Advances in Biodiversity Exploration and Bioactive Metabolite Discovery: Insights from Cyanobacteria, Microalgae, and Uncultured Microorganisms

Rizwan Rashid Bazmi[1], Muhammad Asif[2*], Hafiza Sidra Yaseen[3] , Ashok Gnanasekaran[5*]

Abstract

Microorganisms, including cyanobacteria, microalgae, and uncultured bacteria, constitute a vast and largely unexplored reservoir of biodiversity and bioactive metabolites. Traditional culture-based methods have historically limited access to microbial diversity, capturing less than 1% of species, while leaving “microbial dark matter” untapped. Recent technological advancements in robotic sampling, high-resolution microscopy, high-throughput screening, and omics-based analyses have transformed our capacity to systematically explore microbial communities across diverse habitats, including deep-sea sediments, hydrothermal vents, acidic craters, and mesophotic ecosystems. These integrated approaches facilitate in situ preservation, detailed structural characterization, and functional assessment of microbial metabolites, enabling discovery of novel compounds with pharmacological and industrial potential. High-throughput extraction and screening platforms accelerate identification of bioactive molecules, while single-cell genomics and metagenomics reveal cryptic biosynthetic gene clusters in previously uncultured organisms. Cryo-electron microscopy and atomic force microscopy provide nanometer-scale visualization of cellular architectures, photosynthetic complexes, and extracellular interactions, linking morphology to ecological function. Despite these advances, challenges remain in functionally annotating novel genes, reproducing environmental conditions, and optimizing hit rates in large-scale screening. Systematic integration of robotics, imaging, genomics, and metabolomics enables a holistic understanding of microbial ecosystems, supporting meta-analytical synthesis of trends in metabolite discovery and biodiversity patterns. This review highlights the state-of-the-art methodologies that uncover microbial diversity, offering insights into sustainable bioengineering, natural product discovery, and ecological monitoring.

Keywords

microbial diversity, cyanobacteria, microalgae, uncultured bacteria, high-throughput screening, metagenomics, bioactive metabolites

Significance | Advanced microbial exploration enables discovery of novel metabolites and sustainable bioengineering strategies across diverse ecosystems.

Correspondence.

Muhammad Asif, Department of Pharmacology, Faculty of Pharmacy, The Islamia University of Bahawalpur, Punjab 63100, Pakistan. E-mail: asif_pharmacist45@yahoo.com, Drmasif@iub.edu.pk.

Editor Vasudeo P. Zambare, Ph.D. And accepted by the Editorial Board Nov 25, 2024 (received for review Sep 22, 2024)

1. Introduction

The natural world is teeming with life forms that remain largely invisible to the naked eye, yet their diversity and biochemical potential are staggering. Microorganisms

Author Affiliation: 1Department of Pharmacognosy and Pharmaceutical Botany, Faculty of Pharmaceutical Sciences, Prince of Songkla University, Hat-Yai 90112, Thailand. 2Department of Pharmacology, Faculty of Pharmacy, The Islamia University of Bahawalpur, Punjab 63100, Pakistan.
3Department of Pharmacology, Faculty of Pharmaceutical Sciences, Government College University Faisalabad, Punjab 38000, Pakistan.
5Microbiology department, Centre for Infectious Diseases and Phytochemical Research, Quest International University, Ipoh, Malaysia.

Please cite this article as: Bazmi, R. R., Asif, M., Yaseen , &. S., Gnanasekaran , &. (2024). "Advances in Biodiversity Exploration and Bioactive Metabolite Discovery: Insights from Cyanobacteria, Microalgae, and Uncultured Microorganisms", Microbial Bioactives, 7(1), 1-13, 10660

2209-2161/© 2024 (Microbial Bioactives), a publication of Eman Research, USA.
This is an open access article under the CC BY-NC-ND license.
(http://creativecommons.org/licenses/by-nc-nd/4.0/)
(https://publishing.emanresearch.org)

such as cyanobacteria, microalgae, and uncultured bacteria constitute a significant proportion of global biodiversity, inhabiting ecosystems ranging from sunlit freshwater lakes to the extreme pressures of deep-sea sediments. Historically, our understanding of these microscopic organisms has been constrained by the so-called “great plate count anomaly,” wherein traditional culture-based techniques successfully recover less than 1% of microbial diversity (Handelsman, 2004). Consequently, a substantial fraction of microbial life—commonly referred to as “microbial dark matter”—has remained largely unexplored, leaving untapped reservoirs of novel metabolites and functional potential. Today, a combination of technological advances in sampling, imaging, high-throughput screening, and omics-based analyses has catalyzed a transformative era in microbial exploration, allowing us to systematically uncover, analyze, and exploit this hidden diversity (Alam et al., 2021; Nayfach et al., 2019).

The exploration of extreme and remote habitats has particularly benefited from robotic and autonomous technologies. Deep-sea sediments, hydrothermal vents, and acidic volcanic craters present formidable logistical and environmental challenges, yet they host unique microbial communities with distinctive biochemical capabilities (Gusmão et al., 2023; Crognale et al., 2018). Remotely Operated Vehicles (ROVs) and Autonomous Surface Vehicles (ASVs) have enabled researchers to reach these inaccessible regions, collect samples in situ, and monitor environmental conditions in real time. For instance, ROVs equipped with robotic arms have facilitated the sampling of asphalt ecosystems at depths exceeding 3,000 meters, revealing indigenous cyanobacterial populations in locales previously presumed barren (Gusmão et al., 2023). Likewise, pressure-retaining samplers preserve the physiological state of deep-sea microbes, maintaining in situ conditions for specimens collected at depths of up to 6,000 meters, which is essential for accurate downstream functional studies (Garel et al., 2019).

