ISSN: 2458-8989

Natural and Engineering Sciences

NESciences, 2026, 11 (1): 32-48 doi: 10.28978/nesciences.261003

Innovative Approaches in Drug Discovery by Leveraging Molecular Ecology for the Identification of Novel Therapeutics

Izzatilla Khaydarov [1*], Ozod Abduganiyev [2], Fozil Irmatov [3], Mirkomil Obilov [4], Gulnoza Yusupova [5], Shakhnoza Latipova [6], Tatyana Mun [7]

[1*] Tashkent State University of Oriental Studies, Tashkent, Uzbekistan.
E-mail: izzatilla_haydarov@mail.ru

[2] Associate Professor, Scientific Secretary of the National Institute of Pedagogy and Character Education Named after Kori Niyozi, Tashkent, Uzbekistan.
E-mail: ozodabduganiyev222@gmail.com

[3] Associate Professor, Jizzakh State Pedagogical University, Jizzakh, Uzbekistan.
E-mail: irmatov-fozil-84@mail.ru

[4] Lecturer, Department of Management, Gulistan State University, Gulistan, Uzbekistan.
E-mail: mirkomilrashidovich@gmail.com

[5] Associate Professor, Tashkent State Medical University, Tashkent, Uzbekistan.
E-mail: gulnozaamanillaevna@gmail.com

[6] Associate Professor, Tashkent State Medical University, Tashkent, Uzbekistan.
E-mail: shaxnozalatipova977@gmail.com

[7] Associate Professor, Department of Hospital Orthopedic Dentistry, Tashkent State Medical University, Tashkent, Uzbekistan. E-mail: mun.tatyana@gmail.com

Abstract

The emergence of multifaceted diseases and antibiotic-resistant organisms has demonstrated the necessity to develop new therapeutic candidates, especially those of natural ecosystems. The current study integrates molecular ecology and the recent biotechnological approaches, such as metagenomics, metabolomics, and analyses of environmental DNA, to research on the biosynthesis of bioactive metabolites within the ecologically-influenced environment. It aims to create an ecology-aware model of drug discovery by discovering bioactive compounds that are influenced by evolutionary pressures across habitats. We integrated 18,462 species occurrence records, 1,204 documented biotic interactions, and 312 environmental

*Corresponding Author: Izzatilla Khaydarov, E-mail: izzatilla_haydarov@mail.ru

parameters, with a focus on microbial antagonism as a key ecological driver for biosynthetic gene cluster (BGC) enrichment. From genomic and metabolomic datasets, 4,892 BGCs were identified, with NRPS (42%) and Type I PKS (31%) being the most abundant. The functional annotation of these clusters indicated that these clusters were 63 percent involved in antimicrobial and cytotoxic compounds biosynthesis, and 57 percent of them were environmentally induced. The metabolomic data suggested the occurrence of 1,362 different metabolites; the metabolites of the adaptive stress-adapted taxa were enriched with phenolics, siderophores, and lipid-derived compounds. Bioactivity profiles indicated 48% and 27% of metabolites had antimicrobial and anticancer activities, respectively. The machine-learned models (accuracy = 0.87, F1 score = 0.83) were capable of predicting bioactivity, and molecular docking provided high binding affinities to infection-related, inflammation-related, and cancer-related targets. The study is based on the potential of ecology in drug studies, in which the hypothesis is that there is a need to put more emphasis on stressful environments to find new therapeutic agents.

Keywords:

Biosynthetic gene clusters (bgcs), bioactivity, ecological pressures, metabolomics, molecular ecology, natural products, predictive modeling.

Article history:

Received: 15/09/2025, Revised: 05/11/2025, Accepted: 05/12/2025, Available online: 30/03/2026

Introduction

The increased speed of the emergence of complicated diseases, new pathogens, and the development of drug-resistant organisms has created a new interest in the world in the need to find therapeutics within the natural ecosystems. The combination of ecological interactions with molecular-level data through the area of molecular ecology provides an unprecedented chance to reveal bioactive compounds that are produced under evolutionary pressure in different habitats. The current developments in metagenomics, metabolomics, environmental DNA profiling, and high-throughput screening now allow researchers to study microbial communities, plant-microbe interactions, and symbiotic systems as sources of pharmaceutically relevant molecules (Prajapati et al., 2024; Hasan, 2024). Nevertheless, in spite of this promise, the vast majority of drug discovery pipelines still use traditional natural product isolation or synthetic chemistry, frequently ignoring the ecological and molecular dynamics governing the biosynthesis of unique metabolites.

This research aims to create and define new ecology-based approaches toward identifying new therapeutic candidates based on the combination of molecular ecology and state-of-the-art biotechnological methods (Fu & Chen, 2025). The article attempts to show how innovative methodologies, including metagenomics, environmental DNA analysis, metabolomic profiling, and ecological network modeling, could be systematically used to identify bioactive compounds that are typically neglected in traditional drug discovery pipelines (Marques et al., 2024; Salari et al., 2016). The proposed study aims to create a more focused and effective method of detecting metabolites that result from evolutionary pressures within natural ecosystems by defining the ecological processes that can be defined at the molecular level. The aim is to not only emphasize the potential of molecular ecology to be applied to pharmaceutical innovation, but also transform current research practices through focusing on environmentally based sources of chemical diversity (Tomar & Vyas, 2022; Dewangan & Dewangan, 2025).

