, 1641–1649. doi: 10.1093/genetics/165.4.1641

Dayimu, A. (2022). Forestploter: Create Flexible Forest plot. Available at: https://CRAN.R-project.org/package=forestploter. (Accessed May 25, 2023)

Dean, A. M., and Dykhuizen, D. E. (1988). Fitness effects of amino acid replacements in the fl- galactosidase of Escherichia coZil. *Mol. Biol. Evol.* 5, 469–485.

Dunai, A., Spohn, R., Farkas, Z., Lázár, V., Györkei, Á., Apjok, G., et al. (2019). Rapid decline of bacterial drug-resistance in an antibiotic-free environment through phenotypic reversion. *elife* 8:e47088. doi: 10.7554/eLife.47088

Durão, P., Trindade, S., Sousa, A., and Gordo, I. (2015). Multiple resistance at no cost: Rifampicin and Streptomycin a dangerous liaison in the spread of antibiotic resistance. *Molecular Biology and Evolution* 32, 2675–2680. doi: 10.1093/molbev/msv143

Durão, P., Balbontín, R., and Gordo, I. (2018). Evolutionary mechanisms shaping the maintenance of antibiotic resistance. *Trends Microbiol.* 26, 677–691. doi: 10.1016/j.tim.2018.01.005

Dykhuizen, D. E. (1990). Experimental studies of natural selection in Bacteria. *Annu. Rev. Ecol. Evol. Syst.* 21, 373–398. doi: 10.1146/annurev.es.21.110190.002105

Dykhuizen, D. E., and Hartl, D. L. (1983). Selection in chemostats. *Microbiol. Rev.* 47, 150–168. doi: 10.1128/mr.47.2.150-168.1983

Enne, V. I. (2004). Enhancement of host fitness by the sul2-coding plasmid p9123 in the absence of selective pressure. *Journal of Antimicrobial Chemotherapy* 53, 958–963. doi: 10.1093/jac/dkh217

Enne, V. I., Delsol, A. A., Davis, G. R., Hayward, S. L., Roe, J. M., and Bennett, P. M. (2005). Assessment of the fitness impacts on Escherichia coli of acquisition of antibiotic resistance genes encoded by different types of genetic element. *Journal of Antimicrobial Chemotherapy* 56, 544–551. doi: 10.1093/jac/dki255

Hall, J. P. J., Wright, R. C. T., Harrison, E., Muddiman, K. J., Wood, A. J., Paterson, S., et al. (2021). Plasmid fitness costs are caused by specific genetic conflicts enabling resolution by compensatory mutation. *PLoS Biol.* 19:e3001225. doi: 10.1371/journal.pbio.3001225

Harrer, M., Cuijpers, P., Furukawa, T., and Ebert, D. (2019). Dmetar: Companion R Package for the Guide "Doing Meta-Analysis in R". R Package Version 0.0.9000. Available at: http://dmetar.protectlab.org/. (Accessed Jan 27, 2023)

Harrer, M., Cuijpers, P., Furukawa, T., and Ebert, D. (2022). *Doing Meta-Analysis with R - A Hands-on Guide. 1st* Edn Chapman & Hall. Available at: https://www.taylorfrancis.com/books/mono/10.1201/9781003107347/meta-analysis-mathias-harrer-pim-cuijpers-toshi-furukawa-david-ebert

Higgins, J. P. T., and Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Stat. Med.* 21, 1539–1558. doi: 10.1002/sim.1186

Huseby, D. L., Pietsch, F., Brandis, G., Garoff, L., Tegehall, A., and Hughes, D. (2017). Mutation supply and relative fitness shape the genotypes of ciprofloxacin-resistant Escherichia coli. *Mol Biol Evol* msx052. doi: 10.1093/molbev/msx052

Jian, Z., Zeng, L., Xu, T., Sun, S., Yan, S., Yang, L., et al. (2021). Antibiotic resistance genes in bacteria: occurrence, spread, and control. *J. Basic Microbiol.* 61, 1049–1070. doi: 10.1002/jobm.202100201