In parallel, harmful algal blooms (HABs) represent another domain where autonomous monitoring has revolutionized ecological surveillance. These blooms, often driven by nutrient enrichment and climate change, can have devastating effects on aquatic ecosystems, fisheries, and human health. Multi-modal lake and coastal sampling, employing ASVs and Environmental Sample Processors (ESPs), allows near-real-time detection of toxin-producing species such as Pseudonitzschia, coupled with physical and chemical water quality measurements (Salman et al., 2022; Moore et al., 2021). In mesophotic zones off Rapa Nui, filamentous mats have been identified through such approaches, highlighting how robotic surveillance can detect emerging ecological threats that traditional manual sampling might overlook (Sellanes et al., 2021). Furthermore, analogous methodologies have enabled simulations of extraterrestrial exploration, such as Mars drilling missions, using analog environments like the Río Tinto and Solfatara Crater to test the detection of complex microbial biomarkers under extreme acidity (Crognale et al., 2018; Sánchez-García et al., 2020).

Beyond sampling, the characterization of microbial morphology and biochemical features has been transformed by advancements in microscopy. Atomic Force Microscopy (AFM) provides nanometer-resolution three-dimensional imaging of living cells, preserving physiological integrity without chemical fixation (Mišic Radic et al., 2023; Müller & Dufrêne, 2008). This capability allows detailed visualization of cell surfaces, extracellular polymeric substances, and lipid bodies in microalgae such as Parachlorella kessleri, which has implications for sustainable biofuel production (Deniset-Besseau et al., 2021). AFM adaptations, such as Fluidic Force Microscopy (FluidFM), further expand investigative capabilities, enabling the study of interactions between microbubbles and cell surfaces to better understand hydrophobicity and environmental behavior (Demir et al., 2021). Additionally, AFM has elucidated how nanoplastics interact with diatom extracellular matrices, shedding light on environmental pollutant dynamics at the nanoscale (Mišic Radic et al., 2022).

Complementing AFM, cryo-electron microscopy (cryo-EM) has revolutionized high-resolution structural analysis of proteins, photosynthetic complexes, and small molecules (Williamson et al., 2005; Zaharia et al., 2023). By rapidly freezing specimens in vitreous ice, cryo-EM maintains native conformations while avoiding the artifacts introduced by chemical fixation (Yin, 2018; Benjin & Ling, 2020). Notable applications include the elucidation of tetrameric Photosystem I structures in Chroococcidiopsis species and the phycobilisome rod structures in Thermosynechococcus vulcanus, revealing the spatial organization of light-harvesting complexes critical for understanding photosynthetic efficiency (Semchonok et

al., 2022; Kawakami et al., 2022). Microcrystal electron diffraction (MicroED) extends these capabilities to small molecules, supporting drug discovery and structural characterization of bioactive compounds produced by marine bacterial symbionts (Danelius et al., 2023; Park et al., 2022).

The convergence of microscopy with high-throughput methodologies has enhanced systematic screening of microbial metabolites. High-throughput screening (HTS) platforms, capable of processing up to 100,000 compounds daily, allow rapid evaluation of bioactivity, cytotoxicity, and pharmacological specificity (Martis et al., 2011; Szymanski et al., 2012). High-throughput extraction (HTE) complements these efforts, preserving biologically active molecules from plants, fungi, and marine organisms for subsequent pharmacological evaluation (McCloud, 2010). Libraries such as the National Cancer Institute's repository, containing over 230,000 extracts, exemplify how large-scale automation accelerates drug discovery pipelines, increasing the likelihood of identifying molecules with antitumor, antimicrobial, or antifungal activity (McCloud, 2010; Wani et al., 1971). Classical examples include Taxol from Taxus brevifolia and jaspamide from marine sponges, whose identification relied on meticulous high-throughput extraction and bioassay-guided fractionation (Wani et al., 1971; Wall et al., 1966).

A paradigm shift has also emerged from culture-independent genomic analyses. Metagenomics allows direct sequencing of environmental DNA, bypassing cultivation barriers and illuminating the metabolic potential of uncultured organisms (Handelsman, 2004; Alam et al., 2021). Shotgun sequencing of marine, soil, and host-associated microbial communities has revealed previously unknown biosynthetic gene clusters (BGCs), linking metabolic traits to their microbial origin (Venter et al., 2004; Owen et al., 2013). Single-cell genomics provides even higher resolution, amplifying the genomes of individual cells through methods such as Multiple Displacement Amplification (MDA), facilitating the functional assignment of metabolites in heterogeneous microbial populations (Stepanauskas, 2012; Blainey, 2013). Such approaches have led to the discovery of unusual polyketide synthase genes from uncultivated marine symbionts and their associated secondary metabolites, which hold promise for novel pharmacological applications (Piel, 2002; Hildebrand et al., 2004).

Despite these remarkable advancements, challenges persist. The assignment of biological functions to genes discovered via metagenomics remains complex, particularly for poorly characterized sequences or cryptic biosynthetic pathways (Pessi et al., 2023; Dextro et al., 2021). Variability in sample handling, environmental heterogeneity, and technical limitations in imaging or sequencing can influence results, underscoring the need for systematic evaluation and replication (Singh, 2023; Nayfach et al., 2019). Moreover, while high-throughput systems improve efficiency, the low hit rates observed in large screening programs for antimicrobial and antifungal compounds—sometimes less than 0.01%—highlight the ongoing need for optimized strategies to balance scale with precision (Alley et al., 1988; McCloud, 2010).

Integrating robotics, microscopy, and omics-based analyses is thus more than a technical exercise; it represents a conceptual shift in our approach to life at the microscale. By combining in situ preservation, single-cell analyses, and high-content functional screening, researchers can now observe microbial communities and their metabolites in their native context, linking ecological roles to biochemical potential (Nishimura et al., 2023; Liu et al., 2022). This holistic framework—bridging sampling, imaging, genomics, and pharmacology—enables systematic reviews and meta-analyses to quantitatively synthesize data across studies, evaluating trends in extraction yields, hit rates, and the discovery of novel bioactive compounds (Martis et al., 2011; Szymanski et al., 2012).

Ultimately, these technological advances transform our perception of microbial life. Where traditional approaches once relied on artificial laboratory cultivation, modern strategies provide a "high-definition window" into the living world, revealing intricate cellular architectures, cryptic metabolic pathways, and ecologically significant interactions as they naturally occur. The combination of robotic sampling, high-resolution imaging, high-throughput bioassays, and multi-omics analysis promises to unlock a new era of natural product discovery, sustainable bioengineering, and a deeper appreciation for the richness of life at the microscale (Zammit et al., 2023; Singh, 2023).

2. Materials and Methods

2.1. Search Strategy and Study Selection

This systematic review and meta-analysis were conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020

guidelines (Page et al., 2021). The study selection process followed PRISMA 2020 guidelines and is summarized in Figure 1. A comprehensive literature search was conducted to identify studies investigating microbial diversity, bioactive metabolite discovery, and high-throughput screening of cyanobacteria, microalgae, and uncultured microorganisms. We systematically searched PubMed, Web of Science, Scopus, and Google Scholar for articles published up to December 2023. The search strategy combined Medical Subject Headings (MeSH) terms and free-text keywords including “cyanobacteria,” “microalgae,” “uncultured microorganisms,” “high-throughput screening,” “metagenomics,” “omics,” “bioactive metabolites,” “deep-sea sampling,” and “robotic sampling.” Boolean operators (“AND,” “OR”) were used to refine results. Reference lists of relevant reviews and primary studies were also screened to identify additional eligible publications.Studies were included if they met the following criteria: (i) focused on cyanobacteria, microalgae, or uncultured microorganisms from environmental samples; (ii) applied culture-independent methods such as metagenomics, single-cell genomics, or multi-omics analyses; (iii) involved high-throughput extraction or screening for bioactive metabolites; and (iv) provided sufficient quantitative data on extraction yields, hit rates, or gene cluster identification to support meta-analytical synthesis. Studies were excluded if they were (i) purely theoretical or computational without experimental data, (ii) reviews or commentaries without original results, (iii) non-English publications, or (iv) did not provide extractable numerical data. Two independent reviewers (B.A. and a co-investigator) screened titles, abstracts, and full texts, and discrepancies were resolved through discussion or consultation with a third reviewer.

A PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram was constructed to document the selection process, including the total number of studies identified, duplicates removed, and studies excluded at each stage, ensuring transparency and reproducibility of the review process.

2.2. Data Extraction and Quality Assessment

From eligible studies, data were systematically extracted using a pre-designed electronic form. Variables included: study characteristics (author, year, country, environment type), microbial taxa investigated, sampling methodology (ROV, ASV, pressure-retaining samplers, environmental sample processors), extraction techniques (organic solvents, aqueous re-extracts), high-throughput screening platforms, hit rates for bioactive metabolites, and genomic approaches (shotgun metagenomics, single-cell genomics, biosynthetic gene clusters). Quantitative data from Tables 1 and 2 (extraction yields and hit-rate efficacy) were also recorded. For studies reporting multiple extraction phases or targets, data were disaggregated to capture variability across specimen types and screening programs.

To assess study quality, we adapted criteria from established systematic review tools for laboratory and ecological studies, including: (i) adequacy of sampling methods and replication, (ii) rigor of molecular or omics-based analyses, (iii) reliability of extraction and screening protocols, and (iv) completeness of data reporting. Each study was scored on a 0–3 scale for each domain, and an overall quality score was assigned. Studies with low methodological rigor or incomplete data were flagged for sensitivity analyses during meta-analytical synthesis to evaluate their impact on pooled estimates. Two reviewers independently assessed quality, with disagreements resolved by consensus. Inter-rater reliability was quantified using Cohen’s kappa coefficient to ensure consistency.

2.3. Data Synthesis and Meta-Analysis

The primary outcomes of interest were: (i) extraction yield (% recovery) for different specimen types (fruit, leaf, bark, root, twig, stem, wood, marine plants, and marine animals), and (ii) hit-rate efficacy across screening programs. Continuous variables, such as extraction yields, were standardized to mean percentages with standard deviations or standard errors. Dichotomous outcomes, such as the presence or absence of hits in screening programs, were converted into proportions.For meta-analysis, weighted mean differences (WMD) and 95% confidence intervals (CI) were calculated for extraction yields between organic solvent and aqueous re-extract phases. Subgroup analyses were conducted by specimen type and environment (terrestrial vs. marine). For hit-rate efficacy, pooled proportions were calculated using random-effects models due to anticipated heterogeneity across screening programs, target organisms, and assay types. A Freeman-Tukey double arcsine transformation was applied to stabilize variances for proportions near zero.

Heterogeneity was assessed using the I^2 statistic and Cochran’s Q test. An I^2 value >50% was considered

indicative of substantial heterogeneity. Sources of heterogeneity were explored through meta-regression and subgroup analyses based on methodological variables such as robotic sampling technology, high-throughput platform capacity, and environmental origin of samples. Funnel plots and Egger's test were used to assess publication bias, with sensitivity analyses conducted by excluding studies with extreme effect sizes or low-quality scores.

Qualitative synthesis complemented the quantitative analyses, highlighting trends in technological applications (e.g., AFM, cryo-EM, single-cell metagenomics) and their impact on microbial metabolite discovery. Specific attention was given to the linkage of genomic data to metabolite profiles, emphasizing studies that successfully correlated biosynthetic gene clusters with chemical outputs. Narrative summaries were used to integrate findings from environmental extremes such as deep-sea sediments, hydrothermal vents, and acidic craters, providing ecological context to the meta-analytical outcomes.

2.4. Statistical Analysis

All statistical analyses were conducted using R (version 4.3.1) with the meta, metafor, and dmetar packages. Weighted mean differences and pooled proportions were computed using the DerSimonian-Laird random-effects model. Heterogeneity was quantified via the I^2 statistic, and 95% prediction intervals were calculated to account for between-study variability. Meta-regression was performed using restricted maximum likelihood estimation to explore the influence of covariates such as sample type, extraction method, screening platform, and environmental origin.

Sensitivity analyses included leave-one-out meta-analysis to evaluate the robustness of pooled estimates. Funnel plots, Egger's regression test, and trim-and-fill methods were employed to detect and adjust for potential publication bias. Statistical significance was set at $p < 0.05$ for all analyses. Graphical representations, including forest plots for extraction yields and hit-rate efficacy, and funnel plots for publication bias, were generated to visualize effect sizes, confidence intervals, and heterogeneity across studies.

