

Article

Not peer-reviewed version

Genomic Foundation for the Ecological Dominance of Cosmopolitan *Microcoleus vaginatus* and *Microcystis aeruginosa*

[Jingyi Wei](#), [Hua Li](#)^{*}, Xiaoyu Guo, [Yunzhu Wang](#), [Chunxiang Hu](#)^{*}

Posted Date: 21 April 2026

doi: 10.20944/preprints202604.1526.v1

Keywords: biocrusts; cyanobacteria; comparative genomics; drylands; ecological dominance; freshwater bloom; life-strategy trade-offs; *Microcoleus vaginatus*; *Microcystis aeruginosa*



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Genomic Foundation for the Ecological Dominance of Cosmopolitan *Microcoleus vaginatus* and *Microcystis aeruginosa*

Jingyi Wei ^{1,2}, Hua Li ^{1,*}, Xiaoyu Guo ³, Yunzhu Wang ^{1,2} and Chunxiang Hu ^{1,*}

¹ CAS Key Laboratory of Algal Biology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China

² College of Advanced Agricultural Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

³ Henan Vocational College of Agriculture, Zhengzhou 451450, China

* Correspondence: lih@ihb.ac.cn (H.L.) and cxhu@ihb.ac.cn (C.H.); Tel.: +86-27-68780046.

Abstract

Cyanobacteria dominate ecosystems ranging from oligotrophic deserts to eutrophic lakes, yet it remains unclear whether distantly related species thrive in disparate habitats through shared genomic foundations or divergent specialization. Here, we address this question using *Microcoleus vaginatus*, the pioneer stabilizer of biocrusts, and *Microcystis aeruginosa*, the agent of freshwater blooms worldwide, as contrasting models of terrestrial and aquatic dominance. We assembled a comparative framework of 504 high-quality cyanobacterial genomes, including 132 *M. vaginatus*, 148 *M. aeruginosa*, and 224 reference taxa, and jointly analyzed genome architecture, functional repertoires, and genomic plasticity. Despite phylogenetic separation, both species share high rates of horizontal gene transfer and retain a compact, conserved functional core centered on FAD-dependent oxidoreductases, manganese efflux, and class II aldolases that collectively maintain redox balance, photosynthetic performance, and metabolic robustness. Nevertheless, the two lineages followed contrasting genomic strategies that *M. vaginatus* expands regulatory breadth and stress-resilience gene families, whereas *M. aeruginosa* shows genome streamlining and rapid exploitation. Notably, aquatic *M. vaginatus* strains retain terrestrial genomic scaffolds while gradually rewiring plasticity mechanisms and niche-specific functions. Together, these results reveal a two-tier architecture of cyanobacterial dominance, a conserved survival core coupled with divergent adaptive peripheries. It offers a predictive framework for how cyanobacterial lineages will respond to the global-change pressures.

Keywords: biocrusts; cyanobacteria; comparative genomics; drylands; ecological dominance; freshwater bloom; life-strategy trade-offs; *Microcoleus vaginatus*; *Microcystis aeruginosa*

1. Introduction

Cyanobacteria are among the oldest oxygenic photoautotrophs on Earth and occupy nearly every light-bearing habitat, from open oceans and freshwater systems to desert soils and polar rocks [1–3]. Through oxygenic photosynthesis and biogeochemical cycling of C, N, and S, they continue to shape the atmosphere and the modern biosphere [4,5]. Despite this pervasive success, the genomic logic that allows distantly related cyanobacteria to achieve ecological dominance in strikingly different niches remains an unresolved question in microbial evolution.

Nonetheless, *Microcoleus vaginatus* and *Microcystis aeruginosa* provide an ideal contrast for addressing this question (hereafter referred to as *M.v* and *M.a*, respectively). *M.v* is the canonical pioneer and stabilizer of biological soil crusts (known as Earth's living skin [6]), playing a crucial role in underpinning soil multifunctionality and dryland restoration [7,8]. In contrast, *M.a* is a globally

invasive, harmful-bloom-forming species that periodically overwhelms eutrophic freshwaters and coastal transition zones, impairing water quality, the integrity of the aquatic food web, and public health [9–12]. The two species are separated by more than 1 billion years of phylogenetic divergence and occupy habitats with opposing light, water, and nutrient regimes [13]. However, both achieve monopolistic dominance within their realms. Whether this parallel ecological triumph reflects shared genomic foundations, lineage-specific specialization, or a layered combination of the two mechanisms is not known [14,15].

The ecological success of cyanobacteria is ultimately encoded at the genomic level, where architecture, gene content, and plasticity jointly determine metabolic flexibility, stress tolerance, and competitive capacity [3,16,17]. Transitions between terrestrial and aquatic lifestyles, in particular, typically demand substantial genomic reorganization, ranging from streamlining for metabolic efficiency in nutrient-poor waters to genome expansion to cope with environmental heterogeneity on land [18,19]. These life-history contrasts are sculpted by positive selection, homologous recombination, and gene gain/loss, which together shape pan-genome plasticity and niche specialization [20–22].

Previous work has identified several traits that plausibly contribute to single-niche dominance, including EPS-mediated filament bundling, surface motility, specialized signal transduction, and phycosphere-based C/N exchange in *M.v* [23–25], and colony formation, gas-vesicle buoyancy, toxin production, antiviral systems, and efficient nutrient storage in *M.a* [2,26,27]. However, these studies have remained habitat-bounded, precluding a systematic test of which features are lineage-specific adaptations and which belong to a broader cyanobacterial dominance toolkit. Moreover, although genome streamlining is frequently invoked for aquatic specialists [28,29], many terrestrial cyanobacteria retain the capacity to thrive in liquid culture, and aquatic *M.v* strains are increasingly reported from transitional habitats. It remains unclear whether short-term terrestrial-to-aquatic transitions reshape genomic plasticity and functional repertoires while retaining ancestral terrestrial traits, and whether such transitions pass through identifiable genomic intermediates.

To address these gaps, we assembled a comparative framework comprising 504 high-quality cyanobacterial genomes, including both terrestrial and aquatic ecotypes of bundle-forming *M.v* ($n = 132$) and bloom-forming *M.a* ($n = 148$), along with 224 additional cyanobacterial reference genomes. By integrating analyses of genome architecture, functional gene repertoires, genomic plasticity, and lineage-specific molecular evolution, complemented by metabolic modeling, we aim to resolve (i) how genome architecture underpins ecological dominance across terrestrial and aquatic habitats, (ii) how distinct plasticity mechanisms (HGT, recombination, gene gain/loss) orchestrate lineage-specific genetic innovation, and (iii) how evolutionary processes jointly sustain a conserved functional core while driving divergent adaptive peripheries, including the genomic signatures of aquatic *M.v* as potential terrestrial-to-aquatic intermediates. Together, this study seeks to move cyanobacterial comparative genomics beyond single-habitat descriptions toward a unified, mechanism-based account of cross-habitat dominance and a predictive basis for how these lineages will respond to global-change pressures as desertification and eutrophication.

2. Materials and Methods

2.1. Data Collection and Quality Control

We constructed a comprehensive dataset comprising 280 high-quality genomes, including 132 *M.v* genomes representing both terrestrial and aquatic ecotypes, and 148 *M.a* genomes from aquatic environments. Of these, we previously sequenced 57 *M.v* genomes, and the rest were downloaded from the NCBI database as of March 2024 (**Table S1**). The genome of strain *M.a* CS-567/02-A1 was filtered out due to its low similarity to other *M.a* genomes. We assessed each genome using *CheckM* v1.0.18 to meet stringent quality standards, with $\geq 95\%$ completeness and $\leq 5\%$ contamination [30,31]. In addition, we retrieved 245 cyanobacterial genomes marked as reference in NCBI, of which 224 strains with $> 95\%$ completeness were selected to provide evolutionary context for our comparative

analysis (**Table S1**). To ensure the evolutionary coherence of each focal species, we verified that both *M.v* and *M.a* populations maintain robust intra-species gene flow (**Figure S1**). The homoplasy-to-nonhomoplasy ratio (h/m) across increasing numbers of genome samples was calculated using *ConSpeciFix* v1.3.0 to quantify the extent of gene flow [32].

2.2. Genome Annotation and Architectural Characterization

All retained genomes were uniformly annotated using the *Prokka* v1.12 pipeline [33]. Genome size, GC content, and gene density were calculated using a custom *R* script. The GC content at the third codon position (GC3s) and the codon adaptation index (CAI) were determined using *CodonW* v1.3 (*Peden*, <http://codonw.sourceforge.net/>). Before the formal analysis of genomic datasets, all ribosomal protein genes annotated by *Prokka* were extracted for a preliminary run to generate the *cai.coa* file required for subsequent CAI computation. We predicted tRNA genes using *tRNAscan-SE* v2.0.5 [34]. Repetitive sequences and small RNAs were identified via *RepeatMasker* v4.0.8 (<http://www.repeatmasker.org>) and *Infernal* v1.1.5 against the *Rfam* database [35], respectively. For functional classification, all protein-coding sequences were annotated against the *EggNOG* database using *eggNOG-mapper* v2.1.8 to assign clusters of orthologous groups (COG) categories [36]. Genome-scale metabolic models were reconstructed using *CarveMe* [37], and specific metabolic pathways were further extracted through *COBRAPy* [38].

2.3. Genomic Plasticity and Molecular Evolution Analysis

Orthologous group (OG) clustering was performed using *OrthoFinder* v2.2.7 [39], which employs the diamond algorithm for sequence alignment. Based on the result file *Orthogroups.GeneCount.csv*, which contains gene counts per sample for each OG, we performed 300 permutations for each OG to calculate pan- and core-summary statistics. The pan-genome size was fitted to Heaps' law, and core genome decay was modeled using an exponential decay function [40].

Putative horizontal gene transfer (HGT) regions were predicted using *Alien Hunter*, an application that calculates HGT boundaries using interpolated variable-order motifs [41]. Insertion sequences (IS) and prophage regions were identified by *ISEScan* v1.7.3 and *Phispy* v3.7.8 [42,43]. For defense systems, the restriction-modification (RM) system was screened against *REBASE* using *Diamond* v2.0.8.146 [44,45], and we used a custom *Python* script to identify complete RM systems. Type I systems required all three core subunits (RE, MT, S); Type II and III systems required both RE and MT; Type IIG and Type IV systems were counted directly as functional units. The total number of complete RM systems was summed across all types. *CRISPRCasFinder* v4.2.18 was used to find the CRISPR-Cas genetic architecture in the genomes [46]. We curated the natural competence gene set based on previous studies about natural transformation in Gram-negative bacteria and filamentous cyanobacteria [47,48].

The protein sequences of core OGs present across all strains in each group were aligned using *MAFFT* v7.453 with the parameter `--localpair --maxiterate = 1000` [49], and back-translated to codon alignments based on corresponding nucleotide sequences. We used *KaKs_Calculator* v2.0 to calculate the pairwise non-synonymous substitution rate (*Ka*), synonymous substitution rate (*Ks*), and *Ka/Ks* for core homologous genes using the *YN* algorithm [50]. The substitution rates of paralogous genes were calculated by *wgd* v1 [51].

2.4. Niche-Specific Functional Divergence and Evolution

A custom *R* script was used to identify conserved and niche-specific elements. A count file of orthologous genes or metabolic reactions was used as input, and the data were binarized into presence-absence matrices. Conserved elements were those with $\geq 95\%$ prevalence within each target group. Elements present in $> 20\%$ of the reference group were defined as ubiquitous and removed. Venn diagram was employed to distinguish niche-specific elements and shared elements.

Subsequently, niche-specific and shared OGs were selected as target OGs for comprehensive functional and evolutionary characterization.

For function, we extracted COG annotations of all niche-specific and shared OGs. The top 12 dominant COG terms were retained, and low-frequency categories were combined into the Other category. Reciprocal searches and domain profiling via *eggNOG-mapper* result files were used to distinguish conserved OGs from their potential isozymes. Protein motifs were analyzed using *MEME* suite v5.5.9 [52]. For evolution, corresponding gene coordinates were retrieved from *GFF* files annotated by *Prokka*, and 10-kb flanking regions of target genes were generated using *bedtools* v2.27.1 [53]. A trusted HGT event was defined by genes overlapping with HGT regions identified by *Alien Hunter* and accompanied by either Phispy-identified prophages or IS elements in the 10-kb flanking sequences. Nucleotide sequences of genes from each specific/shared OG were extracted, and gene pairs within each group were analyzed by *KaKs_Calculator* v2.0 to obtain median *Ka/Ks*. For shared OG, median *Ka/Ks* values were evaluated within each group and averaged.

2.5. Statistical Analysis and Visualization

All visualizations were generated using *R* v4.2.2, which included the *ggplot2*, *ggridges*, *gghalves*, *VennDiagram*, and *corrplot* packages. Correlations between genomic features and evolutionary indices were calculated using Spearman's rank correlation. Statistical significance between groups was assessed using the Wilcoxon rank-sum test.

3. Results

3.1. Divergence in Genomic Architecture and Functional Repertoires

Comparative genomic analysis revealed pronounced architectural divergence between the two species, while the *M.v* ecotypes exhibited minimal differentiation (**Figure 1A, 1B**). Both *M.v* ecotypes maintained similar genome sizes (terrestrial strains: 7.36 ± 0.64 Mb, aquatic strains: 7.24 ± 0.53 Mb; Wilcoxon test, $p > 0.05$), significantly larger than the genomes of *M.a* (4.86 ± 0.40 Mb, $p < 0.001$). Despite this, *M.v* exhibits lower gene density (850 genes/Mb) than *M.a* (949 genes/Mb), suggesting a more complex genomic architecture of *M.v* that accommodates additional regulatory elements or mobile genetic content. Furthermore, *M.v* tends to have higher repeat proportions than *M.a*, with the terrestrial strains of *M.v* peaking at the highest repeat ratios, and the aquatic strains showing similarly elevated repeat ratios (**Figure S2**). The total sRNA hit distribution of *M.v* showed a parallel trend, with aquatic strains at higher counts, followed by terrestrial strains, whereas *M.a* was skewed toward lower counts (**Figure S2**).

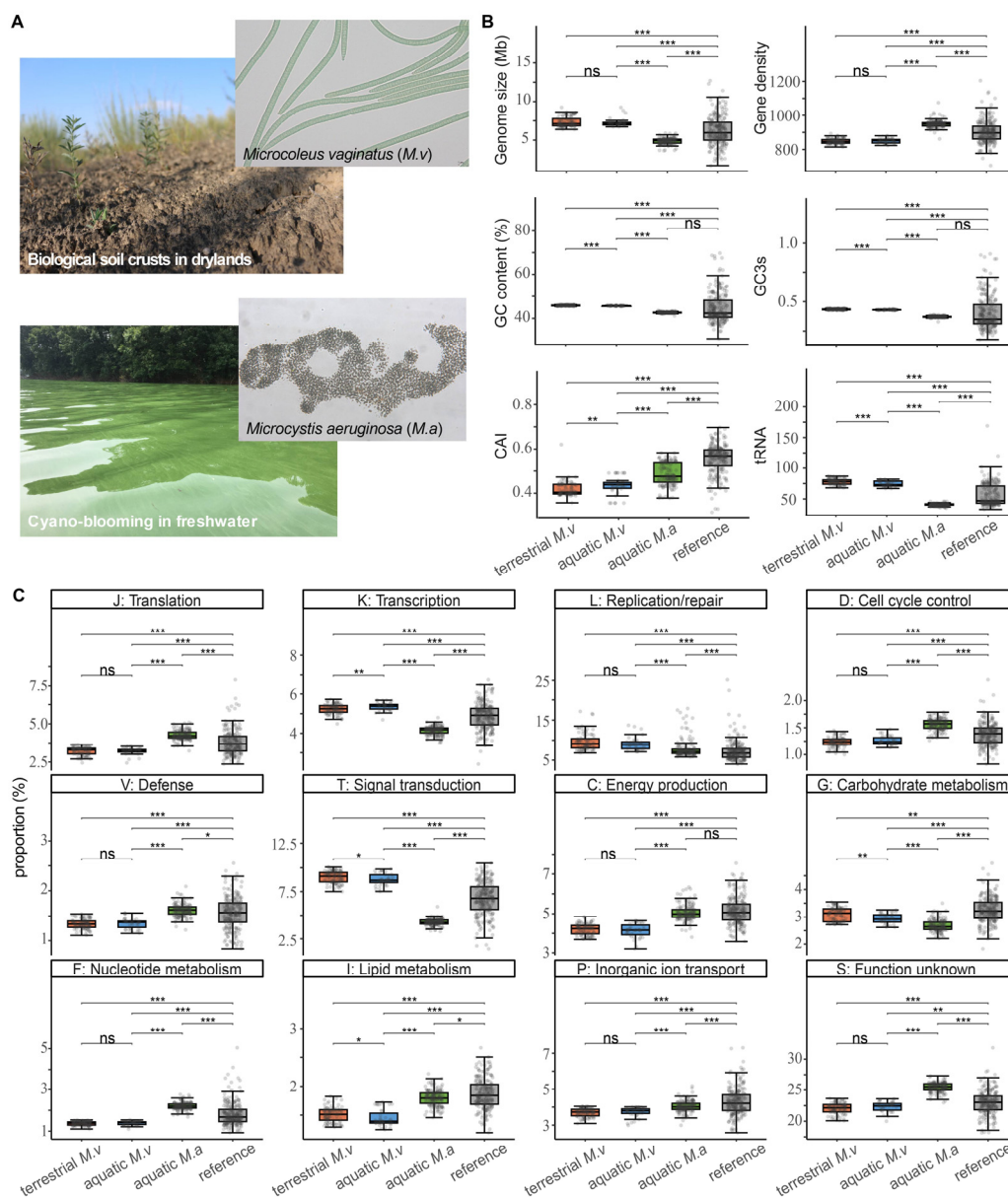


Figure 1. Divergence in genomic architecture and functional gene allocation among focal cyanobacterial lineages. **(A)** Representative photographs of the typical habitats and microscopic morphologies of the focal cyanobacterial species: *Microcoleus vaginatus* (*M.v*) inhabiting biological soil crusts in drylands, and *Microcystis aeruginosa* (*M.a*) forming cyano-blooms in freshwater ecosystems. **(B)** Comparison of genomic features among terrestrial *M.v* ecotypes, aquatic *M.v* ecotypes, *M.a*, and reference cyanobacterial genomes. **(C)** Proportions of genes assigned to specific COG functional categories across the four groups. In the boxplots, the center line represents the median, the box limits indicate the upper and lower quartiles, and whiskers extend to 1.5 times the interquartile range. Only COG categories showing significant divergence in over half of all pairwise comparisons were retained for visualization. Statistical significance was assessed using the Wilcoxon rank-sum test (ns, not significant; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).