Lenski, R. E. (1991). "Quantifying fitness and gene stability in microorganisms" in *Assessing ecological risks of biotechnology (Elsevier)*, 173–192. doi: 10.1016/B978-0-409-90199-3.50015-2

Lenski, R. E., Rose, M. R., Simpson, S. C., and Tadler, S. C. (1991). Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am. Nat.* 138, 1315–1341. doi: 10.1086/285289

Levin, B. R., Perrot, V., and Walker, N. (2000). Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. *Genetics* 154, 985–997. doi: 10.1093/genetics/154.3.985

Liu, Z., Wang, Z., Lu, X., Peng, K., Chen, S., He, S., et al. (2021). Structural Diversity, Fitness cost, and stability of a BlaNDM-1-bearing cointegrate plasmid in Klebsiella pneumoniae and Escherichia coli. *Microorganisms* 9:2435. doi: 10.3390/microorganisms9122435

López-Causapé, C., Cabot, G., del Barrio-Tofiño, E., and Oliver, A. (2018). The versatile mutational resistome of *Pseudomonas aeruginosa*. *Front. Microbiol.* 9:685. doi: 10.3389/fmicb.2018.00685

Marcusson, L. L., Frimodt-Møller, N., and Hughes, D. (2009). Interplay in the selection of fluoroquinolone resistance and bacterial fitness. *PLoS Pathog* 5:e1000541. doi: 10.1371/journal.ppat.1000541

McArthur, A. G., Waglechner, N., Nizam, F., Yan, A., Azad, M. A., Baylay, A. J., et al. (2013). The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.* 57, 3348–3357. doi: 10.1128/AAC.00419-13

Melnyk, A. H., Wong, A., and Kassen, R. (2015). The fitness costs of antibiotic resistance mutations. *Evol. Appl.* 8, 273–283. doi: 10.1111/eva.12196

Merker, M., Barbier, M., Cox, H., Rasigade, J.-P., Feuerriegel, S., Kohl, T. A., et al. (2018). Compensatory evolution drives multidrug-resistant tuberculosis in Central Asia. *elife* 7:e38200. doi: 10.7554/eLife.38200

Nadeem, S. F., Gohar, U. F., Tahir, S. F., Mukhtar, H., Pornpukdeewattana, S., Nukthamna, P., et al. (2020). Antimicrobial resistance: more than 70 years of war between humans and bacteria. *Crit. Rev. Microbiol.* 46, 578–599. doi: 10.1080/1040841X.2020.1813687

Nazarian, P., Tran, F., and Boedicker, J. Q. (2018). Modeling multispecies gene flow dynamics reveals the unique roles of different horizontal gene transfer mechanisms. *Front. Microbiol.* 9:2978. doi: 10.3389/fmicb.2018.02978

Palkovicova, J., Sukkar, I., Delafuente, J., Valcek, A., Medvecky, M., Jamborova, I., et al. (2022). Fitness effects of bla CTX-M-15-harbouring F2:A1:B− plasmids on their native Escherichia coli ST131 H 30Rx hosts. *Journal of Antimicrobial Chemotherapy* 77, 2960–2963. doi: 10.1093/jac/dkac250

Patel, V., and Matange, N. (2021). Adaptation and compensation in a bacterial gene regulatory network evolving under antibiotic selection. *elife* 10:e70931. doi: 10.7554/eLife.70931

Poirel, L., Madec, J.-Y., Lupo, A., Schink, A.-K., Kieffer, N., Nordmann, P., et al. (2018). Antimicrobial resistance in *Escherichia coli*. *Microbiol Spectr* 6:6.4.14. doi: 10.1128/microbiolspec.ARBA-0026-2017

Pope, C. F., McHugh, T. D., and Gillespie, S. H. (2010). "Methods to determine fitness in Bacteria" in *Antibiotic Resistance Protocols Methods in Molecular Biology*. eds. S. H. Gillespie and T. D. McHugh (Totowa, NJ: Humana Press), 113–121.