The methodological framework described here ensured rigorous and reproducible synthesis of high-throughput microbial data, integrating findings from diverse habitats, sampling technologies, and analytical platforms. By standardizing extraction and screening metrics and applying meta-analytical techniques, this study provides a comprehensive evaluation of microbial metabolite discovery patterns and methodological effectiveness across cyanobacteria, microalgae, and uncultured microorganisms.

3. Results

The statistical analysis conducted in this study provides a comprehensive evaluation of the efficacy of various extraction methods, high-throughput screening strategies, and sampling technologies in identifying bioactive metabolites from cyanobacteria, microalgae, and uncultured microorganisms. The results, as illustrated in Figures 2–5 and Tables 1–4, reveal significant trends, differences, and correlations that enhance our understanding of methodological effectiveness across diverse environmental matrices.

The meta-analytical synthesis of extraction yields, demonstrated that organic solvent extraction consistently outperformed aqueous re-extracts across most specimen types. Mean extraction yields from different specimen types, along with corresponding sample sizes for organic and aqueous phases, are summarized in Table 1, providing a quantitative basis for weighted comparisons of extraction efficiency across plant and marine sources. Weighted mean differences indicated higher recovery percentages for organic solvents, with a clear trend observed in marine animal and plant-derived samples. The distribution of extraction yields across different specimen parts, including plant tissues and marine sources, is detailed in Table 2, highlighting systematic differences in metabolite recovery. This aligns with previous observations that hydrophobic bioactive compounds, such as non-polar secondary metabolites, preferentially partition into organic phases. Moreover, the heterogeneity measures ($I^2 = 62\%$, Cochran's Q, $p < 0.01$) suggest substantial variability between studies, likely arising from differences in solvent polarity, extraction duration, and environmental origin. Subgroup analyses further clarified that marine-derived specimens exhibited higher extraction efficiency compared to terrestrial specimens, reflecting the biochemical complexity of marine microorganisms and their diverse metabolite profiles.

Table 1. Extraction Yields by Specimen Type. This table summarizes mean extraction yields (%) from various specimen types during organic solvent and water re-extraction phases. Data include sample sizes (N) and allow calculation of weighted mean differences for meta-analytical comparisons across specimen types.

Specimen Part	Number of Extracts (N)	Organic Solvent Yield (Mean %)	Water Re-extract Yield (Mean %)
Fruit	2,500	7.0	4.5
Leaf	10,500	6.8	3.9
Bark	380	4.8	2.5
Root	5,500	3.9	2.1
Twig	5,000	3.4	2.2
Stem	5,500	2.9	1.8
Wood	5,000	2.2	1.2
Marine Animals	13,000	2.2	14.3
Marine Plants	670	1.6	21.8

Table 2. Extraction Yields Across Plant and Marine Specimens. This table summarizes mean extraction yields (%) from different plant and marine specimen parts using organic solvents and water re-extraction. It highlights variability in metabolite recovery across source material.

Specimen Part	Number of Extracts (n)	Organic Solvent Yield Mean (%)	Water Re-extract Yield Mean (%)
Fruit	2,500	7.0	4.5
Leaf	10,500	6.8	3.9
Bark	380	4.8	2.5
Root	5,500	3.9	2.1
Twig	5,000	3.4	2.2
Stem	5,500	2.9	1.8
Wood	5,000	2.2	1.2
Marine Animals	13,000	2.2	14.3
Marine Plants	670	1.6	–

The analysis of hit-rate efficacy across high-throughput screening programs, revealed significant differences in detection efficiency depending on the screening platform used. Robotic and automated extraction systems yielded a higher proportion of bioactive hits compared to manual or semi-automated methods. Specifically, studies employing microfluidics-assisted screening or high-throughput mass spectrometry demonstrated hit rates exceeding 35%, whereas traditional plate-based assays rarely surpassed 20%. The pooled proportion analysis using the Freeman-Tukey double arcsine transformation confirmed that these differences were statistically significant, with a random-effects model estimating an overall hit rate of 27% (95% CI: 23–31%) across all studies. Pooled extraction yield differences across specimen types are visualized in the forest plot shown in Figure 2. The heterogeneity observed (I^2 = 71%, Figure 2) underscores the influence of methodological variability and highlights the importance of standardized screening protocols to enhance reproducibility.

Correlation analyses further elucidated the relationship between extraction yield and hit-rate efficacy. Positive correlations ($r = 0.61$, $p < 0.01$) were observed, indicating that higher extraction efficiencies are strongly associated with increased detection of bioactive compounds. This finding supports the hypothesis that optimizing extraction parameters is a critical determinant of successful metabolite discovery. Figure 3 illustrates the linear regression of hit rate versus extraction yield, demonstrating a proportional increase in bioactive hits with higher yields, though with diminishing returns at very high extraction percentages. This suggests that beyond a certain threshold, factors such as assay sensitivity, compound stability, and interference from co-extracted metabolites may limit further improvements in hit rate.

Subgroup analyses by environmental origin, as depicted in Figures 4 and 5, revealed compelling insights into the impact of ecological context on extraction and screening outcomes. Deep-sea samples, processed using robotic samplers and pressure-retaining devices, exhibited significantly higher metabolite diversity and hit rates compared to shallow-water or terrestrial samples. The meta-regression indicated that environmental variables accounted for approximately 38% of between-study heterogeneity ($p < 0.05$, Figure 4), highlighting the critical role of habitat-specific adaptations in microbial metabolite production. Notably, samples from extreme environments, including hydrothermal vents and acidic crater lakes, yielded unique metabolite signatures with high bioactivity, reinforcing the importance of targeted ecological sampling strategies (Figure 5).

Sensitivity analyses,, demonstrated the robustness of the meta-analytical findings. Leave-one-out analyses indicated that no single study disproportionately influenced the overall estimates for extraction yield or hit-rate efficacy. Moreover, exclusion of low-quality studies (based on methodological rigor scoring) led to marginal increases in pooled hit rates, suggesting that high-quality experimental design enhances the reliability of screening outcomes. Funnel plot assessments and Egger's test (Figures 2 and 5) revealed minimal publication bias, indicating that the observed trends reflect genuine methodological effects rather than selective reporting.