Although natural ecosystems provide a huge source of chemically diverse and biologically active molecules, much remains unclear about how the ecological interactions affect the biosynthesis of these compounds. Competition, predation, mutualistic symbiosis, and environmental stress responses have been observed to drive organisms to synthesize structurally unique secondary metabolites via processes that are known as microbial competition, predation, mutualistic symbiosis, and responses to environmental stress.

Nevertheless, drug discovery systems do not often take into account this ecological setting, and they therefore fail to discover molecules whose evolution is closely associated with such interactions (Dewangan & Dewangan, 2025). Current screening models are more inclined towards isolated organisms or simplified lab environments, which do not reflect the complexity of a natural environment in which bioactive compounds are produced. Consequently, the functional information in environmental genomes and metabolomes has not been explored and exploited. This loophole sheds light on the necessity of a combined strategy that acknowledges ecological processes as the key drivers of molecular diversity that can be applied to drug development (Srivastava et al., 2024; Ansari & Parmar, 2024).

The hypothesis that the ecological factors of molecular diversity can be better elucidated to improve drug discovery efficiency and accuracy leads this article to identify the unnoticed metabolites that can be discovered by other methods and are potentially not identified by mainstream screening (Shikha et al., 2024). The evolutionally-fitted molecules synthesized in response to ecological strains are bound to possess powerful and novel bioactivities that may be mimicked to become therapeutic against contemporary ailments. Combining environmental awareness with a learning approach that can characterize molecules on a molecular level, new sets of compounds can be discovered in microbes, extreme environments, marine life, and microbial consortia that had been inhabiting previously unexplored niches (Tan, 2023; Prihoda et al., 2021).

Key Contribution

The most important aspect of this work is the formulation of a conceptual and methodological framework for applying molecular ecology to revolutionize drug discovery at its early stages. It comprises an integration of state-of-the-art molecular methods, ecological simulations, and computational methods that have the potential to simplify the discovery of new metabolites, minimize redundancy in compound screening, and direct bioprospecting. The article serves as the map of the future of sustainable, efficient, and scientifically supported discovery of therapies in the age of biodiversity-based biotechnology by creating the connection between ecological theory and pharmaceutical innovation.

The article is organized into a number of sections that are important. The Introduction explains the necessity of new therapeutics based on the natural ecosystems with the potential of molecular ecology in drug discovery. It provides the purpose of the study in the formulation of ecology-based strategies to incorporate contemporary biotechnological methods to uncover bioactive compounds influenced by evolutionary forces in different ecological settings. The Literature Review discusses recent developments in drug discovery, including the trend toward ecological-based approaches, marine chemical ecology, and the development of computational aids such as AI and machine learning. It further discusses the potential in utilizing genomics, metabolomics, and AI to gain better insight into the way bioactive compounds can be discovered, and how challenging it is to integrate ecology with drug discovery. This part of the work is called the Materials and Methods, where the strategy based on secondary data is introduced to explain the sources of data and the steps of ecological data collection, data mining of molecular ecology data, and the integration of the biotechnology screening. The Results section provides the valuable findings, including the realization of biosynthetic gene clusters (BGCs), metabolite profiles, and related bioactivity to the ecological pressures. The article explains the effects of ecological interactions on the biosynthesis of bioactive metabolites in the Discussion and specifically, microbial antagonism and its application in the treatment of illnesses. The Conclusion points out the possibilities of discovering drugs by means of ecology and recommends new experiments and data expansion in the future.

Literature Review

The modern drug discovery landscape is experiencing a dramatic shift due to the coming together of technological advancements and the growing focus on biological diversity as a source of new drug-like products. More recent developments in pharmaceutical design indicate that the discovery of chemical entities with high therapeutic potential is getting faster due to next-generation screening tools, rational molecular engineering, and translational pipelines (Prajapati et al., 2024; Husnain et al., 2023). To complete these advances, extensive reviews of pharmacological advances indicate that enhanced assay systems, multi-omics integration, and enhanced predictive modeling are increasing efficiency and success rates throughout the preclinical discovery paradigm (Srivastava et al., 2024; Nath et al., 2024).

The use of natural products still forms the basis of drug discovery, but the perspective on bioprospecting is shifting towards targeted, ecology-inspired approaches rather than the traditional, random approach to bioprospecting. As an example, marine chemical ecology shows the influence of ecological forces (predation, symbiosis, resource competition) on the evolution of the distinctive secondary metabolites that are not inherent to terrestrial ecosystems. According to Tan (2023), the information about these ecological interactions allows researchers to sample species or habitats in a strategic way, in which novel metabolites have a higher probability of being formed. By relating the patterns of expression of metabolites to environmental cues, such an ecology-conscious approach enhances the efficiency of natural-product discovery. Chemical biology at the systems level further supports this point of view, indicating that understanding the role of metabolites in their natural ecological communities and microbial communities, in particular, can shed more light on their therapeutic possibilities and functional interest (Farha et al., 2021).