Praski Alzrigat, L., Huseby, D. L., Brandis, G., and Hughes, D. (2021). Resistance/fitness trade-off is a barrier to the evolution of MarR inactivation mutants in *Escherichia coli*. *J. Antimicrob. Chemother.* 76, 77–83. doi: 10.1093/jac/dkaa417

Pu, C., Yu, Y., Diao, J., Gong, X., Li, J., and Sun, Y. (2019). Exploring the persistence and spreading of antibiotic resistance from manure to biocompost, soils and vegetables. *Sci. Total Environ.* 688, 262–269. doi: 10.1016/j.scitotenv.2019.06.081

R Core Team (2022). R: A Language and Environment for Statistical Computing. Available at: https://www.R-project.org/. (Accessed May 25, 2023)

Rajer, F., and Sandegren, L. (2022). The role of antibiotic resistance genes in the fitness cost of multiresistance plasmids. *mBio* 13:e03552-21. doi: 10.1128/mbio.03552-21

Roux, D., Danilchanka, O., Guillard, T., Cattoir, V., Aschard, H., Fu, Y., et al. (2015). Fitness cost of antibiotic susceptibility during bacterial infection. *Sci. Transl. Med.* 7:297ra114. doi: 10.1126/scitranslmed.aab1621

Silva, R. F., Mendonça, S. C. M., Carvalho, L. M., Reis, A. M., Gordo, I., Trindade, S., et al. (2011). Pervasive sign epistasis between conjugative plasmids and drug-resistance chromosomal mutations. *PLoS Genet* 7:e1002181. doi: 10.1371/journal.pgen.1002181

Tang, F., Cai, W., Jiang, L., Wang, Z., and Liu, Y. (2022). Large-Scale Analysis of Fitness Cost of tet(X4)-Positive Plasmids in Escherichia coli. *Front. Cell. Infect. Microbiol.* 12:798802. doi: 10.3389/fcimb.2022.798802

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* 36, 1–48. doi: 10.18637/jss.v036.i03

Vogwill, T., and MacLean, R. C. (2015). The genetic basis of the fitness costs of antimicrobial resistance: a meta-analysis approach. *Evol. Appl.* 8, 284–295. doi: 10.1111/eva.12202

Wang, J., Guo, Z.-W., Zhi, C.-P., Yang, T., Zhao, J.-J., Chen, X.-J., et al. (2017). Impact of plasmid-borne oqxAB on the development of fluoroquinolone resistance and bacterial fitness in Escherichia coli. *Journal of Antimicrobial Chemotherapy* 72, 1293–1302. doi: 10.1093/jac/dkw576

Wang, T., Kunze, C., and Dunlop, M. J. (2019). Salicylate increases fitness cost associated with MarA-mediated antibiotic resistance. *Biophys. J.* 117, 563–571. doi: 10.1016/j.bpj.2019.07.005

Wheatley, R., Diaz Caballero, J., Kapel, N., de Winter, F. H. R., Jangir, P., Quinn, A., et al. (2021). Rapid evolution and host immunity drive the rise and fall of carbapenem resistance during an acute *Pseudomonas aeruginosa* infection. *Nat. Commun.* 12:2460. doi: 10.1038/s41467-021-22814-9

Yang, Q. E., MacLean, C., Papkou, A., Pritchard, M., Powell, L., Thomas, D., et al. (2020). Compensatory mutations modulate the competitiveness and dynamics of plasmid-mediated colistin resistance in *Escherichia coli* clones. *ISME J.* 14, 861–865. doi: 10.1038/s41396-019-0578-6

Yano, H., Wegrzyn, K., Loftie-Eaton, W., Johnson, J., Deckert, G. E., Rogers, L. M., et al. (2016). Evolved plasmid-host interactions reduce plasmid interference cost. *Mol. Microbiol.* 101, 743–756. doi: 10.1111/mmi.13407

Zhang, P., Mao, D., Gao, H., Zheng, L., Chen, Z., Gao, Y., et al. (2022). Colonization of gut microbiota by plasmid-carrying bacteria is facilitated by evolutionary adaptation to antibiotic treatment. *ISME J* 16, 1284–1293. doi: 10.1038/s41396-021-01171-x

Frontiers in Microbiology

Explores the habitable world and the potential of microbial life

The largest and most cited microbiology journal which advances our understanding of the role microbes play in addressing global challenges such as healthcare, food security, and climate change.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact