Computational Molecular Biology 2025, Vol.15, No.6, 273-281
http://bioscipublisher.com/index.php/cmb

Research Insight Open Access

Genome Assembly and Comparative Genomics of a Novel Extremophilic Bacterium

Yinghua Chen, Hui Xiang, Zhongqi Wu
Institute of Life Science, Jiyang College of Zhejiang A&F University, Zhuji, 311800, Zhejiang, China
Corresponding author: zhongqi.wu@jicat.org

Computational Molecular Biology, 2025, Vol.15, No.6 doi: 10.5376/cmb.2025.15.0027
Received: 07 Sep., 2025
Accepted: 20 Oct., 2025
Published: 10 Nov., 2025

Preferred citation for this article:
Chen Y.H., Xiang H., and Wu Z.Q., 2025, Genome assembly and comparative genomics of a novel extremophilic bacterium, Computational Molecular Biology, 15(6): 273-281 (doi: 10.5376/cmb.2025.15.0027)

Abstract Extremophiles have evolved unique physiological mechanisms and genomic characteristics in extreme environments such as high salt, high temperature, high pressure, and strong acid. Studying their genetic basis is of great significance for revealing the adaptive evolution mechanisms of microorganisms and discovering functional genes with biotechnological value. This study aims to assemble and annotate the genome of a newly isolated polar bacterium and analyze its metabolic potential, environmental adaptation mechanism and evolutionary characteristics through comparative genomics. Functional annotations reveal that the genome of this strain is rich in key functional genes related to salt tolerance, heat resistance, heavy metal resistance, etc. In the comparative analysis with other known polar bacteria, conserved core gene clusters, species-specific gene islands, and the expansion of gene families in response to environmental stress were discovered. Case studies show that it has application potential in the development of industrial enzymes and the construction of synthetic biology platforms. This study provides a new genome-level perspective for understanding the adaptation mechanisms of polar microorganisms and lays a foundation for their functional exploration and application development.

Keywords Polar microorganisms; Genome assembly; Comparative genomics; Salt resistance; Phylogeny

1 Introduction

Not all lives are keen on the comfortable environment like a greenhouse. Some microorganisms, on the contrary, prefer high temperatures, extreme cold, strong acids, heavy salts and even radiation. These places are not suitable for most life forms, but they are common for extremist microorganisms. In environments like hot springs, polar ice caps and alkaline salt lakes, they not only survive tenaciously but also play a "behind-the-scenes role" in maintaining the ecosystem cycle. Although they seem far away from our lives, these organisms offer a window to understand the limits of life and are very inspiring for the exploration of celestial life, the development of biotechnology, and even the research of environmental restoration (Arias et al., 2023; Gomez et al., 2024).

However, to figure out exactly what skills they have relied on to "survive", mere observation is far from enough. Methods such as genome assembly and comparative analysis are the true keys that enable us to enter the internal structure of their genetics. In the past, problems such as high GC content and numerous genomic duplications were indeed troublesome. Now, with high-throughput sequencing and long-read technologies, even these "tough nuts to crack" can be successfully tackled (Dong, 2024). By comparing the genomes of these extreme bacteria, not only can we identify genes related to membrane stability, DNA repair or stress response, but also some new families and mechanisms that were not previously noticed may be unearthed. Some may even change our understanding of microbial diversity and evolutionary patterns (Zhang et al., 2021a).

This study will utilize the most advanced sequencing and bioinformatics methods to assemble and analyze the genome of a novel extreme microorganism isolated from a unique extreme environment, construct a high-quality genomic sequence, conduct comparative genomics analysis with related extreme microorganisms, and identify the genetic determinants of its extreme tolerance. The scientific significance of this research lies in expanding the catalogue of extremist microorganism genomes and revealing the molecular adaptation mechanisms that may have an impact on the application of biotechnology and evolutionary biology.

2 Sample Collection and Sequencing Strategies

2.1 Sampling from extreme environments and bacterial isolation

These "durable" microorganisms can not be found everywhere. Most of the true extreme bacteria grow in places like hot springs, salt lakes and acidic mines - the conditions are harsh, but they just survive well. However, extracting bacteria from these environments is not as simple as just scooping up a ladle of water. We have to find a way to restore them to their original growth state; otherwise, they will "die" as soon as they enter the laboratory. Often, it is necessary to carefully cultivate under simulated native conditions in order to select the truly tolerant batch. Although it takes a lot of effort, only in this way can pure and stable strains be obtained for subsequent genomic analysis. These operations also lay the foundation for us to understand how they withstand extreme coercion (Verma et al., 2024).

2.2 DNA extraction and selection of sequencing platforms (Illumina, Nanopore, PacBio, etc.)

Extracting DNA from these microorganisms is easier said than done. For instance, if the cell wall is particularly hard or the sample contains some strange environmental impurities, conventional methods may not work. After successfully extracting high-quality DNA, the next step is to select a sequencing platform. The combination of short-read high-precision Illumina and long-read ONT or PacBio with strong coverage is currently the most common hybrid strategy. Especially when dealing with samples with unstable GC content or many repetitive sequences, using only one platform often yields mediocre results. Typically, the research will first use long read length (ONT or PacBio) for preliminary assembly, and then use Illumina for fine-tuning. The overall effect is stable and a considerable amount of budget is saved (Goldstein et al., 2018; Neal-McKinney et al., 2021).

2.3 Data quality control and preprocessing approaches

After the measurement, the data cannot be used directly. Quality control is the crucial step next. For instance, first, the connectors need to be removed, low-quality reads filtered out, and contaminated fragments mixed in filtered out. All these tasks must be done thoroughly one by one. Otherwise, errors are likely to occur during the subsequent assembly. Especially when multiple sequencing platforms are used in combination, it is necessary to carefully examine the error rate and coverage distribution. Some long-read platforms themselves have insertion or missing issues. Using short-read high-fidelity data for error correction is one of the common operations. Nowadays, most processes can be automated. Basically, from raw data to available assembly data, the entire set of preprocessing can be seamlessly connected. This is particularly important for the research object of extremist microorganisms (De Maio et al., 2019; Olagoke et al., 2025).

3 Genome Assembly and Quality Assessment

3.1 Genome assembly strategies (de novo, hybrid, etc.)

The genome assembly of extremist microorganisms usually adopts a hybrid strategy, combining short-read sequencing (such as Illumina) with long-read sequencing platforms (such as Oxford Nanopore Technologies (ONT) or PacBio). Hybrid assembly tools, especially Unicycler, combine the high precision of short reads with the long-distance continuity of long reads, thereby generating more complete and continuous genomes than de novo assembly based solely on a single technology. This method can effectively analyze the common complex genomic regions, repetitive sequences and structural variations in extremist microorganisms, thereby achieving chromosome-level assembly and improving the accuracy and completeness of the assembly. Studies comparing different strategies have shown that hybrid assembly is superior to pure short read or long read assembly in terms of continuity and genetic integrity, while also balancing sequencing costs and DNA initiation quantity requirements (Wick et al., 2017; Chen et al., 2020).

3.2 Evaluation of contigs/scaffolds and statistics on N50 and GC content

Assembly quality is usually evaluated using indicators such as the number of overlapping groups or scaffolds, N50 values, and the distribution of GC content. A higher N50 value indicates greater continuity, reflecting longer assembly sequences and better representing genomic structure. The genomes of extremely thermophilic bacteria usually exhibit different GC contents, which poses challenges to assembly algorithms; Therefore, evaluating the GC content is helpful for verifying the accuracy of assembly and detecting potential deviations. Compared with

assembly using only short-read sequences, hybrid assembly usually results in fewer overlapping groups and higher N50 values, indicating a more complete and continuous genome. Monitoring the consistency of GC content with the expected value of this species can further support the reliability of assembly (Zhang et al., 2021b).

3.3 Application of BUSCO, QUAST and other tools for completeness assessment

The integrity and quality assessment of genomic assembly mainly rely on tools such as BUSCO and QUAST, which respectively provide indicators of biological and technical significance. BUSCO estimates the integrity and redundancy of the genome by evaluating nearly ubiquitous single-copy direct homologous genes, thereby gaining a deeper understanding of the integrity of genetic content. QUAST reports assembly statistics, including overlapping group counts, N50 values, and incorrect assembly rates, thereby enabling the detection of structural errors. Other tools such as CheckM2 utilize machine learning to predict integrity and contamination, especially for metagenomic assembled genomes. Combining these assessment methods can ensure that the assembled genome is structurally accurate and biologically complete, which is crucial for subsequent comparative genomics and functional analysis of extremist microorganisms (Manni et al., 2021; Chklovski et al., 2022).

4 Genome Annotation and Functional Analysis

4.1 Coding sequence prediction and structural annotation (Prokka, RAST, etc.)

To understand the "genetic ledger" of an extremist microorganism, it is usually necessary to start with which coding regions it has. Often, researchers directly use toolkits like Prokka and RAST, which not only saves time but also facilitates standardization in the later stage. The operation interfaces of these programs may not be overly complicated, but in fact, they integrate multiple sets of gene prediction models and databases behind the scenes, capable of generating a complete set of annotation results at once, including rRNA, tRNA, and even hypothetical proteins. For instance, Prokka has been repeatedly used in various bacterial genomes. Stability and speed are its advantages (De Almeida et al., 2023). However, even if the tools are powerful, the "assumed proteins" automatically annotated still need to be guarded against - they often represent unknown functions, which is precisely the most attractive part of extremist microorganisms.

4.2 Functional annotation and database comparisons (COG, KEGG, Pfam)

The list of genes derived from automatic annotation does not specify exactly what these genes do. This step still depends on the results of database comparisons, such as the commonly used COG, KEGG and Pfam. They respectively focus on different levels such as functional classification, metabolic networks, and domain recognition. When used, they are like jigsaw puzzles, filling in the blank Spaces of physiological processes with individual genes. Like KEGG pathway mapping, it can help people identify the node genes involved in key reactions, while the structural analysis of Pfam can reveal the conserved modules in proteins (Sohail et al., 2025). Interestingly, many times such comparisons will unearth some unique "module combinations" of extremist microorganisms, which are often functionally related to energy metabolism, nutrient utilization or environmental response.

4.3 Identification of special functional genes (salt tolerance, thermotolerance, heavy metal resistance, etc.)

Not every extremist microorganism has "dramatic" stress resistance genes in its genome, but as long as it can survive in high-salt or high-temperature environments, those genes responsible for resistance functions are mostly not far away. Heat shock proteins, compatible solute synthases, metal transport pumps... These names may not look new, but once they appear in a certain strain in the form of family expansion or unique combinations, they are worth taking a second look. Comparative analysis can also reveal in which bacteria they are "standard" and in which they are "introduced from outside", which is crucial for understanding adaptation strategies (Srivastava et al., 2017; Wang et al., 2025). Sometimes, a metal-resistant protein is not just for "survival"; it may also become a "candidate star" in later biotechnology.

5 Comparative Genomic Analysis

5.1 Whole-genome alignment with related extremophilic bacteria

Sometimes, identifying the "special features" of an extremist microorganism doesn't require it to speak for itself - a genomic comparison with its "relatives" can reveal where it has retained its traditions and where it has embarked

on a unique evolutionary path. Especially with the high-quality genomes pieced together by hybrid sequencing, structural changes such as insertions, deletions, and even gene rearrangements become particularly clear. For instance, the consensus pan-chromosome assembly method of *Acinetobacter baumannii* has revealed "flexible regions" related to environmental adaptability. Such strategies are also applicable to the study of microorganisms living in extreme environments (Chan et al., 2015; Gould and Henderson, 2023). Often, it is those easily overlooked genomic islands and resistance elements that are truly the key to explaining adaptability.

5.2 Identification of core and variable genomic regions

The requirements for bacteria in extreme environments are not merely to "survive", but to "survive and live well". So, in their genomes, apart from the core areas that maintain basic life activities, there are also many "mobile" modules hidden. These variable regions are not present in every strain; they are more like "additional configurations" tailored for certain ecological conditions. Polar bacilli such as Haemophilus erythrosalis exhibit an open pan-genome - the stable core region is maintained by homologous recombination, while the dynamic helper parts are strongly influenced by horizontal gene transfer, including many fragments from plasmids or gene islands (Figure 1) (González-Torres and Gabaldon, 2018). These "mobile components" may not be used every day, but once the environment changes, they become the trump card for survival.

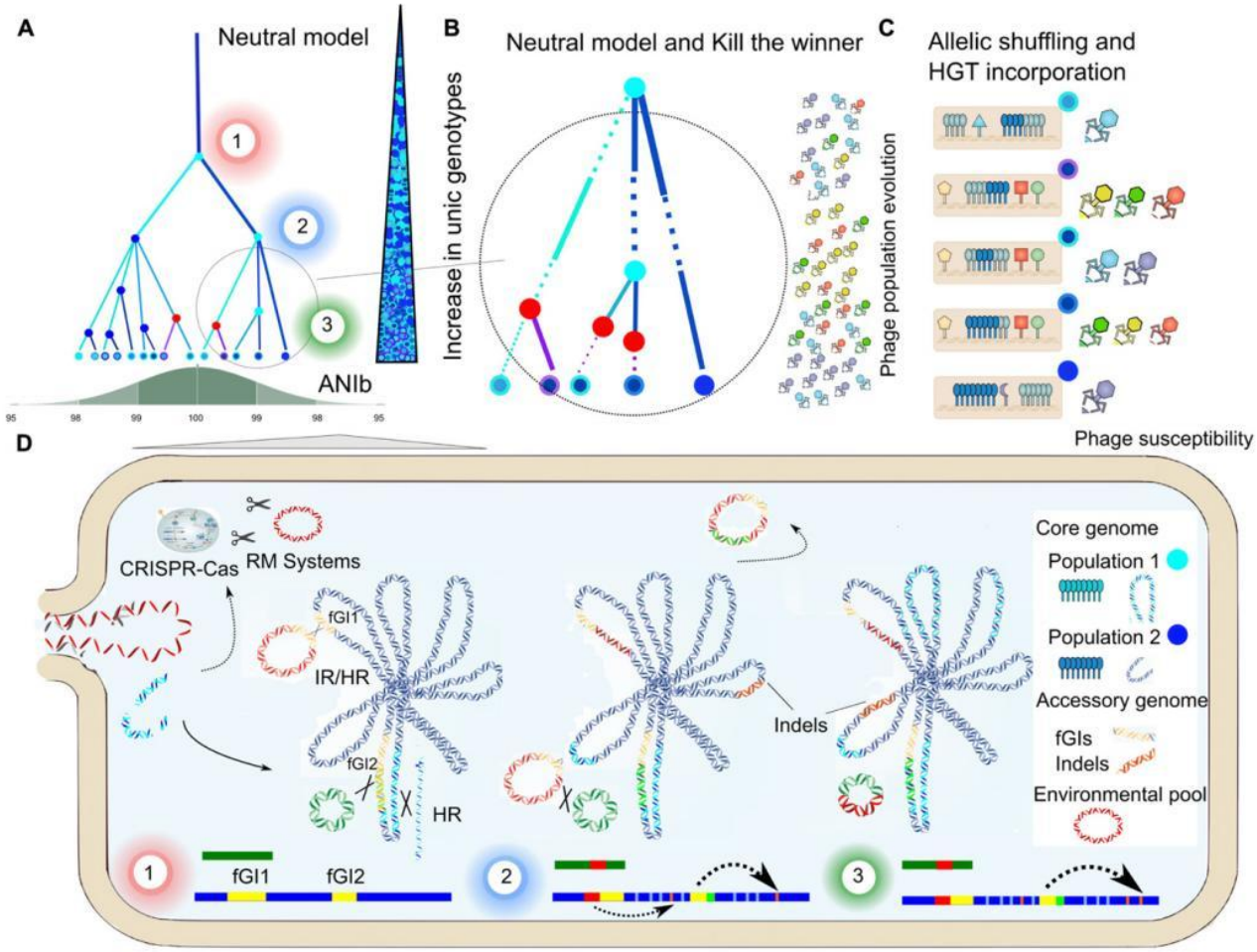

Figure 1 Genome dynamics and ecological models (Adopted from González-Torres and Gabaldón, 2018)

5.3 Gene family expansion, contraction, and species-specific gene analysis

Not all genetic families remain unchanged during evolution; some grow stronger while others gradually fade away. For instance, those gene groups related to heat resistance, metal resistance or osmotic pressure regulation often become "expansion households" in extremophiles. However, in contrast, some less necessary pathways will also experience functional contraction. This contractional-expansion is not random but closely related to the selective pressure of the environment. By comparing the genomes of different species, some species-specific metabolic

pathways and resistance mechanisms can be unearthed, which are all evidence of their "solo" operation in the ecological niche (Gonzalez-Torres and Gabaldon, 2018). In other words, the changes in gene families have written their "evolutionary diaries" of adapting to the environment.

6 Phylogenetic and Evolutionary Analysis

6.1 Phylogenetic tree construction based on 16S rRNA and single-copy core genes

To determine which branch the newly discovered extreme thermophilic bacteria belong to in the systematic classification, relying solely on traditional markers such as 16S rRNA is often not detailed enough. However, as an "old tool" for bacterial classification, it has not been phased out, especially when combined with multiple single-copy core genes for analysis, it can fill many resolution gaps. The phylogenetic relationships revealed in this way can not only identify kinship but are also often linked to the ecological adaptability of the bacterial species. The new strain forms a stable branch with other thermophilic bacteria that live in similar extreme environments. It is worth noting that some "traces" left by horizontal gene transfer can still be faintly seen in this clustering - they might be the driving forces behind the diversification of metabolic capabilities (Cooper et al., 2022; Shen et al., 2024).

6.2 Synteny analysis and detection of genome rearrangement events

Sometimes, when phylogenetic trees fail to tell us something, genomic structure can make up for it. In the collinearity comparison of this strain with other extremophiles, the gene arrangement in some regions was highly consistent, but many regions were disrupted, such as inversions, insertions, and even large-scale translocations. This kind of rearrangement is not merely about "different arrangements"; in many cases, it actually reflects a shift in evolutionary strategies. For instance, genomic islands and movable elements often cluster in these rearrangement regions. Many studies suspect that they are related to the flexibility of the genome - especially when facing environmental stress, they may be the "accelerators" of rapid evolution (Li et al., 2020; Neubert et al., 2021).

6.3 Adaptive evolution and identification of positively selected genes

Genome-wide screening sometimes reveals some "restless" genes that are particularly active in metabolism, stress response or perception of environmental changes, and often bear traces of positive selection. For instance, some genes related to heavy metal resistance or osmotic pressure tolerance are significantly more prone to mutation than other conserved regions, which precisely explains why these microorganisms can thrive in extreme environments. Furthermore, some genes are not inherited vertically but are "borrowed" from other microorganisms. That is to say, horizontal transfer is also involved in the rapid process of functional innovation. Overall, vertical genetics, horizontal transfer and natural selection interweave like three forces, jointly shaping the evolutionary profile of these extreme microorganisms (Safari et al., 2025).

7 Case Studies: Applications of Extremophile Genomes

7.1 Industrial enzyme development from hot spring or salt lake strains

Some microorganisms from hot springs or salt lakes, which originally live in extreme environments, have enzymes in their bodies that possess extraordinary "resistance" properties: they can withstand high temperatures, endure strong salt, and remain calm even at extreme pH values. This characteristic precisely meets some industrial demands, such as high-demand scenarios like biofuels, food processing and pharmaceuticals. In fact, many heat-resistant proteases, lipases and amylases that are currently in use were initially discovered through genomic analysis. Especially after the quality of genome assembly has improved, researchers can not only locate the genes of these enzymes, but also identify the regulatory regions that co-occur with them and even complete expression modules. Interestingly, some microorganisms from soda lakes have surprisingly demonstrated carbohydrate metabolism pathways adapted to high-salt conditions, and even metagenomic data can reveal their catalytic potential (Verma et al., 2022; Mangoma et al., 2024).

7.2 Engineering microorganisms using stress-resistance genes from extreme environments

Not all microorganisms can survive in environments with high salt or heavy metals. Those that can withstand them mostly carry specific anti-stress genes in their bodies. Such genes are now increasingly being "borrowed" to

modify industrial bacteria. It's not that traditional strains are bad, but when facing extreme process conditions, they are too "delicate". After the emergence of tools like CRISPR/Cas9, transplanting these stress-resistant genes is no longer a difficult task. For instance, the attempt on *Halomonas*, a halophilic bacterium, is quite typical: through gene editing, it performed better under non-sterile and highly saline conditions, with a significant increase in product yield. This approach of transferring the "talent" of extreme bacteria to engineered bacteria is precisely a breakthrough point in the progress of microbial application (Figure 2) (Qin et al., 2018).

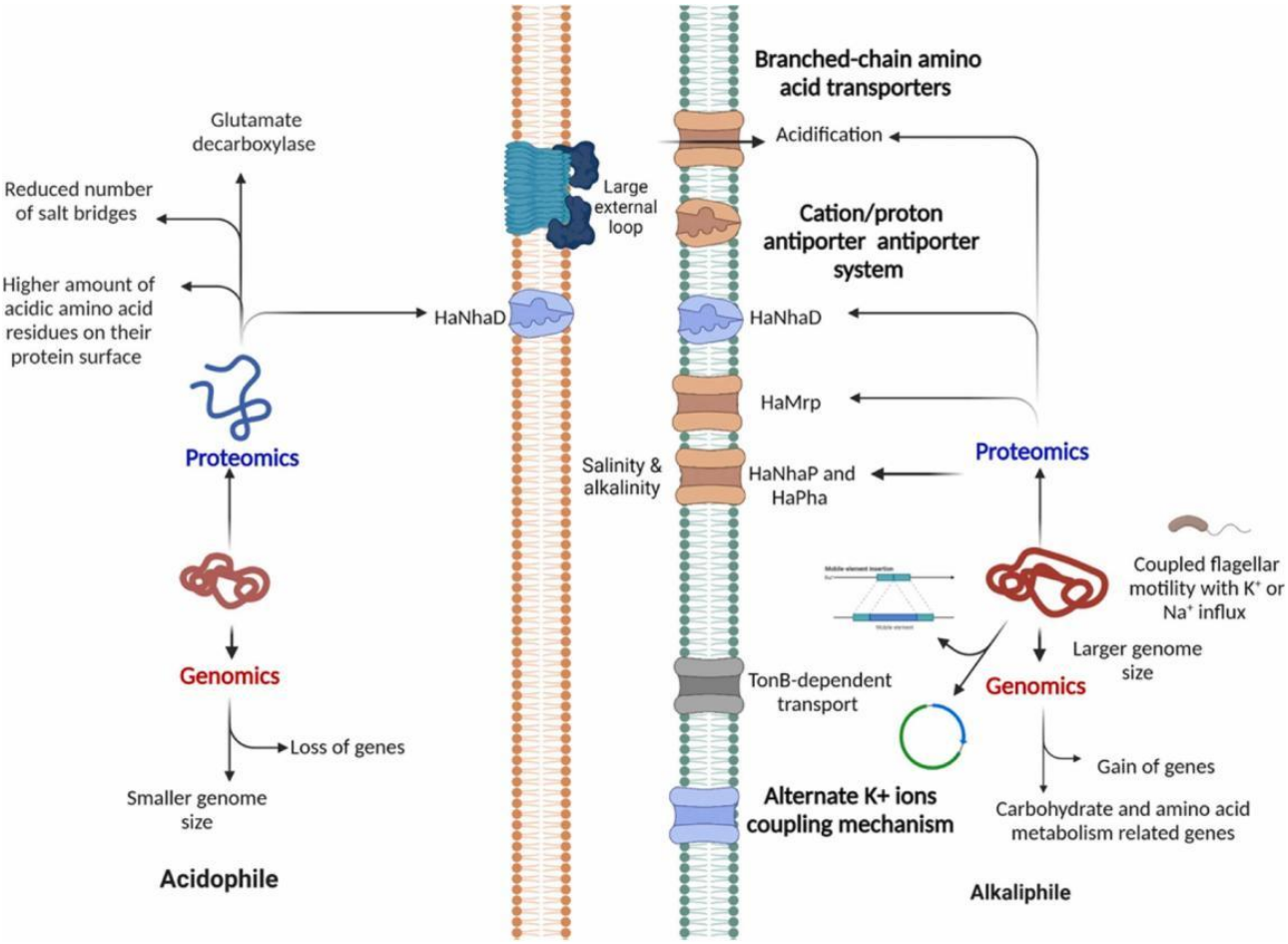

Figure 2 Genomic adaptations of acidophiles and alkaliphiles (Adopted from Salwan and Sharma, 2022)

7.3 Strategies for bioremediation and synthetic biology platforms in harsh environments

When it comes to "problem areas" like acidic mine drainage and saline-alkali land, ordinary microorganisms simply can't handle them. However, some polar bacteria can survive quite well in these environments. Their genomes are like an instruction manual, telling us how these bacteria deal with heavy metals and toxic substances. Often, it is the new genes or adaptive mutations that have been horizontally transferred from the genome that are at work. Researchers are attempting to unearth this adaptability and transform it into materials for building "customized" biological platforms. The purpose of designing these platforms is not only to degrade pollutants, but also to utilize them to synthesize some special and economically valuable substances. The logic is actually quite simple: By integrating the solutions provided by nature into synthetic biology systems in an eclectic manner, it is possible to obtain both sustainable and efficient processing or production pathways (Salwan and Sharma, 2022).

8 Conclusion and Future Perspectives

When assembling the genomes of extremist microorganisms, what is often first exposed is the set of "physiological codes" that enable them to survive in harsh environments - such as resilience genes, peculiar metabolic pathways, and specialized tools for DNA repair. However, such discoveries are not isolated. After comparison, many core functions are indeed very conservative, but there are always some gene modules that seem to be "borrowed" through horizontal gene transfer mixed in. This combination precisely constructs the way they

adapt to specific ecological niches. Phylogenetic relationships may be far apart, but convergent evolution still brings them back together - perhaps because these extreme environments exert similar survival pressures (which is quite evident in homology analysis). So, these pieced together genomes are not merely data; they serve as a window to understand their potential functions and application values.

Of course, not all problems can be solved by "competing on genes". Often, no matter how clear the measurement is, it still gets stuck when facing a genome with high GC content, complex structure or dense repetitive sequences. Although existing sequencing and assembly methods are making progress, they still fail to capture certain key variations or low-abundance regions adequately. Some structural information that affects adaptability may leave no trace at all. Moreover, the number of functional genes verified through experiments is too small, making subsequent annotation and interpretation difficult. To break through these blind spots, in addition to the continuous optimization of the hybrid sequencing strategy, the "caliber" of the reference database must also keep up, so as to truly restore the complex and complete genetic blueprint of these microorganisms.

If the early genome is a "map", then the next step undoubtedly depends on multiple omics working together. Just having maps is not enough. Real-time dynamic data - such as transcriptional responses, protein expression, and metabolic flow maps - must all be superimposed to see how they respond to environmental stimuli and maintain their homeostasis. The relationships among different "groups" of data are complex and difficult to sort out by intuition. At this time, algorithms such as machine learning become particularly important. They can identify those key variables in the chaos, such as potential biomarkers or certain specific adaptation mechanisms. This entire set of integrated measures is not merely about helping scientific research "see clearly", but also paves the way for the future design of engineered strains with specific functions, and makes the application of extremist microorganisms in industrial or synthetic biology fields more practical and feasible.