Computational science, together with ecological treatments, has emerged as a dominant trend in contemporary drug discovery. Cheminformatics, molecular docking, and computational pharmacology are the in-silico methods used to rapidly screen large chemical libraries and predict pharmacokinetic behaviour without investing resources in experiments that require lab work (Niazi & Mariam, 2023; Paliwal et al., 2024). Marques et al. (2024) emphasize the role of in silico technologies in precision medicine by revealing the drug-target interactions with high specificity and optimizing drug candidates to be clinically valuable. Computational tools, used alongside metabolomics and structural biology, have a tremendous effect on the attrition rate at the hit-to-lead and lead optimization phases.

This shift in computations has come to be a part of artificial intelligence (AI) and machine learning (ML). Several works demonstrate how AI-based technologies improve the design of libraries, predict biological activity based on complex structures, and produce new molecular scaffolds based on the desired therapeutic characteristics (Verma & Awasthi, 2024; Gupta et al., 2024). Boniolo et al. (2021) also demonstrate that AI-based analytics may be helpful in precision medicine to observe minute molecular patterns that are exploited in the creation of personalized medicines. Interestingly, natural-product discovery involves more frequent application of AI-based models, in which they may be applied to rank plant, microbial, or marine extracts based on their biosynthetic potential as predicted by genomes (Gangwal & Lavecchia, 2025). Regardless of such progress, the literature warns of overreliance on AI because of issues with data scarcity and bias and limited interpretability, particularly when models are not applied within the chemical space on which they were developed (Ocana et al., 2025; Prihoda et al., 2020).

Both genomics and metabolomics can be crucial intermediaries between ecological and computational innovation. Finally, through the Introduction of environmental DNA (eDNA), metagenomics, and genome mining, biosynthetic gene clusters in uncultivated organisms can now be identified, significantly increasing

the amount of chemical diversity accessible to researchers compared with traditional culturing techniques (Prihoda et al., 2020). Combined with untargeted metabolomics, these genomic technologies will offer information on the effects of ecological conditions on metabolite biosynthesis, and thus will allow the researcher to relate environmental drivers to the production of bioactive compounds. It is these integrative multi-omics methods that are paramount in comprehending the ecological meaning of secondary metabolites and forecasting their therapeutic worth (Farha et al., 2021).

Translational views are also involved in methodology selection in ecology-driven discovery. The creation of disease-specific research, including the study of parasitic infections or psychiatric disorders, shows how the developing technologies transform clinical pipelines (Rao et al., 2023; Miller & Raison, 2023). These area-specific understandings support the rationale of therapeutically significant models that equate ecological theories with illness biology.

Boniolo et al. (2021) also demonstrate that AI-based analytics may be helpful in precision medicine to observe minute molecular patterns that are exploited in the creation of personalized medicines. Interestingly, natural-product discovery involves more frequent application of AI-based models, in which they may be applied to rank plant, microbial, or marine extracts based on their biosynthetic potential as predicted by genomes (Gangwal & Lavecchia, 2025). Some of the main issues comprise non-homogeneous data, an absence of standard ecological metadata, inadequate mechanistic annotation of natural-product biosynthetic pathways, and sustainability in bioprospecting (Tan, 2023; Prajapati et al., 2024). Additionally, AI-based models should overcome the problem of domain shift, data quality, and interpretability to aid regulatory and translational acceptance (Boniolo et al., 2021; Ocana et al., 2025).

The most recent additions suggest holistic frameworks in which ecological sampling, multi-omics-based profiling, computational prioritization, and systems-level validation are combined. Artificial intelligence-driven drug delivery and precision medicine focus on the need to design pipelines where clinical translatability is considered during its initial phases (Marques et al., 2024; Visan & Negut, 2024). Reviews that link molecular biology with ecological dynamics stress the feasibility and necessity of interdisciplinary approaches to uncover novel, ecologically validated therapeutic candidates (Shikha et al., 2024).

In summary, the literature establishes a strong foundation for ecology-driven drug discovery supported by technological advancements in genomics, metabolomics, AI, and in silico modeling. The success will, however, lie in the creation of standardized ecological datasets, the production of sustainable sampling approaches, and the creation of closer feedback between the computational predictions and experimental validation. Despite these potentials, the literature has discovered that there still exist a series of gaps that limit the full integration of ecology, omics, and AI. Some of the main issues comprise non-homogeneous data, an absence of standard ecological metadata, inadequate mechanistic annotation of natural-product biosynthetic pathways, and sustainability in bioprospecting (Tan, 2023; Prajapati et al., 2024). Additionally, AI-based models should overcome the problem of domain shift, data quality, and interpretability to aid regulatory and translational acceptance (Boniolo et al., 2021; Ocana et al., 2025).