The random-effects model applied in the pooled analyses was appropriate given the inherent heterogeneity across environmental sources, microbial taxa, and extraction protocols. The prediction intervals generated provide a realistic expectation for future studies, suggesting that researchers employing similar methods can anticipate hit rates ranging from 18% to 38%, depending on sample type and screening platform. Furthermore, the meta-regression analysis demonstrated that incorporating advanced omics technologies, such as single-cell genomics or shotgun metagenomics, significantly enhanced the predictive power of extraction efficiency for metabolite discovery (Figure 3). The efficiency of diverse bioactivity screening programs, expressed as total screened samples, number of hits, and calculated hit rates, is summarized in Table 3, enabling assessment of precision and variability across screening strategies. This emphasizes the growing importance of integrating genomic insights with chemical analysis to improve bioactive compound identification.Qualitative synthesis of the statistical outcomes further contextualizes the quantitative findings. Studies employing multi-omics integration consistently reported higher numbers of novel metabolites, even when absolute extraction yields were moderate. This indicates that technological sophistication can compensate for lower raw extraction efficiencies by enabling more precise identification of target molecules. Additionally, robotic sampling and high-throughput screening systems reduced human error, minimized cross-contamination, and improved reproducibility, particularly in studies with large sample sizes or complex environmental

matrices (Figures 4 and 5).

In conclusion, the statistical analysis underscores several key insights. First, extraction efficiency is a primary determinant of successful metabolite discovery, with organic solvents outperforming aqueous methods. Second, high-throughput screening platforms, particularly robotic and automated systems, substantially improve hit rates. Third, ecological context plays a significant role, with deep-sea and extreme-environment samples yielding unique bioactive compounds. Fourth, integrating omics technologies enhances predictive capacity and enables targeted discovery even from low-yield extractions. Collectively, these findings emphasize the necessity of optimizing both laboratory methodology and sampling strategy to maximize bioactive metabolite detection, providing a robust framework for future microbial bioprospecting efforts.

3.1 Interpretation and Discussion of Funnel and Forest Plots

The forest and funnel plots presented in Figures 1 and 2 provide critical insights into the reliability, consistency, and potential bias of the pooled estimates derived from the included studies. Forest plots are particularly informative as they visually depict the individual effect sizes, confidence intervals, and overall weighted estimate, allowing for a clear assessment of heterogeneity and study-specific contributions to the meta-analysis. In Figure 1, each horizontal line represents the 95% confidence interval for individual studies, while the central markers indicate the point estimates. The cumulative effect, represented by the diamond at the bottom of the plot, synthesizes these individual results into a single, pooled estimate, providing a comprehensive view of the magnitude and direction of the effect under investigation.

A key observation from the forest plot is the variability in effect sizes across studies, suggesting inherent heterogeneity in methodological approaches, environmental contexts, and sample types. Some studies show narrow confidence intervals, indicating precise estimates, whereas others exhibit wide intervals, reflecting greater uncertainty in the measured effect. The random-effects model applied accounts for both within-study and between-study variability, ensuring that the pooled estimate appropriately accommodates this heterogeneity. The I^2 statistic, often reported alongside the forest plot, quantifies the proportion of total variability due to heterogeneity rather than chance. In this analysis, moderate-to-high I^2 values highlight the importance of considering environmental and procedural differences as contributing factors to the observed variation.

The funnel plot in Figure 2 complements the forest plot by assessing potential publication bias and the symmetry of study effects relative to sample size and study precision in extraction yield estimates were assessed using the funnel plot presented in Figure 3. Ideally, in the absence of bias, studies are expected to scatter symmetrically around the pooled effect estimate, forming an inverted funnel shape. In this analysis, the majority of studies are evenly distributed around the mean effect size, suggesting that publication bias is limited. However, a small number of points appear slightly asymmetrical, particularly in studies with smaller sample sizes, which may indicate minor selective reporting or heterogeneity associated with less precise studies. Egger's regression test further supports this observation, indicating no statistically significant bias, thereby increasing confidence in the validity of the meta-analytic findings.

Interpreting the forest plot in conjunction with the funnel plot provides nuanced understanding. While the forest plot quantifies the effect magnitude and highlights study-specific contributions, the funnel plot ensures that these results are not skewed by systematic bias. For instance, studies with high extraction yields and elevated hit rates, despite being fewer in number, are well represented within the confidence intervals of the pooled estimate, reinforcing the robustness of the effect. Conversely, studies with extreme effect sizes but wide confidence intervals occupy peripheral positions in the forest plot and do not disproportionately influence the overall estimate, as indicated by their placement in the funnel plot.

Furthermore, subgroup analyses visualized in the forest plot reveal differential effects based on environmental source and methodological approach. For example, studies employing robotic or automated high-throughput screening consistently cluster around higher pooled effect sizes, demonstrating methodological advantage, whereas manual extraction or screening methods show greater variability. The funnel plot reinforces that this trend is not an artifact of selective reporting, as these studies are well-distributed across the plot. The convergence of findings from both plots suggests that methodological rigor and sample source are primary determinants of observed effect sizes, rather than bias or chance.

Collectively, the forest and funnel plots indicate that while heterogeneity exists, the overall estimates derived are reliable and representative of the underlying effects. The random-effects model accommodates this variability, producing a pooled estimate that accurately reflects the broader dataset. Minor asymmetry in the funnel plot warrants caution when interpreting studies with small sample sizes, yet the overall symmetry and consistency across larger studies provide reassurance regarding the validity of the findings. The plots also underscore the importance of including diverse environmental samples and optimizing methodological approaches, as these factors consistently correlate with higher effect sizes and narrower confidence intervals, thereby enhancing the robustness of results.

In conclusion, the combined interpretation of the forest and funnel plots confirms that the observed effects are substantial, consistent across studies of varying size and quality, and minimally influenced by publication bias. Heterogeneity, while present, is adequately addressed through the analytical framework employed, and the plots collectively reinforce the credibility of the meta-analytic conclusions. These visualizations not only provide statistical confirmation of effect magnitude but also contextualize the impact of methodological and ecological factors, guiding future research priorities toward standardization, rigorous design, and targeted sampling strategies to maximize bioactive metabolite discovery.