GC content patterns were consistent with genome size (**Figure 1B**). The terrestrial and aquatic *M.v* ecotypes displayed significantly higher GC content ($45.83 \pm 0.20\%$ and $45.67 \pm 0.11\%$, respectively) than *M.a* ($42.64 \pm 0.28\%$, $p < 0.001$). This disparity was particularly prominent at GC3s, where *M.v* maintained consistently elevated levels (~ 0.43), contrasting sharply with *M.a* (0.37 ± 0.01). For translation metrics, *M.v* exhibited a lower CAI than *M.a* ($p < 0.001$). Moreover, *M.v* ecotypes

harbored elevated tRNA numbers (median: 76), far exceeding *M.a* (median: 41) and the upper quartile of the reference genomes.

Functional annotation revealed distinct patterns of gene allocation (**Figure 1C**). *M.v* genomes were enriched in transcriptional regulation (K), signal transduction (T), and carbohydrate metabolism (G). In contrast, *M.a* displayed a pronounced bias toward translation (J), cell cycle control (D), energy production and conversion (C), inorganic ion transport (P), and defense-related functions (V). Both *M.v* (terrestrial and aquatic strains) and *M.a* lineages maintained significantly higher proportions of L-class genes (replication, recombination, and repair) than reference genomes. Correlation analysis demonstrated that the relative abundance of L-class genes was positively associated with the proportion of IS elements ($p < 0.001$, **Figure S3**). Genome-scale metabolic model reconstruction further delineated distinct metabolic boundaries (**Table S2**). *M.v* lineages possessed metabolic features such as biotin biosynthesis, nitrite reductases [NAD(P)H], amide hydrolysis, and specific stress response signaling coupled with lipid and peptidoglycan remodeling capabilities. Given the aquatic transition, aquatic *M.v* strains additionally possessed pathways for organic nitrogen assimilation and aromatic metabolism, including urea hydrolysis and tryptamine synthesis. In contrast, *M.a* possessed high-affinity potassium transporter, methionine salvage cycles, histidine biosynthesis, energy storage pathways, reactive nitrogen detoxification, and osmoadaptation mechanisms.

3.2. Genomic Plasticity and Defense System Trade-Offs

Pan-genome accumulation curves revealed that all three lineages maintain open pan-genome structures (**Figure S4**), as indicated by power-law parameters (γ) ranging from 0.237 to 0.269, reflecting ongoing horizontal gene acquisition and loss. They maintained substantially smaller core genomes relative to pan-genome sizes, with *M.a* exhibiting the most streamlined core genome, reflecting intensive adaptation to stable aquatic conditions. The open pan-genome structures necessitate corresponding genomic plasticity mechanisms to integrate incoming genetic material, a context in which the two focal species have evolved distinct genome management strategies. The proportion of putative horizontal gene transfer (HGT) regions was markedly higher in *M.v* and *M.a* than in reference genomes (**Figure 2A**), indicating pervasive horizontal acquisition. While the genome fraction occupied by IS remained broadly comparable across the three lineages, the composition-level diversity of IS differed, as reflected in IS family counts (**Figure S5**). In contrast, significant enrichment of prophages was observed in the *M.a* genomes.

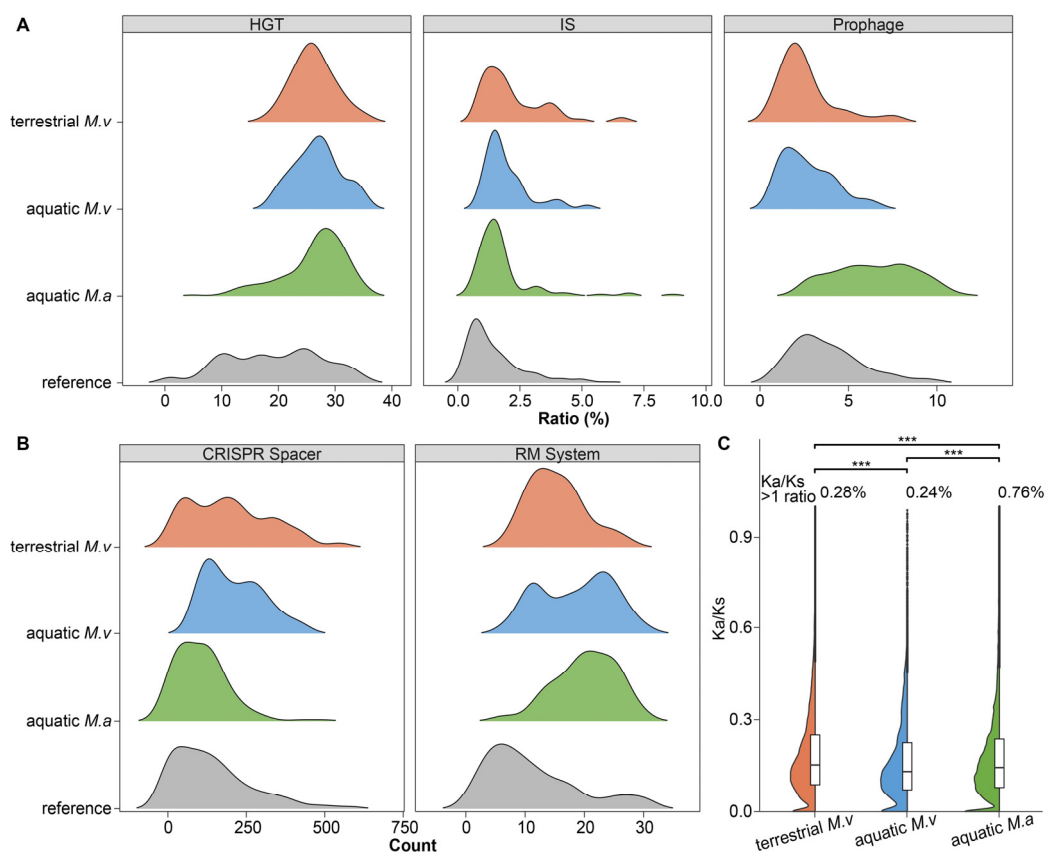


Figure 2. Mechanisms of genomic plasticity and molecular evolutionary signatures. (A) Ridgeline plots illustrating the distribution of the genomic ratio (%) occupied by putative HGT regions, IS, and prophages across terrestrial *M.v* ecotypes, aquatic *M.v* ecotypes, *M.a*, and reference genomes. (B) Distribution of CRISPR spacer counts and complete RM system counts, reflecting the diversity and investment in defense systems across the studied lineages. (C) Violin plot showing the distribution of pairwise non-synonymous to synonymous substitution rate ratios (Ka/Ks) for core OGs in terrestrial *M.v*, aquatic *M.v*, and aquatic *M.a*. The center line within the boxes indicates the median values. The percentage of gene pairs under positive selection ($Ka/Ks > 1$) is noted above each distribution. Asterisks *** indicate significant differences ($p < 0.001$) between groups based on the Wilcoxon rank-sum test.

To further assess DNA uptake as an additional route supporting HGT, we screened genomes for a competence-associated gene set (Table S3). Core components of the pilus/DNA uptake machinery were widely conserved across lineages, indicating broad retention of the basic transformation apparatus. Notably, the minor pilin *fimT*, the pilus assembly factor *pilO*, and the membrane-associated DNA-binding receptor *comEA* were generally present in *M.v* but nearly absent from *M.a*. In contrast, the pilin *pilX* was nearly restricted to *M.a*, but undetectable in *M.v*; and the secretin *pilQ* was absent in *M.v* but present in 32% of *M.a*. These genes were present in part of the reference genomes, ranging from 19% to 75%.

Defense systems modulate HGT rates and constitute crucial regulatory factors of genomic plasticity. *M.v* displayed heterogeneous CRISPR spacer distributions yet minimal investment in RM systems. Detailed system analysis revealed that Type II RM systems were most abundant across all three lineages. However, *M.a* showed significant enrichment for complete Type I and Type II G systems, whereas *M.v* maintained a greater number of Type IV systems (Figure S6). Both aquatic lineages showed higher restriction enzyme abundance than terrestrial *M.v* ($p < 0.01$). Notably, aquatic *M.v* occupied intermediate positions across these metrics. The retention of elevated repetitive content, combined with increasing prophage integration, suggests that aquatic *M.v* could be progressively remodeling its genome to match aquatic selective pressures while retaining terrestrial adaptive features.