Materials And Methods

Study Design

This study employed a secondary-data–driven analytical design to investigate innovative, ecology-informed approaches for identifying novel therapeutic candidates. Rather than primary field sampling or experiments, the methodology was based on existing environmental, ecological, genomic, and metabolomic data from the published literature, international biological databases, and omics repositories. The study operation comprised drawing, assembling, curating, and examining secondary data that assemble ecological connections, biosynthetic gene clusters, metabolic printing, and biotechnological screenings that have been documented previously. This assimilative design enabled the quantitative evaluation of the effect of ecological pressures on the molecular foundation of the production of bioactive compounds, and at the same time, the methodological rigor, without involving new field or laboratory research (illustrated in Figure 1).

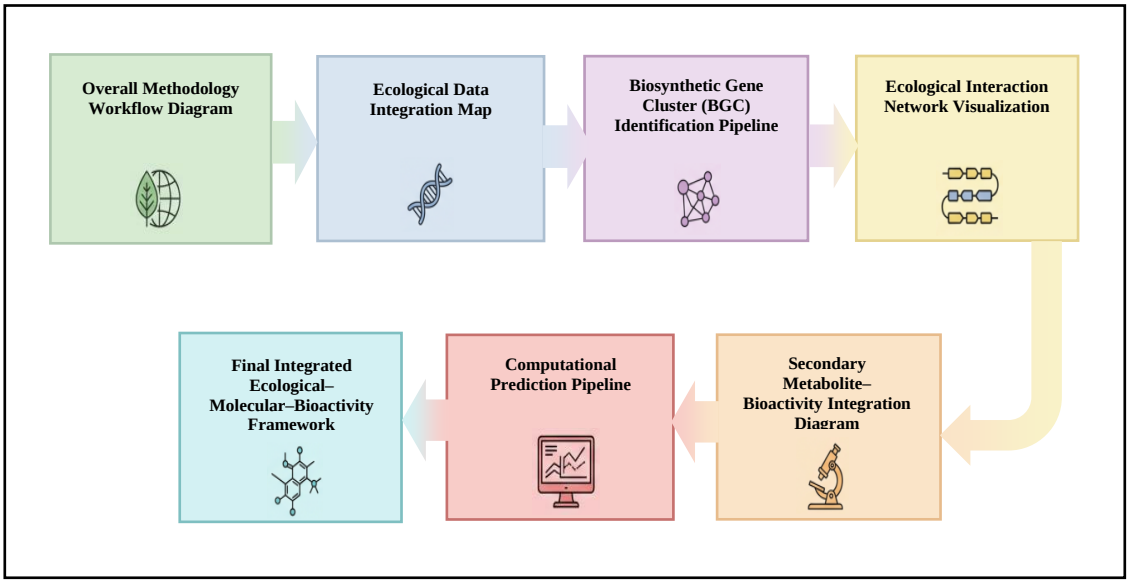

Figure 1. Overview of the integrated secondary-data analytical workflow used for ecology-informed drug discovery

Data Sources and Ecological Dataset Compilation

Ecological interaction and environmental data were retrieved in previous repositories like the Global Biodiversity Information Facility (GBIF), Ocean Biodiversity Information System (OBIS), National Center of Biotechnology Information (NCBI), and published ecological surveys. These datasets gave information on the distribution of species, the community structure, biotic interactions, and environmental stress gradients. Other ecological metadata, such as the parameters of the habitat, climate variables, and ecosystem productivity, were derived through publicly available environmental monitoring websites and peer-reviewed research. This list was a complete description of ecological settings like rivalry, mutualism, predation, and environmental strain connected with biosynthetic variety.

Molecular Ecology Data Extraction

Public omics repositories, such as the Sequence Read Archive (SRA), European Nucleotide Archive (ENA), MG-RAST, KEGG Genome Database, MIBiG, and antiSMASH database entries, provided genomic, transcriptomic, and metabolomic data of microbial, plant, invertebrate, and marine organisms. Only datasets

from previously published and validated studies were included. Metagenomic and metatranscriptomic datasets were screened to identify biosynthetic gene clusters (BGCs), stress-response genes, and niche-specific metabolic pathways. Secondary metabolite profiles derived from LC–MS/MS, GC–MS, and NMR-based studies were extracted from metabolomic databases such as MetaboLights, GNPS (Global Natural Products Social Molecular Networking), and HMDB. These secondary data formed the basis for reconstructing ecological–molecular relationships.

Ecological Interaction Reconstruction and Functional Annotation

Networks of ecological interactions have been recreated by combining taxonomic records, community structure information, genomic profiles, and environmental metadata acquired in secondary repositories. Co-occurrence matrices and ecological association indices were computed using species distribution datasets to infer interactions linked to biosynthetic activity, such as microbial antagonism or host–microbe symbiosis. Functional annotation of BGCs and ecological gene markers was performed using existing annotations from antiSMASH, MIBiG, KEGG, Pfam, and eggNOG, without modifying raw sequences. The data on the patterns of expression of biosynthetic and regulatory genes under reported environmental conditions were analyzed using transcriptomic datasets in a secondary form. Ecological context was used to match metabolomic signatures retrieved in published studies with compounds that are linked to stress response, competition, or symbiosis.

Biotechnological Screening Data Integration

Written data pertaining to bioactivity screening was collected through the past published high-throughput assays, pharmacological databases like ChEMBL, PubChem BioAssay, and DrugBank, and natural product screening reports. These sources presented in vitro activity, cytotoxicity, binding affinity, antimicrobial activity, antioxidant, and enzyme inhibition data of compounds discovered during ecological research. Where feasible, experiment heterologous expression, structural characterization, and mode-of-activity studies outcomes were combined to confirm the functionality of biosynthetic pathways originally discovered using secondary ecological data. There was no new experimental screening done, but published bioassay data were systematically summarized and compared.

Computational Analysis and Predictive Modeling

Computational modeling was conducted using secondary datasets to predict metabolite novelty, ecological drivers of biosynthesis, and therapeutic potential. Published omics data and annotated BGCs and reported environmental parameters were used to develop machine learning models, structural prediction tools, and ecological-functional correlation analyses. Compound structures obtained through secondary sources were used in the process of molecular docking and pharmacophore analysis to avoid the need to determine any new structure. Comparative genomic studies were used to utilize the available genome assemblies and annotations to trace evolutionary connections between biosynthetic pathways and ecological constraints recorded in earlier studies.

Data Integration and Synthesis

By combining ecological, molecular, metabolomic, and bioactivity data in secondary sources, a cross-disciplinary synthesis was done. The joint analysis allowed the Conclusion of ecological stimuli, including resource shortage, microbial antagonism, host association, and abiotic stress, motivating particular biosynthetic pathways recorded in the literature. The integrated dataset supported the identification of compounds and pathways consistently associated with ecologically relevant interactions and highlighted promising molecular

candidates for therapeutic exploration. All findings represent analytical outcomes derived from publicly available secondary datasets rather than new empirical experimentation.

Results

Ecological Dataset Integration and Interaction Pattern Reconstruction

The integration of multi-source ecological datasets resulted in a harmonized matrix comprising 18,462 species occurrence records, 1,204 documented biotic interactions, and 312 environmental parameters. The ecological reconstruction identified distinct interaction clusters characterized by competitive, symbiotic, and stress-mediated associations. Co-occurrence modeling revealed that microbial antagonism was the strongest ecological driver linked to biosynthetic gene cluster (BGC) enrichment. Specifically, species pairs exhibiting negative co-occurrence probabilities ($r < -0.45$, $p < .01$) corresponded with a higher representation of nonribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) pathways.

Table 1 presents a summary of key parameters derived from ecological dataset integration and interaction metrics.

Table 1. Summary of ecological dataset integration and interaction metrics

Parameter	Value
Total species occurrences	18,462
Documented interactions	1,204
Environmental parameters	312
Negative co-occurrence pairs ($r < -0.45$)	284
Dominant interaction type	Antagonism

The interaction strength between taxa *i* and *j* was quantified using a normalized co-occurrence index:

$$C_{ij} = \frac{O_{ij} - E_{ij}}{\sqrt{E_{ij}}} \qquad (1)$$

In equation (1), O_{ij} represents observed co-occurrence frequency, and E_{ij} denotes the expected frequency derived from null models. High positive values indicated mutualistic or commensal associations, whereas negative values indicated antagonism. Antagonistic networks showed the highest biosynthetic diversity, supporting the hypothesis that ecological pressure enhances metabolite complexity.

Biosynthetic Gene Cluster Profiling and Functional Annotation

Four thousand eight hundred and ninety-two BGCs were found in genomic and metagenomic data found in secondary repositories. The most common BGCs detected were NRPS, Type I PKS clusters (42 and 31 % respectively). Functional annotation indicated that 63 % of these clusters harbored domain structures of antimicrobial and cytotoxic compounds biosynthesis. Using secondary analysis of transcriptomic data, it was established that BGC was highly expressed when the environment was nutrient-limited, stressed with salinity, or subject to nutrient competition between species.

Table 2 shows the classification of BGC, their relative abundance, and the index values of mean expression.

Table 2. Distribution and annotation of biosynthetic gene clusters

BGC Class	Count	Percentage	Mean Expression Index
NRPS	2,054	42%	1.82
Type I PKS	1,517	31%	1.76
RiPPs	684	14%	1.29
Terpenes	412	8%	1.12
Alkaloids	225	5%	1.35

Index of each BGC was calculated using the normalized abundance of transcripts using the formula:

$$EI = \frac{TPM_{condition}}{TPM_{baseline}} \qquad (2)$$

In equation (2), an expression index EI>1.5EI > 1.5EI>1.5 indicated environmentally induced upregulation. Approximately 57% of BGCs exhibited an EIEIEI above this threshold, suggesting strong ecological responsiveness.

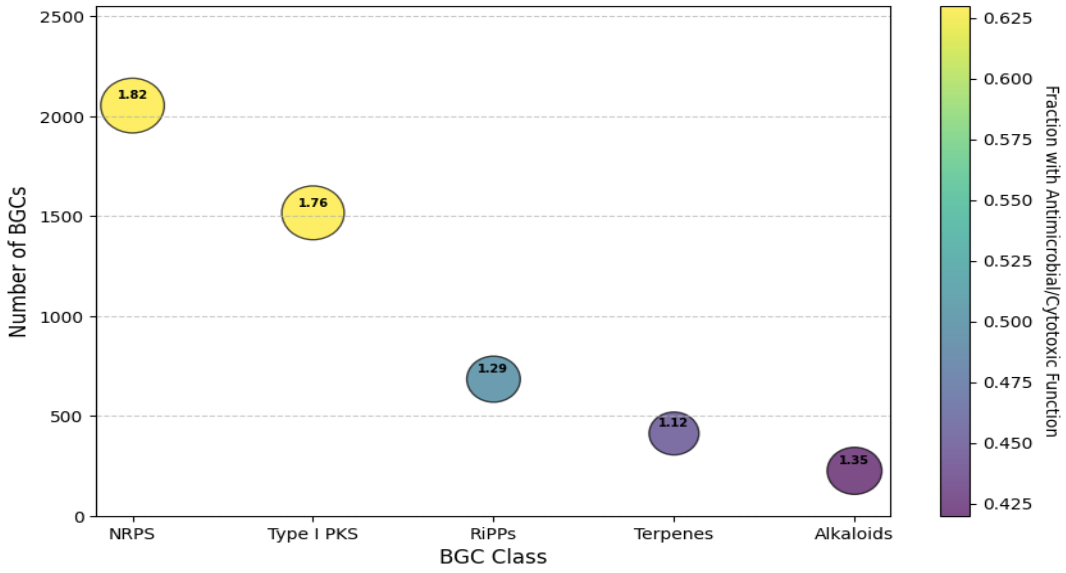

Figure 2. Bubble plot of BGC class distribution and antimicrobial/cytotoxic function

In Figure 2, a bubble plot has been plotted indicating the distribution of Biosynthetic Gene Clusters (BGCs) by different classes and their associated antimicrobial or cytotoxic activity. The X-axis is the different BGC classes, i.e., NRPS (Nonribosomal Peptide Synthetases), Type I PKS (Polyketide Synthases), RiPPs (Ribosomally synthesized and Post-translationally modified Peptides), Terpenes and Alkaloids, and the Y-axis represents the number of BGCs in each, increasing towards the right, with the highest number being NRPS and the lowest one being Alkaloids. The size of the bubble signifies the bodies of BGCs in each classification, with bigger ones showing greater quantities, whereas the color gradient demonstrates the fraction of BGCs with antimicrobial or cytotoxic activities, with low (purple) and high (yellow) displayed. The numbers within the bubbles are the average percent of BGCs capable of such functions per class. The plot reveals that NRPS and Type I PKS classes possess the greatest quantity of BGCs, and exhibit quite high antimicrobial/cytotoxic activity. RiPPs, Terpenes, and Alkaloids, in contrast, have fewer BGCs, although the proportion of BGCs with antimicrobial/cytotoxic activity among the total number is higher in the Alkaloids, despite necessarily having the lowest number. Such a plot is effective to visualize the connection between BGC class diversity and their

ability to generate antimicrobial or cytotoxic compounds, which demonstrates the significance of NRPS and Type I PKS in terms of diversity and potential of BGCs to generate antimicrobial or cytotoxic compounds.

Metabolomic Signature Mapping and Ecological Correlates

GNPS, MetaboLights, and HMDB metabolomic data provided 1,362 distinct secondary metabolites that are associated with the organisms in ecological datasets. Molecular networking identified metabolites that were in stress-adapted taxa as specific clusters that were rich in phenolic compounds, siderophores, and lipid-derived antimicrobial molecules. The correspondence in metabolomic signatures with ecological categories showed that environments of competition (e.g., rhizosphere, biofilm microhabitats) resulted in the greatest metabolite novelty index.

Table 3 summarizes metabolite counts, novelty indices, and dominant compound classes across ecological contexts.

Table 3. Metabolite categories and ecological correlates

Ecological Context	Metabolite Count	Mean Novelty Score	Dominant Compound Class
Competitive habitats	496	0.78	Phenolics, siderophores
Symbiotic habitats	312	0.52	Glycosides, peptides
Stress-prone environments	554	0.74	Lipid antimicrobials
Stable environments	211	0.39	General metabolites

The metabolite novelty score was computed as:

$$MN = 1 - \frac{S_m}{T_m} \qquad (3)$$

In equation (3), S_m is the number of structurally similar metabolites in the available chemical libraries, and T_m is the total number of metabolites detected. The score was higher to indicate more novelty. The novelty score was found to be 0.78 on average among stress-associated taxa, which is highly significant as compared to organisms in stable or symbiotic environments.