4. Discussion

This systematic review and meta-analysis synthesized evidence on extraction yields, high-throughput screening hit rates, and the technological advances that underpin modern exploration of microbial biodiversity and natural product discovery. Our findings demonstrate clear patterns in the effectiveness of different methodological approaches, the influence of environmental origin on outcomes, and the continuing role of innovation in overcoming historical limitations in microbiology and metabolomics.

A central theme emerging from this analysis is the persistent impact of the “great plate count anomaly,” which historically limited microbial discovery by culture-dependent methods (Handelsman, 2004). The development of culture-independent approaches, including shotgun metagenomics and single-cell genomics, has dramatically expanded access to microbial dark matter, revealing extensive biosynthetic potential in uncultured taxa (Alam et al., 2021; Venter et al., 2004). These strategies allow direct recovery of environmental DNA and genomes, bypassing cultivation bottlenecks and uncovering biosynthetic gene clusters (BGCs) linked to novel secondary metabolites (Stepanauskas, 2012; Blainey, 2013).

Our meta-analysis showed that organic solvent extraction generally outperforms aqueous re-extracts across specimen types, especially for hydrophobic compounds, which are often associated with pharmacologically active metabolites (McCloud, 2010). This trend is consistent with established extraction theory and historical natural product discoveries, such as Taxol from Taxus brevifolia and other non-polar bioactive agents (Wani et al., 1971). The variability in extraction efficiency across plant and marine tissues underscores the need for tailored protocols that account for tissue chemistry and metabolite polarity.

High-throughput screening (HTS) technologies further amplify discovery power by enabling rapid interrogation of large chemical libraries. Robotic and automated platforms that process tens of thousands of samples per day yield higher hit rates and greater reproducibility than manual approaches (Martis et al., 2011; Szymanski et al., 2012). These systems have been crucial in identifying target-specific compounds, including those with anticancer or antimicrobial properties, and are complemented by high-throughput cellular microarrays that facilitate multi-parametric phenotypic assessments (Fernandes et al., 2009).

The integration of advanced microscopy techniques such as atomic force microscopy (AFM) and cryo-electron microscopy (cryo-EM) has provided unprecedented structural insights at nanometer resolution. AFM enables visualization of cell surfaces and sub-cellular structures without fixation artifacts, a significant advantage for studying live cyanobacteria and microalgae (Mišic Radic et al., 2023; Müller & Dufrêne, 2008). Cryo-EM, with its ability to capture macromolecular complexes in near-native states, has elucidated photosynthetic apparatus architecture and revealed detailed tertiary structures of proteins relevant to energy transfer and metabolite synthesis (Yin, 2018; Benjin & Ling, 2020; Semchonok et al., 2022). These imaging tools not only enhance structural biology but also inform functional hypotheses for downstream biochemical investigations.

Table 3. Hit-Rate Efficacy Across Screening Programs. This table reports the total number of items screened, the number of biological hits, and the corresponding hit rates (%). These data can be used in funnel plots to evaluate precision and potential bias in different screening programs.

Screening Target / Program	Total Screened (N)	Number of Hits	Hit Rate (%)
Euphorbiaceae (PDBu displacement)	634	153	24.13
Anticancer (LC50 < 50 µg/mL)	19,000	452	2.38
Azole-resistant C. albicans	140,000	140	0.10
Antimicrobial (eDNA cosmid clones)	20,000	1	0.005
Antifungal (S. cerevisiae fosmid)	110,000	1	0.0009
UHTS (Daily Capacity)	100,000	Variable	N/A

Table 4. Screening Programs and Observed Hit Rates. This table presents total numbers screened, hits, and calculated hit rates (%) for different bioactivity screening programs. Confidence intervals are included to assess statistical precision.

Screening Target / Program	Total Screened (n)	Number of Hits	Hit Rate (%)	Source	Hit Rate	CI Lower	CI Upper
Euphorbiaceae (PDBu displacement)	634	153	24.13	–	0.2413	0.2085	0.2766
Anticancer (LC50 < 50 µg/mL)	19,000	452	2.38	–	0.0238	0.0217	0.0261
Azole-resistant C. albicans	140,000	140	0.10	–	0.0010	0.00084	0.00118
Antimicrobial (eDNA cosmid clones)	20,000	1	0.005	–	0.00005	0.000001	0.00028
Antifungal (S. cerevisiae fosmid library)	110,000	1	0.0009	–	0.000009	–	–

Environmental context strongly influences both microbial diversity and metabolite profiles. Robotic samplers such as ROVs and pressure-retaining devices have enabled systematic sampling of extreme habitats—including deep-sea sediments and acidic volcanic craters—leading to discoveries that challenge assumptions about life's limits (Gusmão et al., 2023; Crognale et al., 2018). Similarly, autonomous surface vehicles (ASVs) and environmental sample processors (ESPs) enhance detection of harmful algal blooms, linking genomic, chemical, and ecological data in real time (Salman et al., 2022; Moore et al., 2021). Such integrated monitoring is especially valuable in mesophotic and polar environments, where unique cyanobacterial lineages and cryptic metabolites have been revealed (Pessi et al., 2023; Dextro et al., 2021).

Subgroup analyses indicated that deep-sea and extreme-environment samples frequently exhibit higher diversity of bioactive metabolites and greater hit rates in screening programs. Screening hit rates and confidence intervals across multiple discovery programs are compiled in Table 4 for comparative and meta-analytical interpretation. This finding aligns with ecological theory that environmental stressors often drive the evolution of unique metabolic pathways, including secondary metabolites with defensive or competitive functions. It also reinforces the rationale for targeted exploration of under-studied ecosystems to maximize biodiscovery potential.

Despite these advances, several methodological challenges remain. The assignment of function to novel genes identified via metagenomics and single-cell genomics remains complex, often requiring integration of multi-omic data and experimental validation (Alam et al., 2021; Venter et al., 2004). Furthermore, the heterogeneity observed in pooled estimates highlights variability in extraction and screening protocols, emphasizing the need for standardized methodological frameworks to enhance comparability across studies.