3.3. Niche-specific Functional Divergence and Adaptation

Nucleotide-substitution rate analysis of core orthologs further supported the evolution of ecological adaptation. Median K_s values declined progressively from terrestrial *M.v* (0.120) through aquatic *M.v* (0.103) to *M.a* (0.083) (**Figure S7A**). While K_a/K_s ratios indicated predominant purifying selection (**Figure 2C**), *M.v* maintained significantly higher median K_a/K_s values. Paralogous pairs showed lower K_s but higher proportions under positive selection in *M.v* ecotypes than in *M.a* and the references (**Figures S7B, S7C**).

To identify niche-specific adaptation in functional repertoires, we screened OG clusters that were either unique to or shared among lineages. After filtering OGs present in > 20% of reference genomes to exclude universally conserved functions, *M.a* possessed 343 exclusive OGs, substantially exceeding those of individual *M.v* ecotypes (**Figure 3A**). By contrast, both terrestrial and aquatic ecotypes of *M.v* shared 584 OGs that form the stable adaptive foundation for *Microcoleus* ecological success across terrestrial and aquatic systems. Functional distribution analysis revealed that the massive shared gene pool of *M.v* showed primary investment in signal transduction (T) and cell wall/membrane biogenesis (M) (**Figure 3B**). Exclusive OGs of *M.a* were predominantly unknown or unmapped. A gene annotated as the cell envelope-related transcriptional attenuator was shared exclusively by aquatic *M.v* and *M.a*, despite being detected in most terrestrial *M.v* genomes ($n = 81$) with a lower than 95% conserved threshold (**Table S4**). Three OGs shared across all lineages further illustrate the fundamental conserved requirements for the ecological dominance of cyanobacteria, including a FAD-dependent oxidoreductase, a manganese efflux pump (*mntP*), and a class II aldolase with an adducin N-terminal domain.

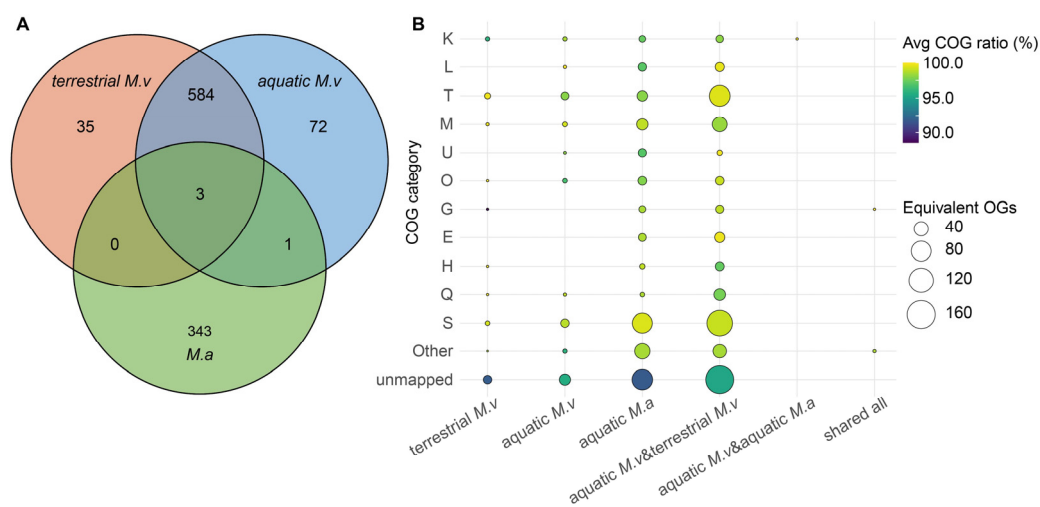


Figure 3. Distribution patterns and functional characterization of niche-specific and shared OGs. **(A)** Venn diagram displaying the number of conserved OGs (present in $\geq 95\%$ of strains within a group) that are exclusive to or shared among terrestrial *M.v*, aquatic *M.v*, and *M.a*. OGs present in > 20% of the reference genomes were filtered out to highlight non-ubiquitous function. **(B)** Bubble plot mapping the COG functional category distribution (Y-axis) across unique and shared OG subsets (X-axis). Bubble size represents the number of equivalent OGs, and the color gradient indicates the average COG category ratio (%) within each OG. The columns 'terrestrial *M.v*', 'aquatic *M.v*', and 'aquatic *M.a*' show OGs unique to each specific ecotype or species, driving their respective niche specializations. The column 'aquatic *M.v* & terrestrial *M.v*' reveals the stable adaptive foundation and conserved core genes of the *Microcoleus vaginatus* across different habitats. The column 'aquatic *M.v* & aquatic *M.a*' represents OGs shared exclusively by the two aquatic lineages, reflecting potential convergent adaptations to aquatic environments. The column 'all-shared' identifies the three universal OGs maintained across all focal lineages, representing the fundamental requirements for ecological dominance in both terrestrial and aquatic ecosystems.

Given the possibility that distinct gene families may annotate to identical functions, we performed reciprocal searches on these conserved functional OGs and identified potential isozyme gene families (Table S4). Multiple OGs encoding FAD-dependent oxidoreductases are widely conserved across cyanobacteria. Domain profiling revealed that the homolog conserved in the *M.v* and *M.a* lineages possesses the FAD_binding_3 and Trp_halogenase domains (Figure 4). Another OG characterized by these domains was only present in two *M.a* strains. For the manganese efflux pump, additional OGs identified by the search were detected in only a limited subset of reference genomes. Regarding the class II aldolase, one additional OG was found to be conserved in *M.v* but absent in *M.a*, while another OG, missing in both lineages, was present in some reference genomes. All three aldolase OGs harbor the aldolase II catalytic domain. However, motif analysis revealed substantial structural variation. OG0002425 comprises a compact, conserved catalytic core of relatively small size. While preserving similar catalytic segments, OG0002653 incorporates an additional alkaline-enriched motif8 alongside an aromatic and hydrophobic motif1, and OG0010356 is characterized by a significant C-terminal extension (Figure S8).