Acknowledgments

We are grateful to colleagues for their critically reading the manuscript and providing valuable feedback that improved the clarity of the text. We also thank to the two anonymous peer reviewers for their revision suggestions on this study.

Conflict of Interest Disclosure

The authors affirm that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Arias P., Butler J., Randhawa G., Soltysiak M., Hill K., and Kari L., 2023, Environment and taxonomy shape the genomic signature of prokaryotic extremophiles, Scientific Reports, 13(1): 16105.
https://doi.org/10.1101/2023.05.24.542097

Chan A., Sutton G., DePew J., Krishnakumar R., Choi Y., Huang X., Beck E., Harkins D., Kim M., Lesho E., Nikolich M., and Fouts D., 2015, A novel method of consensus pan-chromosome assembly and large-scale comparative analysis reveal the highly flexible pan-genome of *Acinetobacter baumannii*, Genome Biology, 16(1): 143.
https://doi.org/10.1186/s13059-015-0701-6

Chen Z., Erickson D., and Meng J., 2020, Benchmarking hybrid assembly approaches for genomic analyses of bacterial pathogens using Illumina and Oxford Nanopore sequencing, BMC Genomics, 21(1): 631.
https://doi.org/10.1186/s12864-020-07041-8

Chklovski A., Parks D., Woodcroft B., and Tyson G., 2022, CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning, Nature Methods, 20(8): 1203-1212.
https://doi.org/10.1038/s41592-023-01940-w

Cooper Z., Rapp J., Shoemaker A., Anderson R., Zhong Z., and Deming J., 2022, Evolutionary divergence of *Marinobacter* strains in cryopeg brines as revealed by pangenomics, Frontiers in Microbiology, 13: 879116.
https://doi.org/10.3389/fmicb.2022.879116

De Almeida F., De Campos T., and Pappas G., 2023, Scalable and versatile container-based pipelines for de novo genome assembly and bacterial annotation, F1000Research, 12: 1205.
https://doi.org/10.12688/f1000research.139488.1

De Maio N., Shaw L., Hubbard A., George S., Sanderson N., Swann J., Wick R., AbuOun M., Stubberfield E., Hoosdally S., Crook D., Peto T., Sheppard A., Bailey M., Read D., Anjum M., Walker A., and Stoesser N., 2019, Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes, Microbial Genomics, 5(9): e000294.
https://doi.org/10.1099/mgen.0.000294

Dong Z.Y., 2024, Impact of Bt applications on soil microbial communities, Bt Research, 15(6): 276-283.
http://dx.doi.org/10.5376/bt.2024.15.0028

Goldstein S., Beka L., Graf J., and Klassen J., 2018, Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing, BMC Genomics, 20(1): 23.
https://doi.org/10.1186/s12864-018-5381-7

Gomez S., Sic W., Haridas S., Labutti K., Eichenberger J., Kaur N., Lipzen A., Barry K., Goodwin S., Gribskov M., and Grigoriev I., 2024, Comparative genomics of the extremophile *Cryomyces antarcticus* and other psychrophilic Dothideomycetes, Frontiers in Fungal Biology, 5: 1418145.
https://doi.org/10.3389/ffunb.2024.1418145

González-Torres P., and Gabaldón T., 2018, Genome variation in the model halophilic bacterium *Salinibacter ruber*, Frontiers in Microbiology, 9: 1499.
https://doi.org/10.3389/fmicb.2018.01499

Gould A., and Henderson J., 2023, Comparative genomics of symbiotic *Photobacterium* using highly contiguous genome assemblies from long read sequences, Microbial Genomics, 9(12): 001161.
https://doi.org/10.1099/mgen.0.001161

Li L., Liu Z., Zhang M., Meng D., Liu X., Wang P., Li X., Jiang Z., Zhong S., Jiang C., and Yin H., 2020, Insights into the metabolism and evolution of the genus *Acidiphilium*, a typical acidophile in acid mine drainage, mSystems, 5(6): 10.1128/msystems. 00867-20.
https://doi.org/10.1128/msystems.00867-20

Mangoma N., Zhou N., and Ncube T., 2024, Metagenome-assembled genomes provide insight into the microbial taxonomy and ecology of the Buhera soda pans, Zimbabwe, PLoS ONE, 19(12): e0299620.
https://doi.org/10.1371/journal.pone.0299620

Manni M., Berkeley M., Seppey M., and Zdobnov E., 2021, BUSCO: assessing genomic data quality and beyond, Current Protocols, 1(12): e323.
https://doi.org/10.1002/cpz1.323