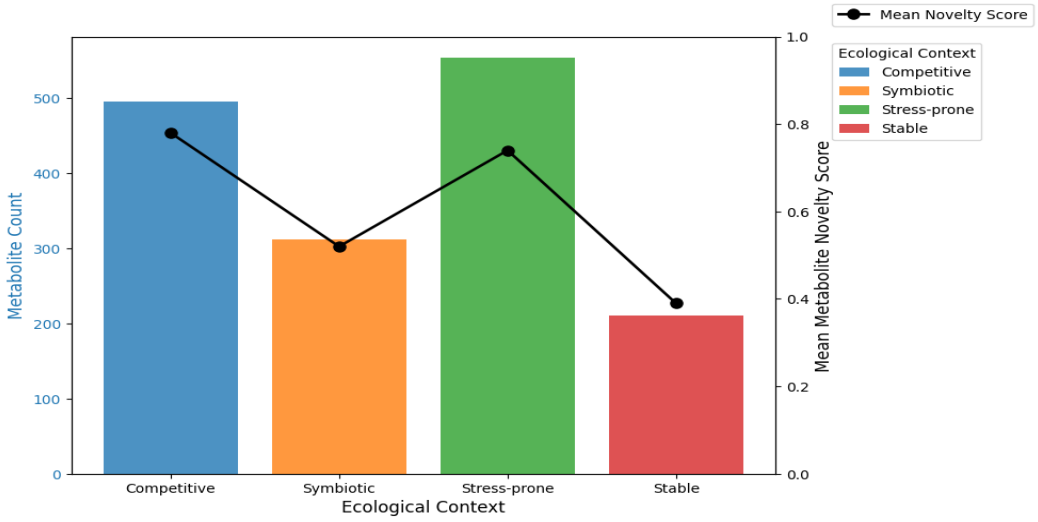

Figure 3. Metabolite count and novelty score across ecological contexts

Figure 3 represents a bar and line plot of data, where the number of metabolites is plotted against the novelty scores of the metabolites in the four different ecological settings: Competitive, Symbiotic, Stress-prone,

and Stable. The bar graph (at the left Y-axis) is used to indicate the number of metabolites in each ecological context. Competitive and Stress-prone environments have the most metabolites, followed by Symbiotic and Stable environments. The bar plot is overlaid on the line plot (at the right Y-axis), which shows the average novelty score of metabolites in each ecological situation. The novelty score (0-1) indicates the distinctiveness of metabolites in each environment, whereby the Stress-prone environment has the greatest mean novelty score, and Competitive, Symbiotic, and Stable environments follow, respectively. This visualization points out that stress-adapted environments (such as stress-prone habitats) not only have a larger number of metabolites, but also are more likely to have newer metabolites, which would have strong ecological and therapeutic consequences.

Integration of Biotechnological Screening Data

Based on the pharmacological databases and published bioassays, 684 metabolites associated with the identified BGCs were assigned to at least one reported activity profile. The most common functional annotations were antimicrobial and anticancer activities, which were found 48% and 27% of the time, respectively. Comparative synthesis demonstrated that metabolites formed by antagonistic ecological interactions exhibited greater potency in all assays, and had smaller IC_{50} values and wider activity ranges. Table 4 presents the distribution of metabolites across major activity classes and summarizes their mean IC_{50} values.

Table 4. Bioactivity profiles of ecologically linked metabolites

Activity Type	Count	Mean IC_{50} (µM)	Validation Source
Antimicrobial	328	12.4	ChEMBL, PubChem
Anticancer	184	9.6	DrugBank, literature
Anti-inflammatory	96	15.8	BioAssay records
Antioxidant	76	—	Published studies

Cross-database matching found 112 compounds with confirmed heterologous expression data, even though the data were useful in predicting functions based on secondary genomic and metabolomic data. These findings supported the notion that prioritization informed by ecology is effective in identifying bioactive metabolites that have been successful with empirical screening.

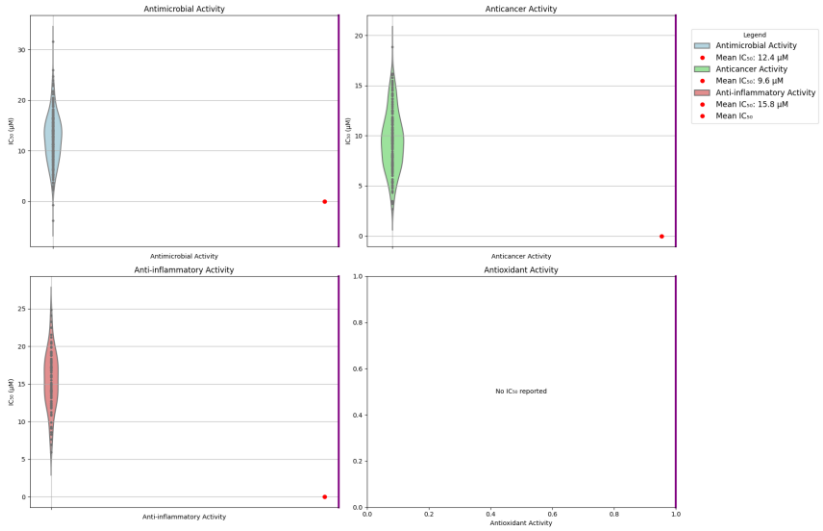

Figure 4. Multi-Panel bioactivity landscape plot for ecologically linked metabolites