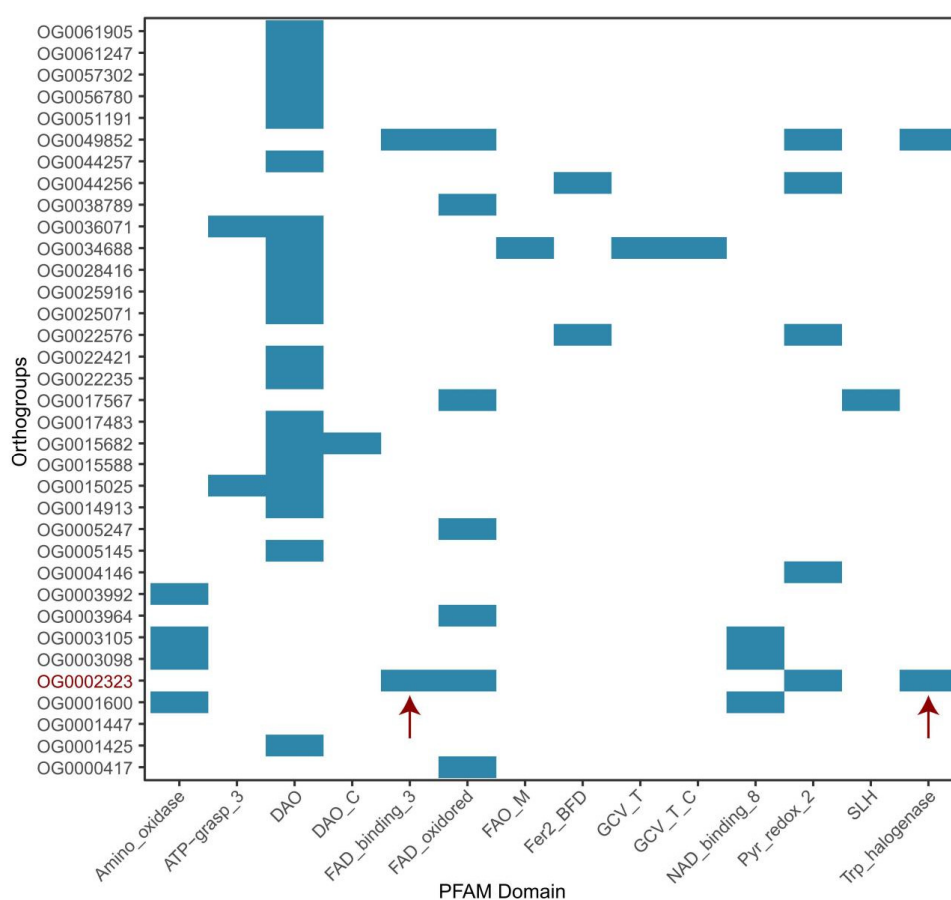


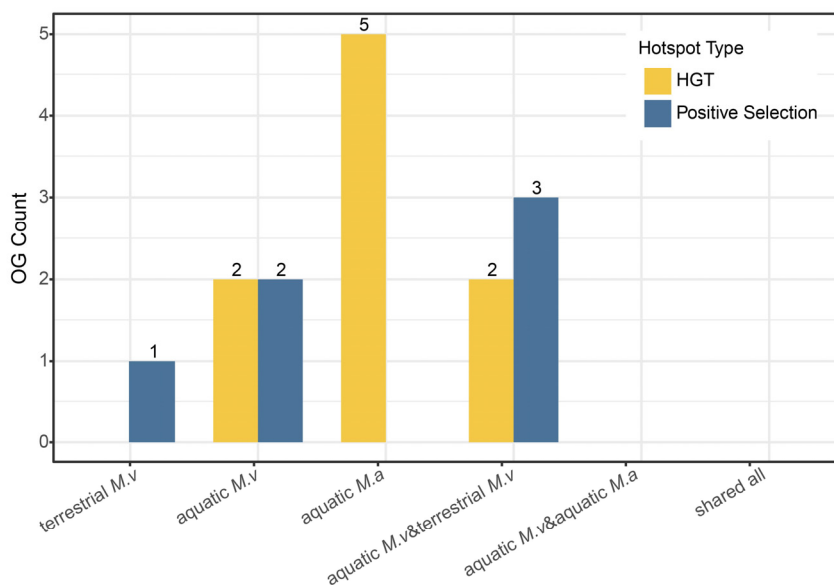
Figure 4. Comparative domain profiling of FAD-dependent oxidoreductase OGs. The heatmap illustrates the distribution of PFAM domains (X-axis) across various OGs (Y-axis) annotated as FAD-dependent oxidoreductases. Blue rectangles indicate the presence of a specific domain within an OG. The OG highlighted in red (OG0002323) represents the universally conserved core identified across all focal lineages. Red arrows highlight the unique domain architecture of OG0002323, which harbors both the FAD_binding_3 and Trp_halogenase domains, distinguishing it from other potential isozymes that lack this functional combination.

3.4. Evolutionary Drivers of Niche-Specific Genes

To further elucidate niche-adaptive strategies, we analyzed the evolutionary drivers of niche-specific gene families, including HGT, gene duplication, and selection pressure. Correlation analysis revealed significant synergistic relationships among genomic processes (Figure S9). The positive

correlation between HGT, multicopy expansion, and K_s values ($p < 0.001$) indicated that niche-specific genes persist through dosage-driven stabilization and subsequent functional refinement rather than representing transient adaptive responses. The strong positive correlation between median K_a/K_s and the $K_a/K_s > 1$ ratio ($r = 0.57$, $p < 0.001$) validated median K_a/K_s as a reliable indicator of adaptive evolutionary intensity.

Hotspot screening based on HGT ratio (> 0.5) or $K_a/K_s > 1$ revealed distinct evolutionary trajectories, with genes bearing minimal reference genome counts emerging as high-priority targets for unique niche specialization (**Figure 5**). *M.v* adaptation centered primarily on the positive selection of lineage-specific genes. A striking example is OG0006890, which encodes arginyl-tRNA synthetase (*argS*) and was detected in aquatic *M.v* with a reference count of only 1, yet exhibits extreme positive selection (K_a/K_s ratio of 47.56%). The IS605 OrfB family transposase (Ref = 3), shared between both *M.v* ecotypes, also shows strong positive selection. Beyond these positive selection hotspots, horizontal acquisition further expanded the *M.v* adaptive repertoire. Terrestrial and aquatic *M.v* maintain horizontally acquired elements crucial for stress response, including a gene encoding a bacterial stress protein and *grpE*, which participates in the response to hyperosmotic and heat shock by preventing the aggregation of stress-denatured proteins. Alternatively, aquatic *M.a* relied predominantly on HGT-mediated acquisition of specialized metabolic modules rather than lineage-specific positive selection. Notably, both aquatic lineages independently acquired distinct orthologous groups annotated as clan AA aspartic protease via horizontal transfer. HGT hotspot genes showed substantially higher multicopy representation and broader distribution across reference genomes than positive selection hotspots, reflecting two complementary adaptive strategies.



OG	Hotspot type	Ref	Group	Ka/Ks	Ka/Ks >1 ratio	HGT ratio	Multicopy ratio	Catetogy	Description
OG0000131	HGT	38	aquatic <i>M.a</i>	0.14	0.10	0.64	0.99	-	-
OG0000151	HGT	25	aquatic <i>M.a</i>	0.12	0.15	0.61	0.99	S	-
OG0000325	HGT	44	aquatic <i>M.a</i>	0.14	0.90	0.52	0.95	O	clan AA aspartic protease
OG0003789	HGT	37	aquatic <i>M.a</i>	0.13	0.84	0.56	0.06	C	acyl-CoA dehydrogenase
OG0004059	HGT	35	aquatic <i>M.a</i>	0.26	0.16	0.63	0.00	S	-
OG0002770	HGT	4	aquatic <i>M.v</i>	0.35	3.59	0.54	0.74	-	-
OG0003494	HGT	33	aquatic <i>M.v</i>	0.39	15.11	0.64	0.47	O	clan AA aspartic protease
OG0000296	HGT	43	aquatic <i>M.v</i> & terrestrial <i>M.v</i>	0.16	0.07	0.64	1.00	T	bacterial stress protein
OG0005404	HGT	7	aquatic <i>M.v</i> & terrestrial <i>M.v</i>	0.30	0.23	0.54	0.00	O	grpE, response to hyperosmotic and heat shock
OG0005665	Positive selection	7	aquatic <i>M.v</i>	5.40	10.12	0.00	0.00	-	-
OG0006890	Positive selection	1	aquatic <i>M.v</i>	1.07	47.56	0.01	0.06	J	argS, arginyl-tRNA aminoacylation
OG0005774	Positive selection	3	terrestrial <i>M.v</i>	1.17	66.16	0.06	0.00	-	-
OG0005366	Positive selection	4	aquatic <i>M.v</i> & terrestrial <i>M.v</i>	1.01	45.96	0.01	0.03	-	-
OG0005570	Positive selection	4	aquatic <i>M.v</i> & terrestrial <i>M.v</i>	1.19	46.25	0.02	0.00	-	-
OG0005631	Positive selection	3	aquatic <i>M.v</i> & terrestrial <i>M.v</i>	1.10	57.71	0.15	0.00	L	transposase, IS605 OrfB family

Figure 5. Evolutionary hotspots driving niche-specific adaptation. The top bar chart shows the number of evolutionary hotspot OGs identified across different lineages and their intersections. Hotspots are categorized into two types: HGT hotspots (yellow), defined by an HGT ratio > 0.5, and positive selection hotspots (blue), characterized by a median $Ka/Ks > 1$. The bottom table shows detailed evolutionary and functional metrics for representative hotspot OGs.

4. Discussion

By comparing the genomes of *M.v* and *M.a*, this study identified a set of conserved keystone mechanisms that underlie the evolution of distinct genomic properties to achieve ecological dominance. Although both *M.v* and *M.a* are increasingly recognized as species complexes encompassing substantial cryptic diversity [7,54], the main goal of this study is not to clarify their taxonomic boundaries. Rather, we focus on the genomic basis driving the ecological success of this important and dominant group of cyanobacteria.

4.1. Genomic Plasticity as a Prerequisite for Ecological Dominance

Our findings suggest that the ecological dominance of both lineages relies on high genomic plasticity buffered by robust maintenance systems. *M.v* and *M.a* maintain open pan-genomes and exhibit HGT ratios well above those of reference strains, indicating that continuous acquisition of exogenous DNA is a shared prerequisite for occupying diverse niches. Repair-associated genes are

significantly enriched across all focal lineages. They are positively correlated with IS elements, supporting the idea that the mutational and structural burden imposed by active gene flow must be counter-balanced by enhanced genome-integrity investment [55,56]. Despite shared plasticity and a repair imperative, the routes by which the two lineages acquire and manage exogenous DNA diverge in ways that match their habitats. *M.v* retains a comparatively complete natural competence toolkit suited for active DNA scavenging in biocrusts (**Table S3**). The universal presence of the minor pilin *fimT*, whose arginine-rich patches facilitate DNA binding to the type IV pilus [57], together with the inner-membrane assembly factor *pilO* and the periplasmic receptor *comEA*, suggests a high-efficiency DNA uptake and translocation platform [48]. This constitutive competence likely enables *M.v* to exploit extracellular DNA pools under nutrient limitation and frequent physicochemical stress in soils.