Neal-McKinney J., Liu K., Lock C., Wu W., and Hu J., 2021, Comparison of MiSeq, MinION, and hybrid genome sequencing for analysis of *Campylobacter jejuni*, Scientific Reports, 11(1): 5676.
https://doi.org/10.1038/s41598-021-84956-6

Neubert K., Zuchantke E., Leidenfrost R., Wuenschiers R., Grützke J., Malorny B., Brendebach H., Dahouk A., Homeier T., Hotzel H., Reinert K., Tomaso H., and Busch A., 2021, Testing assembly strategies of *Francisella tularensis* genomes to infer an evolutionary conservation analysis of genomic structures, BMC Genomics, 22(1): 822.
https://doi.org/10.1186/s12864-021-08115-x

Olagoke O., Aziz A., Zhu L., Read T., and Dean D., 2025, Whole-genome automated assembly pipeline for *Chlamydia trachomatis* strains from reference, *in vitro* and clinical samples using the integrated CtGAP pipeline, NAR Genomics and Bioinformatics, 7(1): lqae187.
https://doi.org/10.1093/nargab/lqae187

Qin Q., Ling C., Zhao Y., Yang T., Yin J., Guo Y., and Chen G., 2018, CRISPR/Cas9 editing genome of extremophile *Halomonas* spp., Metabolic Engineering, 47: 219-229.
https://doi.org/10.1016/j.ymben.2018.03.018

Safari M., Butler J., Randhawa G., Hill K., and Kari L., 2025, Life at the extremes: maximally divergent microbes with similar genomic signatures linked to extreme environments, bioRxiv, 4: 657665.
https://doi.org/10.1101/2025.06.04.657665

Salwan R., and Sharma V., 2022, Genomics of prokaryotic extremophiles to unfold the mystery of survival in extreme environments, Microbiological Research, 264: 127156.
https://doi.org/10.1016/j.micres.2022.127156

Shen L., Liu Y., Chen L., Lei T., Ren P., Ji M., Song W., Lin H., Su W., Wang S., Rooman M., and Pucci F., 2024, Genomic basis of environmental adaptation in the widespread poly-extremophilic *Exiguobacterium* group, The ISME Journal, 18(1): wrad020.
https://doi.org/10.1093/ismejo/wrad020

Sohail H., Naveed M., Aziz T., Mohamed R., and Al-Joufi F., 2025, Whole-genome analysis of *Bacillus subtilis* MBBL2 genomic characterization and comparative genomics, Functional and Integrative Genomics, 25(1): 1-13.
https://doi.org/10.1007/s10142-025-01684-0

Srivastava R., Patel V., Sharma A., Srivastava A., Srivastava A., and Saxena A., 2017, De novo assembly, functional annotation and comparative alignment of whole genome of a halo-tolerant *Exiguobacterium profundum* PHM11 with related genomes, Canadian Journal of Biotechnology, 1(Special): 124.
https://doi.org/10.24870/cjb.2017-a110

Verma D., Joshi S., Ghimire P., Mishra A., and Kumar V., 2024, Developments in extremophilic bacterial genomics: a post next generation sequencing era, Ecological Genetics and Genomics, 32: 100255.
https://doi.org/10.1016/j.egg.2024.100255

Verma D., Kumar V., and Satyanarayana T., 2022, Genomic attributes of thermophilic and hyperthermophilic bacteria and archaea, World Journal of Microbiology and Biotechnology, 38(8): 135.
https://doi.org/10.1007/s11274-022-03327-z

Wang Y., Zhao M., Wang Z., Luo X., Wang C., and Guo B., 2025, Whole genome sequence data of *Comamonas sediminis* FS4_11, a Fumonisin B1-Transforming Bacterium, Using Hybrid Nanopore-Illumina Sequencing, Data in Brief, 2025: 111829.
https://doi.org/10.1016/j.dib.2025.111829

Wick R., Judd L., Gorrie C., and Holt K., 2017, Completing bacterial genome assemblies with multiplex MinION sequencing, Microbial Genomics, 3(10): e000132.
https://doi.org/10.1099/mgen.0.000132

Zhang P., Jiang D., Wang Y., Yao X., Luo Y., and Yang Z., 2021a, Comparison of de novo assembly strategies for bacterial genomes, International Journal of Molecular Sciences, 22(14): 7668.
https://doi.org/10.3390/ijms22147668

Zhang Z., Liu G., Chen Y., Xue W., Ji Q., Xu Q., Zhang H., Fan G., Huang H., Jiang L., and Chen J., 2021b, Comparison of different sequencing strategies for assembling chromosome-level genomes of extremophiles with variable GC content, iScience, 24(3): 102219.
https://doi.org/10.1016/j.isci.2021.102219