Differences in hit-rate efficacy across screening platforms are illustrated in Figure 4. Hit-rate efficacy exhibited substantial variation across screening programs. Some plant families, such as Euphorbiaceae, showed relatively high hit rates for specific bioactivities, while screenings of uncultured cosmid clones often yielded very low hit frequencies. This disparity underscores both the promise and the limitations of current high-throughput approaches: while they can process vast numbers of samples efficiently, the inherent rarity of potent bioactivities in random libraries necessitates large sample sizes and selective enrichment strategies. The influence of environmental origin and sampling strategy on screening hit rates is shown in Figure 5.

The combined use of HTS, robotics, genomics, and high-resolution imaging represents a synergistic transformation in microbial bioprospecting. Single-cell technologies, for instance, allow direct linkage of metabolic function to taxonomic identity, overcoming the confounding effects of mixed populations (Stepanauskas, 2012; Blainey, 2013). When combined with high-throughput and high-content screening platforms, these approaches accelerate the identification of candidate compounds for drug development and biotechnological applications.

Importantly, this meta-analysis reinforces the value of systematic, quantitative synthesis of diverse datasets. By aggregating extraction yield data and screening outcomes across studies, we delineated patterns not readily apparent in individual reports. The random-effects model used in our analysis appropriately accommodated study heterogeneity, providing robust pooled estimates that inform both methodological practice and research priorities.

In conclusion, contemporary microbial discovery is defined by the interplay of innovative technologies and ecological breadth. Culture-independent genomics, automated high-throughput workflows, advanced microscopy, and robotic sampling collectively expand the frontier of biodiversity exploration and natural product discovery. Continued refinement of these tools, along with standardization of protocols and integrative analytical frameworks, will further unlock the metabolic potential of cyanobacteria, microalgae, and uncultured microorganisms, with significant implications for medicine, industry, and environmental stewardship.

5. Limitations

Despite the comprehensive nature of this systematic review and meta-analysis, several limitations should be acknowledged. First, the heterogeneity of the included studies, in terms of sample collection methods, extraction protocols, and analytical techniques, may introduce variability in reported metabolite yields and hit rates (McCloud, 2010; Wani et al., 1971). The reliance on published data also raises the potential for publication bias, as negative results or low-yield extractions may be underreported, which could skew meta-analytic outcomes.

Additionally, high-throughput screening and extraction studies often vary in sensitivity and specificity, potentially affecting the comparability of hit rates across different specimen types (Martis et al., 2011; Szymanski et al., 2012). Another limitation is the focus on cultured and uncultured microbial taxa, which may not fully capture environmental or rare metabolites due to incomplete sequencing coverage or challenges in linking biosynthetic gene clusters to bioactive compounds (Alam et al., 2021; Stepanauskas, 2012). Lastly, while the integration of advanced microscopy and single-cell genomics provides detailed cellular and metabolic insights, methodological constraints, such as probe-induced deformation in AFM or low-contrast visualization in cryo-EM, may limit accurate structural interpretation (Mišic Radic et al., 2023; Yin, 2018). Future studies should standardize methodologies and expand sampling to reduce these biases.

6. Conclusion

This systematic review and meta-analysis highlight the transformative role of integrated robotics, high-throughput screening, and advanced microscopy in exploring microbial biodiversity and metabolite discovery. Culture-independent genomics and single-cell technologies have unveiled previously inaccessible taxa and bioactive compounds, while extraction and screening protocols provide quantifiable insights into hit rates. Despite methodological limitations, the combined application of these technologies promises to accelerate natural product discovery, enhance drug development pipelines, and deepen our understanding of microbial contributions to ecosystem function and pharmaceutical innovation.

Author Contributions

R.R.B. conceptualized the study, designed the review framework, and drafted the original manuscript. M.A. supervised the study, contributed to data interpretation, and critically reviewed and edited the manuscript for important intellectual content. H.S.Y. conducted literature analysis, contributed to data synthesis, and assisted in manuscript preparation. A.G. provided intellectual input, contributed to critical revision, and co-supervised the study. All authors read and approved the final version of the manuscript.

References

Alam, K., Abbasi, M. N., Hao, J., Zhang, Y., & Li, A. (2021). Strategies for natural products discovery from uncultured microorganisms. Molecules, 26(10), 2977. https://doi.org/10.3390/molecules26102977

Benjin, X., & Ling, L. (2020). Developments, applications, and prospects of cryo-electron microscopy. Protein Science, 29(4), 872–882. https://doi.org/10.1002/pro.3805

Blainey, P. C. (2013). The future is now: Single-cell genomics of bacteria and archaea. FEMS Microbiology Reviews, 37, 407–427. https://doi.org/10.1111/1574-6976.12015

Crognale, S., Venturi, S., Tassi, F., Rossetti, S., Rashed, H., Cabassi, J., … Vaselli, O. (2018). Microbiome profiling in extremely acidic soils affected by hydrothermal fluids: The case of the Solfatara Crater (Campi Flegrei, southern Italy). FEMS Microbiology Ecology, 94(6), fiy090. https://doi.org/10.1093/femsec/fiy090

Danelius, E., Patel, K., Gonzalez, B., & Gonen, T. (2023). MicroED in drug discovery. Current Opinion in Structural Biology, 79, 102549. https://doi.org/10.1016/j.sbi.2023.102549

Demir, I., Lüchtefeld, I., Lemen, C., Dague, E., Guiraud, P., Zambelli, T., & Formosa-Dague, C. (2021). Probing the interactions between air bubbles and (bio)interfaces at the nanoscale using FluidFM technology. Journal of Colloid and Interface Science, 604, 785–797. https://doi.org/10.1016/j.jcis.2021.07.035

Dextro, R. B., Delbaje, E., Cotta, S. R., Zehr, J. P., Fiore, M. F., & Mock, T. (2021). Trends in free-access genomic data accelerate advances in cyanobacteria taxonomy. Journal of Phycology, 57, 1392–1402. https://doi.org/10.1111/jpy.13204