In contrast, *M.a* lacks these core competence receptors, indicating reduced reliance on direct transformation in more stable aquatic settings. Instead, *M.a* uniquely retains *pilX*, a factor promoting pilin polymerization that is often linked to biofilm and colony formation [58], suggesting that *M.a* has repurposed its pilus machinery to support its colonial lifestyle. The high prophage content, which provides superinfection exclusion and auxiliary metabolic genes [59,60], further supports a shift toward phage-mediated transduction as a major HGT route in *M.a*. Although *M.a* harbors diverse CRISPR-Cas systems [27], it shows reduced spacer representation but elevated investment in restriction-modification systems relative to *M.v*, suggesting distinct defense trade-offs that shape HGT permissiveness. Heterogeneous soil habitats may favor long-term immune memory against intermittent threats, whereas bloom-forming aquatic populations may benefit more from constitutive barrier defenses and prophage-mediated exclusion under sustained viral pressure [61,62].

4.2. Universal retention Requirements Across Ecological Boundaries

Beyond the divergent plasticity strategies observed between terrestrial- and aquatic-dominant cyanobacteria, we identify a minimal conserved core that likely constrains ecological success across habitats. The conservation of a FAD-dependent oxidoreductase is consistent with a shared requirement to maintain redox homeostasis in oxygenic phototrophs, where photosynthetic electron transport inevitably generates reactive oxygen species under fluctuating light and nutrient conditions [63]. Likewise, the *mntP* regulates the tolerance to oxidative stress and photosystem II function through manganese homeostasis [64,65]. It protects cyanobacterial cells from intracellular toxicity under excess conditions [66], preventing oxidative damage and ensuring adequate manganese cofactors for efficient water splitting. Furthermore, the universal presence of a class II aldolase OG suggests the selection for maintaining robust carbon-flux capacity through central metabolism. Although the function of the N-terminal domain of adducin in cyanobacteria has not been fully characterized, its participation in cytoskeleton assembly in eukaryotes implies a potential role on the cell wall/membrane binding [67], population adhesion, or maintenance of cell morphology, contributing to their bundle or colony forming lifestyle [23,26]. The conservation of these specific OGs over their isozymes appears to be driven by structural efficiency and functional indispensability. For instance, while multiple aldolase isozymes exist, OG0002425 was selected as the conserved core likely due to its compact catalytic structure, whereas its isozymes possess C-terminal extensions or additional motifs that may impose higher metabolic costs due to niche-specific functions not required for general dominance [68]. Meanwhile, the conserved FAD-dependent oxidoreductase is characterized by its Trp_halogenase domain. Tryptophan halogenase catalyzes the first step in the biosynthesis of pyrrolnitrin, an antibiotic with broad-spectrum anti-fungal activity [69].

4.3. Divergent Investment Upon Universal Chassis

Upon this conserved physiological foundation, *M.v* and *M.a* exhibit profoundly divergent genomic architectures and functional investment patterns, reflecting two distinct evolutionary routes for resource allocation. *M.v* invests in architectural complexity and regulatory plasticity, manifested through larger and repeat-rich genomes, lower gene density with an emphasis on signal transduction

and transcriptional regulation, and higher GC content, which grant the metabolic flexibility required to rapidly respond to unpredictable soil micro-environmental pulses and stress [70]. It reflects adaptive responses to frequent desiccation-rehydration cycles and to high irradiance in biological soil crusts [15,24]. Genome-scale metabolic model reconstruction indicates that *M.v* lineages primarily allocate metabolic capacity to cofactor autonomy and redox/structural resilience, including biotin biosynthesis, NAD(P)H-dependent nitrite reduction, amide hydrolysis, and stress-response signaling coupled with lipid and peptidoglycan remodeling, consistent with selection for rapid physiological reconfiguration under brief nutrient pulses and recurrent envelope damage in soils (**Table S2**). Despite the high metabolic cost of an additional gene and elevated GC content [43,46], the expanded tRNA pool, coupled with a refined strategy of codon usage bias, decouples translation from stringent optimization and facilitates rapid cell growth [71,72]. Evolutionary analysis indicates that *M.v* drives adaptation primarily through intense positive selection on rare, lineage-specific genes and paralogous subfunctionalization [48], effectively fine-tuning the intrinsic cellular machinery to withstand localized stressors. Although tRNA synthetases form core components of the translation machinery, whose functional mutations are typically lethal, a recent study showed that mutation of the PheS aminoacyl tRNA synthetase increases bacterial tolerance to disinfectants [73]. Positive selection on the IS605 OrfB family transposase also establishes a direct link between genomic plasticity mechanisms and adaptive fine-tuning, indicating that evolvability itself is evolvable [74].

However, *M.a* embodies a specialized exploiter strategy [75]. Through genome streamlining and translation-biased functional allocation, *M.a* prioritizes replication speed and metabolic efficiency to achieve rapid biomass accumulation in eutrophic water bodies. Metabolism reconstructions highlight *M.a*-specific investments in high-affinity potassium transporters, osmoadaptation, reactive nitrogen detoxification, and energy storage pathways, alongside methionine salvage cycles and histidine biosynthesis, which collectively support fast homeostasis, detoxification, and surge-capture growth in fluctuating ionic and redox conditions of blooms. Moreover, bloom water bodies typically sustain high viral abundance and rapid infection turnover, providing a plausible ecological basis for the elevated representation of defense-related functions in *M.a* [62]. Rather than incremental fine-tuning, *M.a* predominantly relies on the horizontal acquisition of specialized metabolic modules followed by multicopy expansion, consistent with its high diversity in the accessory genome [76]. The coupling between HGT and multicopy has also been observed in *Staphylococcus* [77], representing dosage-amplified deployment to enable the rapid scaling of specific pathways to exploit sudden resource availability or mount immediate competitive defenses.

4.4. Aquatic-Specific Adaptation and Gradual Ecological Transition

The transition from terrestrial to aquatic habitats highlights both convergent evolution and the persistence of ancestral traits. We observed a case of potential convergent retention in which a cell envelope-related transcriptional attenuator is shared exclusively by all aquatic focal lineages. It is also present in most terrestrial *M.v* genomes, but at lower levels of conservation (**Table S4**). The homolog has been reported to catalyze the final step in cell wall teichoic acid biosynthesis [78], reflecting common selective pressures on the cell envelope integrity of aquatic microorganisms. Furthermore, both aquatic lineages independently acquired distinct clan AA aspartic proteases via HGT, which are believed to contribute to adhesion and invasion of host tissues by degrading cell-surface structures [79].

Importantly, the genomic profile of aquatic *M.v* reveals that ecological transition is not an immediate architectural overhaul. Its genome size, GC content, repeat proportion, and broad functional investment remain strongly anchored to the terrestrial state. However, its plasticity-regulatory networks, competence machinery, prophage load, and defense system allocation exhibit distinct intermediate characteristics that progressively shift toward an aquatic phenotype. This pattern aligns with evolutionary discordance among genes and the notion that ecological diversity exceeds evolutionary diversity [80,81]. It suggests a stepwise evolutionary model in which genomic architecture is conserved during initial colonization, serving as a stable, stress-resilient buffer. In

contrast, genomic plasticity mechanisms and niche-specific functional repertoires are progressively rewired to match the viral dynamics and nutrient profiles of the new aquatic habitat. Such gradual remodeling ensures ecological continuity, allowing dominant terrestrial lineages to bridge environmental boundaries without incurring the fitness costs of abrupt genomic contraction. Aquatic *M.v* expanded its substrate spectrum by investing more in urea hydrolysis and tryptamine synthesis. The former is in line with previous studies finding that the *ureC* gene is enriched in riparian zones [82], while the latter is recognized as an endogenous signaling molecule that coordinates the response of plants to environmental stress, which can improve the stress resistance of cyanobacteria by enhancing the antioxidant defense system [83].

5. Conclusions

Taken together, our results advance a unified mechanistic framework for cyanobacterial ecological dominance in which genomic plasticity provides the prerequisite for adaptability, conserved functional cores supply the foundational requirements for phototrophic life, and lineage-specific innovations serve as the competitive differentiator. The conserved mechanisms identified here offer targets for predicting cyanobacterial responses to environmental perturbations, and the distinct HGT and defense strategies illuminate how microbes navigate virus-driven arms races. Future work should prioritize testing the functional significance of lineage-specific genes under field conditions, elucidating the temporal dynamics of genomic transitions during colonization, and extending these principles to other cyanobacterial taxa to assess the generality of the stepwise transition model for microbial ecological success across environmental boundaries.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. The manuscript is accompanied by supplemental materials with 9 Figures and 4 Tables.

Author Contributions: Conceptualization, J.Y.W. and C.X.H.; methodology, J.Y.W.; formal analysis, J.Y.W.; investigation, J.Y.W. and X.Y.G.; writing-original draft preparation, J.Y.W.; writing-review and editing, J.Y.W. and H.L.; visualization, J.Y.W. and H.L.; supervision, H.L. and C.X.H.; funding acquisition, H.L. and C.X.H.

Funding: This work was supported by the National Natural Science Foundation of China (32370125, 32571881, and 32430005), the Natural Science Foundation for Distinguished Young Scholars of Hubei Province (2022CFA105).

Data Availability Statement: The genomes reported in this study are publicly available in the NCBI database as described in Table S1. The scripts supporting the findings in this study are deposited on GitHub (<https://github.com/rosemed/comparative-genomics-between-Microcoleus-vaginatus-Microcystis-aeruginosa>).

Acknowledgments: We are grateful for the technical support provided by the Freshwater Algae Culture Collection at the Institute of Hydrobiology (FACHB) and the Analysis and Testing Center of IHB. The Supercomputing Center of CAS, Wuhan Branch, assists with the sequencing and statistical analyses.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Sanchez-Baracaldo P, Hayes PK, Blank CE. Morphological and habitat evolution in the Cyanobacteria using a compartmentalization approach. Review. *Geobiology*. 2005; 3(3): 145-65. doi:10.1111/j.1472-4669.2005.00050.x
2. Whitton B, Potts M. *The ecology of cyanobacteria. Their diversity in time and space*. Dordrecht: Kluwer Academic Publishers; 2000.
3. Chen MY, Teng WK, Zhao L, et al. Comparative genomics reveals insights into cyanobacterial evolution and habitat adaptation. *ISME J*. 2021; 15(1): 211-27. doi:10.1038/s41396-020-00775-z

4. Garcia-Pichel F, Belnap J, Neuer S, Schanz FJAS. Estimates of global cyanobacterial biomass and its distribution. *Algological Studies*. 2003; 109: 213-27.
5. Zehr J, Bench S, Carter B, et al. Globally Distributed Uncultivated Oceanic N₂-Fixing Cyanobacteria Lack Oxygenic Photosystem II. *Science*. 2008; 322(5904): 1110-2. doi:10.1126/science.1165340
6. Bowker MA, Maestre FT, Eldridge D, et al. Biological soil crusts (biocrusts) as a model system in community, landscape and ecosystem ecology. *Biodivers Conserv*. 2014; 23(7): 1619-37. doi:10.1007/s10531-014-0658-x
7. Stanojkovic A, Skoupy S, Johannesson H, Dvorak P. The global speciation continuum of the cyanobacterium *Microcoleus*. *Nat Commun*. 2024; 15(1): 2122. doi:10.1038/s41467-024-46459-6
8. Li H, Huo D, Wang W, et al. Multifunctionality of biocrusts is positively predicted by network topologies consistent with interspecies facilitation. *Mol Ecol*. 2020; 29(8): 1560-73. doi:10.1111/mec.15424
9. Harke MJ, Steffen MM, Gobler CJ, et al. A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, *Microcystis* spp. *Harmful Algae*. 2016; 54: 4-20. doi:10.1016/j.hal.2015.12.007
10. Huo D, Gan N, Geng R, et al. Cyanobacterial blooms in China: diversity, distribution, and cyanotoxins. *Harmful Algae*. 2021; 109: 102106. doi:10.1016/j.hal.2021.102106
11. Lakshmikandan M, Li M, Pan B. Cyanobacterial Blooms in Environmental Water: Causes and Solutions. *Current Pollution Reports*. 2024; 10(4): 606-27. doi:10.1007/s40726-024-00322-w
12. Tatters A, Howard M, Nagoda C, Busse L, Gellene A, Caron D. Multiple Stressors at the Land-Sea Interface: Cyanotoxins at the Land-Sea Interface in the Southern California Bight. *Toxins*. 2017; 9(3): 95. doi:10.3390/toxins9030095
13. Shih PM, Wu D, Latifi A, et al. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *PNAS*. 2013; 110(3): 1053-8. doi:10.1073/pnas.1217107110
14. Yamamichi M. How does genetic architecture affect eco-evolutionary dynamics? A theoretical perspective. *Phil Trans R Soc B*. 2022; 377(1855): 20200504. doi:10.1098/rstb.2020.0504
15. Murik O, Oren N, Shotland Y, et al. What distinguishes cyanobacteria able to revive after desiccation from those that cannot: the genome aspect. *Environ Microbiol*. 2017; 19(2): 535-50. doi:10.1111/1462-2920.13486
16. Christmas NAM, Anesio AM, Sanchez-Baracaldo P. The future of genomics in polar and alpine cyanobacteria. *FEMS Microbiol Ecol*. 2018; 94(4): doi:10.1093/femsec/fiy032
17. Li C, Liao H, Xu L, et al. The adjustment of life history strategies drives the ecological adaptations of soil microbiota to aridity. *Mol Ecol*. 2022; 31(10): 2920-34. doi:10.1111/mec.16445
18. Muraille E. Diversity Generator Mechanisms Are Essential Components of Biological Systems: The Two Queen Hypothesis. *Front Microbiol*. 2018; 9: 223. doi:10.3389/fmicb.2018.00223
19. Sriswasdi S, Yang CC, Iwasaki W. Generalist species drive microbial dispersion and evolution. *Nat Commun*. 2017; 8(1): 1162. doi:10.1038/s41467-017-01265-1
20. Ellegren H, Galtier N. Determinants of genetic diversity. *Nat Rev Genet*. 2016; 17(7): 422-33. doi:10.1038/nrg.2016.58
21. Chu X, Li S, Wang S, Luo D, Luo H. Gene loss through pseudogenization contributes to the ecological diversification of a generalist *Roseobacter* lineage. *ISME J*. 2021; 15(2): 489-502. doi:10.1038/s41396-020-00790-0
22. Wheatley RM, MacLean RC. CRISPR-Cas systems restrict horizontal gene transfer in *Pseudomonas aeruginosa*. *ISME J*. 2021; 15(5): 1420-33. doi:10.1038/s41396-020-00860-3
23. Garcia-Pichel F, Wojciechowski M. The Evolution of a Capacity to Build Supra-Cellular Ropes Enabled Filamentous Cyanobacteria to Colonize Highly Erodible Substrates. *PLoS One*. 2009; 4(11): e7801. doi:10.1371/journal.pone.0007801
24. Rajeev L, da Rocha UN, Klitgord N, et al. Dynamic cyanobacterial response to hydration and dehydration in a desert biological soil crust. *ISME J*. 2013; 7(11): 2178-91. doi:10.1038/ismej.2013.83
25. Couradeau E, Giraldo-Silva A, De Martini F, Garcia-Pichel F. Spatial segregation of the biological soil crust microbiome around its foundational cyanobacterium, *Microcoleus vaginatus*, and the formation of a nitrogen-fixing cyanosphere. *Microbiome*. 2019; 7: 55. doi:10.1186/s40168-019-0661-2
26. Xiao M, Li M, Reynolds C. Colony formation in the cyanobacterium *Microcystis*. *Biol Rev*. 2018; 93(3): 1399-420. doi:10.1111/brv.12401

27. Yang C, Lin F, Li Q, Li T, Zhao J. Comparative genomics reveals diversified CRISPR-Cas systems of globally distributed *Microcystis aeruginosa*, a freshwater bloom-forming cyanobacterium. *Front Microbiol.* 2015; 6: 394. doi:10.3389/fmicb.2015.00394
28. Jackrel SL, White JD, Evans JT, et al. Genome evolution and host-microbiome shifts correspond with intraspecific niche divergence within harmful algal bloom-forming *Microcystis aeruginosa*. *Mol Biol Evol.* 2019; 28(17): 3994-4011. doi:10.1111/mec.15198
29. Swan BK, Tupper B, Sczyrba A, et al. Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc Natl Acad Sci U S A.* 2013; 110(28): 11463-8. doi:10.1073/pnas.1304246110
30. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 2015; 25(7): 1043-55. doi:10.1101/gr.186072.114
31. Bowers R, Kyrpides N, Stepanauskas R, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol.* 2017; 35(8): 725-31. doi:10.1038/nbt.3893
32. Bobay L, Ellis B, Ochman H. ConSpeciFix: classifying prokaryotic species based on gene flow. *Bioinformatics.* 2018; 34(21): 3738-40. doi:10.1093/bioinformatics/bty400
33. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014; 30(14): 2068-9. doi:10.1093/bioinformatics/btu153
34. Chan Patricia P, Lin Brian Y, Mak Allysia J, Lowe Todd M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* 2021; 49(16): 9077-96. doi:10.1093/nar/gkab688%J Nucleic Acids Research
35. Ontiveros-Palacios N, Cooke E, Nawrocki Eric P, et al. Rfam 15: RNA families database in 2025. *Nucleic Acids Res.* 2024; 53(D1): D258-D67. doi:10.1093/nar/gkae1023%J Nucleic Acids Research
36. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol.* 2021; 38(12): 5825-9. doi:10.1093/molbev/msab293%J Molecular Biology and Evolution
37. Machado D, Andrejev S, Tramontano M, Patil KR. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Research.* 2018; 46(15): 7542-53. doi:10.1093/nar/gky537
38. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRAPy: CONstraints-Based Reconstruction and Analysis for Python. *BMC Syst Biol.* 2013; 7(1): 74. doi:10.1186/1752-0509-7-74
39. Emms D, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019; 20(1): 238. doi:10.1186/s13059-019-1832-y
40. Tettelin H, Riley D, Cattuto C, Medini D. Comparative genomics: the bacterial pan-genome. *Curr Opin Microbiol.* 2008; 11(5): 472-7. doi:10.1016/j.mib.2008.09.006
41. Vernikos GS, Parkhill J. Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics.* 2006; 22(18): 2196-203. doi:10.1093/bioinformatics/btl369
42. Xie Z, Tang H. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics.* 2017; 33(21): 3340-7. doi:10.1093/bioinformatics/btx433
43. Akhter S, Aziz RK, Edwards RA. PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res.* 2012; 40(16): e126. e126. doi:10.1093/nar/gks406
44. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 2015; 12(1): 59-60. doi:10.1038/nmeth.3176
45. Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE: a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* 2022; 51(D1): D629-D30. doi:10.1093/nar/gkac975%J Nucleic Acids Research