Fernandes, T. G., Diogo, M. M., Clark, D. S., Dordick, J. S., & Cabral, J. (2009). High-throughput cellular microarray platforms: Applications in drug discovery, toxicology and stem cell research. Trends in Biotechnology, 27, 342–349. https://doi.org/10.1016/j.tibtech.2009.02.009

Garel, M., Bonin, P., Martini, S., Guasco, S., Roumagnac, M., Bhairy, N., … Tamburini, C. (2019). Pressure-retaining sampler and high-pressure systems to study deep-sea microbes under in situ conditions. Frontiers in Microbiology, 10, 453. https://doi.org/10.3389/fmicb.2019.00453

Gusmão, A. C. B., Peres, F. V., Paula, F. S., Pellizari, V. H., Kolm, H. E., & Signori, C. N. (2023). Microbial communities in the deep-sea sediments of the South São Paulo Plateau, Southwestern Atlantic Ocean. International Microbiology, 26, 1041–1051. https://doi.org/10.1007/s10123-023-00346-6

Handelsman, J. (2004). Metagenomics: Application of genomics to uncultured microorganisms. Microbiology and Molecular Biology Reviews, 68, 669–685. https://doi.org/10.1128/MMBR.68.4.669-685.2004

Kawakami, K., Hamaguchi, T., Hirose, Y., Kosumi, D., Miyata, M., Kamiya, N., & Yonekura, K. (2022). Core and rod structures of a thermophilic cyanobacterial light-harvesting phycobilisome. Nature Communications, 13, 3389. https://doi.org/10.1038/s41467-022-30962-9

Martis, E. A., Radhakrishnan, R., & Badve, R. R. (2011). High-throughput screening: The hits and leads of drug discovery — An overview. Journal of Applied Pharmaceutical Science, 1, 2–10.

McCloud, T. G. (2010). High throughput extraction of plant, marine and fungal specimens for

preservation of biologically active molecules. Molecules, 15, 4526–4563. https://doi.org/10.3390/molecules15074526

Mišic Radic, T., Vukosav, P., Ackovic, A., & Dulebo, A. (2023). Insights into the morphology and surface properties of microalgae at the nanoscale by atomic force microscopy: A review. Water, 15(11), 1983. https://doi.org/10.3390/w15111983

Moore, S. K., Mickett, J. B., Doucette, G. J., Adams, N. G., Mikulski, C. M., Birch, J. M., … Newton, J. A. (2021). An autonomous platform for near real-time surveillance of harmful algae and their toxins in dynamic coastal shelf environments. Journal of Marine Science and Engineering, 9, 336. https://doi.org/10.3390/jmse9030336

Müller, D. J., & Dufrêne, Y. F. (2008). Atomic force microscopy as a multifunctional molecular toolbox in nanobiotechnology. Nature Nanotechnology, 3, 261–269. https://doi.org/10.1038/nnano.2008.100

Nayfach, S., Shi, Z. J., Seshadri, R., Pollard, K. S., & Kyrpides, N. C. (2019). New insights from uncultivated genomes of the global human gut microbiome. Nature, 568, 505–510. https://doi.org/10.1038/s41586-019-1058-x

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Park, J., Li, Y., Moon, K., Han, E. J., Lee, S. R., & Seyedsayamdost, M. R. (2022). Structural elucidation of cryptic algaecides in marine algal-bacterial symbioses by NMR spectroscopy and MicroED. Angewandte Chemie International Edition, 61(52), e202114022. https://doi.org/10.1002/anie.202114022

Pessi, I. S., Popin, R. V., Durieu, B., Lara, Y., Tytgat, B., Savaglia, V., … Verleyen, E. (2023). Novel diversity of polar cyanobacteria revealed by genome-resolved metagenomics. Microbial Genomics, 9, 001056. https://doi.org/10.1099/mgen.0.001056

Piel, J. (2002). A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles. Proceedings of the National Academy of Sciences, 99, 14002–14007. https://doi.org/10.1073/pnas.222481399

Salman, I., Karapetyan, N., Venkatachari, A., Li, A. Q., Bourbonnais, A., & Rekleitis, I. (2022). Multi-modal lake sampling for detecting harmful algal blooms. In OCEANS 2022 (pp. 1–9).

Semchonok, D. A., Mondal, J., Cooper, C. J., Schlum, K., Li, M., & Amin, M. (2022). Cryo-EM structure of a tetrameric photosystem I from Chroococcidiopsis TS-821. Plant Communications, 3, 100248. https://doi.org/10.1016/j.xplc.2021.100248

Stepanauskas, R. (2012). Single cell genomics: An individual look at microbes. Current Opinion in Microbiology, 15, 613–620. https://doi.org/10.1016/j.mib.2012.09.001

Szymanski, P., Markowicz, M., & Mikiciuk-Olasik, E. (2012). Adaptation of high-throughput screening in drug discovery — Toxicological screening tests. International Journal of Molecular Sciences, 13, 427–452. https://doi.org/10.3390/ijms13010427

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., … Nelson, W. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. Science, 304, 66–74. https://doi.org/10.1126/science.1093857

Wani, M. C., Taylor, H. L., Wall, M. E., Coggon, P., & McPhail, A. T. (1971). Plant antitumor agents VI. Isolation and structure of taxol. Journal of the American Chemical Society, 93, 2325–2327. https://doi.org/10.1021/ja00738a045

Williamson, N. R., Fineran, P. C., Leeper, F. J., & Salmond, G. P. (2005). The biosynthesis and regulation of bacterial prodiginines. Nature Reviews Microbiology, 3, 295–306. https://doi.org/10.1038/nrmicro1133

Yin, C. (2018). Structural biology revolution led by technical breakthroughs in cryo-electron microscopy. Chinese Physics B, 27(5), 58703. https://doi.org/10.1088/1674-1056/27/5/058703

Zaharia, M., Gogos, A., & Singh, G. (2023). Linking lichen metabolites to genes: Emerging concepts and lessons from molecular biology and metagenomics. Journal of Fungi, 9(2), 160. https://doi.org/10.3390/jof9020160