46. Couvin D, Bernheim A, Toffano-Nioche C, et al. CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* 2018; 46(W1): W246-W51. doi:10.1093/nar/gky425
47. Nies F, Mielke M, Pochert J, Lamparter T. Natural transformation of the filamentous cyanobacterium *Phormidium lacuna*. *PLoS One.* 2020; 15(6): e0234440. doi:10.1371/journal.pone.0234440
48. Averbhoff B, Kirchner L, Pfefferle K, Yaman D. Natural transformation in Gram-negative bacteria thriving in extreme environments: from genes and genomes to proteins, structures and regulation. *Extremophiles.* 2021; 25(5-6): 425-36. doi:10.1007/s00792-021-01242-z
49. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013; 30(4): 772-80. doi:10.1093/molbev/mst010
50. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics.* 2010; 8(1): 77-80. doi:10.1016/S1672-0229(10)60008-3
51. Chen H, Zwaenepoel A. Inference of Ancient Polyploidy from Genomic Data. In: Van de Peer Y, ed. *Polyploidy: Methods and Protocols.* Springer US; 2023:3-18.
52. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. *Nucleic Acids Res.* 2015; 43(W1): W39-W49. doi:10.1093/nar/gkv416%J Nucleic Acids Research
53. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010; 26(6): 841-2. doi:10.1093/bioinformatics/btq033%J Bioinformatics
54. Perez-Carrascal OM, Terrat Y, Giani A, et al. Coherence of *Microcystis* species revealed through population genomics. *ISME J.* 2019; 13(12): 2887-900. doi:10.1038/s41396-019-0481-1
55. Kim S, Cho C-S, Han K, Lee J. Structural variation of AluElement and human disease. *Genomics & informatics.* 2016; 14(3): 70-7.
56. White M, Allers T. DNA repair in the archaea-an emerging picture. *FEMS Microbiol Rev.* 2018; 42(4): 514-26. doi:10.1093/femsre/fuy020
57. Braus SAG, Short FL, Holz S, Stedman MJM, Gossert AD, Hospenthal MK. The molecular basis of FimT-mediated DNA uptake during bacterial natural transformation. *Nat Commun.* 2022; 13(1): 1065. doi:10.1038/s41467-022-28690-1
58. Hélaïne S, Carbonnelle E, Prouvensier L, Beretti J, Nassif X, Pelicic V. PilX, a pilus-associated protein essential for bacterial aggregation, is a key to pilus-facilitated attachment of *Neisseria meningitidis* to human cells. *Mol Microbiol.* 2005; 55(1): 65-77. doi:https://doi.org/10.1111/j.1365-2958.2004.04372.x
59. Sontheimer E, Davidson A. Inhibition of CRISPR-Cas systems by mobile genetic elements. *Curr Opin Microbiol.* 2017; 37: 120-7. doi:10.1016/j.mib.2017.06.003
60. Middelboe M, Traving S, Castillo D, Kalatzis P, Glud R. Prophage-encoded chitinase gene supports growth of its bacterial host isolated from deep-sea sediments. *ISME J.* 2025; 19(1): wraf004. doi:10.1093/ismejo/wraf004
61. Koonin E, Makarova K. Origins and evolution of CRISPR-Cas systems. *Phil Trans R Soc B.* 2019; 374(1772): 20180087. doi:10.1098/rstb.2018.0087
62. Chen T, Xiong Y, Zhang J, et al. Temporal dynamics, microdiversity, and ecological functions of viral communities during cyanobacterial blooms in Lake Taihu. *NPJ Biofilms Microbiomes.* 2025; 11(1): 178. doi:10.1038/s41522-025-00771-1
63. Trisolini L, Gambacorta N, Gorgoglione R, et al. FAD/NADH Dependent Oxidoreductases: From Different Amino Acid Sequences to Similar Protein Shapes for Playing an Ancient Function. *Journal of Clinical Medicine.* 2019; 8(12): 2117. doi:10.3390/jcm8122117
64. Peng W, Xu Y, Yin Y, et al. Biological characteristics of manganese transporter MntP in *Klebsiella pneumoniae*. *mSphere.* 2024; 9(7): doi:10.1128/msphere.00377-24
65. Eisenhut M. Manganese Homeostasis in Cyanobacteria. *Plants-BASEL.* 2020; 9(1): 18. doi:10.3390/plants9010018
66. Bosma E, Rau M, van Gijtenbeek L, Siedler S. Regulation and distinct physiological roles of manganese in bacteria. *FEMS Microbiol Rev.* 2021; 45(6): fuab028. doi:10.1093/femsre/fuab028

67. Matsuoka Y, Li X, Bennet V. Adducin: structure, function and regulation. *Cell Mol Life Sci.* 2000; 57(6): 884-95. doi:10.1007/PL00000731
68. Held T, Klemmer D, Lässig M. Survival of the simplest in microbial evolution. *Nat Commun.* 2019; 10(1): 2472. doi:10.1038/s41467-019-10413-8
69. Hammer PE, Hill DS, Lam ST, Van Pée KH, Ligon JM. Four genes from *Pseudomonas fluorescens* that encode the biosynthesis of pyrrolnitrin. *Appl Environ Microbiol.* 1997; 63(6): 2147-54. doi:10.1128/aem.63.6.2147-2154.1997
70. Liu Q, Liu HC, Zhou YG, Xin YH. Microevolution and Adaptive Strategy of Psychrophilic Species *Flavobacterium bomense* sp. nov. Isolated From Glaciers. *Front Microbiol.* 2019; 10: 1069. doi:10.3389/fmicb.2019.01069
71. Dong H, Nilsson L, Kurland CG. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J Mol Biol.* 1996; 260(5): 649-63. doi:10.1006/jmbi.1996.0428
72. Weissman JL, Hou S, Fuhrman JA. Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. *PNAS.* 2021; 118(12): e2016810118. doi:10.1073/pnas.2016810118
73. Chen M, Cui R, Hong S, et al. Broad-spectrum tolerance to disinfectant-mediated bacterial killing due to mutation of the PheS aminoacyl tRNA synthetase. *PNAS.* 2025; 122(6): e2412871122. doi:10.1073/pnas.2412871122
74. Wagner A. Adaptive evolvability through direct selection instead of indirect, second-order selection. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution.* 2022; 338(7): 395-404. doi:10.1002/jez.b.23071
75. Morris JJ, Lenski RE, Zinser ER. The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss. Article. *mBio.* 2012; 3(2): e00036-12. doi:10.1128/mBio.00036-12
76. Dick GJ, Duhaime MB, Evans JT, et al. The genetic and ecophysiological diversity of *Microcystis*. *Environ Microbiol.* 2021; 23(12): 7278-313. doi:10.1111/1462-2920.15615
77. Chan C, Beiko R, Ragan M. Lateral transfer of genes and gene fragments in *Staphylococcus* extends beyond mobile elements. *J Bacteriol.* 2011; 193(15): 3964-77. doi:10.1128/JB.01524-10
78. Koksharova O, Popova A, Plyuta V, Khmel I. Four new genes of cyanobacterium *Synechococcus elongatus* PCC 7942 are responsible for sensitivity to 2-Nonanone. *Microorganisms.* 2020; 8(8): 1234. doi:10.3390/microorganisms8081234
79. Monika S, Malgorzata B, Zbigniew O. Contribution of Aspartic Proteases in Candida Virulence. Protease Inhibitors against Candida Infections. Review. *Curr Protein Pept Sci.* 2017; 18(10): 1050-62. doi:10.2174/1389203717666160809155749
80. Sharp PM, Shields DC, Wolfe KH, Li WH. Chromosomal location and evolutionary rate variation in enterobacterial genes. *Science.* 1989; 246(4931): 808-10. doi:10.1126/science.2683084
81. Rubin IN, Ispolatov Y, Doebeli M. Maximal ecological diversity exceeds evolutionary diversity in model ecosystems. *Ecol Lett.* 2023; 26(3): 384-97. doi:10.1111/ele.14156
82. Fisher KA, Yarwood SA, James BR. Soil urease activity and bacterial ureC gene copy numbers: Effect of pH. *Geoderma.* 2017; 285: 1-8. doi:https://doi.org/10.1016/j.geoderma.2016.09.012
83. Khandelwal A, Patel A, Tiwari S, Prasad SM. Tryptamine: a novel signaling molecule alleviating salt-induced toxicity by enhancing antioxidant defense and PSII photochemistry in *Anabaena* PCC7120. *Arch Microbiol.* 2025; 208(1): 64. doi:10.1007/s00203-025-04599-3

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.