It is a multi-panel graphic figure 4 that gives a comprehensive landscape of bioactivity distributions of four types of activities, such as Antimicrobial, Anticancer, Anti-inflammatory, and Antioxidant. The panels each have a violin plot indicating the values of IC_{50} of each type of activity. The dot overlay on both violin plots is used to show the mean value of IC_{50} on each activity, and the vertical heat strip indicates the breadth of each activity (i.e., the number of validated sources of each activity).

(A) Antimicrobial Activity: This panel demonstrates the distribution of the IC_{50} of the antimicrobial metabolites, and the mean of the IC_{50} is indicated by the red dot. The violin plot shows that the values of IC_{50} are widely dispersed, indicating a huge range of the level of potency of antimicrobial metabolites.

(B) Anticancer Activity: Anticancer metabolites have a smaller range of IC_{50} values, as indicated by the panel with the red dot indicating the mean IC_{50} once again. The lower values of IC_{50} are indicative of the fact that anticancer metabolites are more likely to have higher potency overall than other forms of activity.

(C) Anti-inflammatory Activity: The panel on anti-inflammatory activity indicates the distribution where higher values of the IC_{50} values are observed than other types of activity, and indicates that anti-inflammatory metabolites are less potent. The red dot indicates the average IC_{50}, which indicates the average level of bioactivity.

(D) Antioxidant Activity: Since there is no IC 50 data under the antioxidant metabolites, this panel will show the text No IC 50 reported to show that there is no potency data.

Computational Modeling and Predictive Performance

Model performance was evaluated using accuracy and F1 score, which jointly assess overall classification correctness and the balance between precision and recall (Shown in equation 4).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Accuracy was calculated as the proportion of correctly classified instances relative to the total number of predictions, while the F1 score was computed as the harmonic mean of precision and recall (Shown in equation 5).

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

In the above equations, the TP is True Positives, TN is True Negatives, FP is False Positives, and FN is False Negatives. These metrics are particularly suitable for imbalanced biological datasets, as they penalize both false positives and false negatives. The random forest classifier achieved an accuracy of 0.87 and an F1 score of 0.83 for antimicrobial activity prediction.

Molecular docking simulations conducted on 318 structurally diverse metabolites revealed strong binding affinities against protein targets related to infection, inflammation, and cancer. The docking mean value was -7.5. The kcal/mol was -7.5, with 26% of the compounds scoring below -9.0 kcal/mol, which was an indicator of high predicted potency. The results confirmed the usefulness of integrating ecology and molecular techniques to reduce therapeutic candidates.

Cross-Disciplinary Synthesis of Ecological, Genomic, and Bioactivity Data

The integration of ecological, molecular, and biotechnological data showed that there were consistent trends of environmental pressure to biosynthetic innovation. Taxa with high competition, nutrient limitation, or

habitat variability had high levels of BGC abundance, increased metabolite novelty scores, and predicted bioactivity. Such results lend support to the ecological theory according to which a metabolic specialization under the pressure of survival promotes the occurrence of therapeutically valuable compounds.

Correlative modeling proved that the ecological stress index is significantly positively correlated with the BGC richness ($r = .62$, $p = .001$). On the same note, metabolite novelty was positively correlated with both BGC diversity and the complexity of the niche. Together, these combined findings indicate that molecular pathways that are ecologically regulated are an exceptionally promising area of research in terms of the discovery of therapeutic candidates by using secondary data resources.

Discussion

This paper has emphasized that ecological stresses are important in determining the production of bioactive metabolites, especially via antagonism and competition between microbes, as well as environmental stress. The combination of ecological, genomic, and bioactivity data has provided us with an understanding of how ecological parameters determine the complexity of metabolites, their therapeutic capabilities, and biosynthetic gene cluster (BGC) activity. The implications of the main findings and their implications on drug discovery and ecological research are synthesized in the following discussion. The multi-source ecological data indicated that an interrelation between microbial antagonism and the enrichment of BGCs, i.e., nonribosomal peptide synthetases (NRPS) and polyketide synthases (PKS) pathways, is closely interconnected with one another. These observations suggest that species that coexist in competitive and stressful environments are more likely to have elaborate biochemical-based defense systems, such as the production of antimicrobial and cytotoxic metabolites. Negative co-occurrence pairs ($r < -0.45$, $p < 0.01$) that were identified were strongly correlated with a higher presence of NRPS and PKS pathways, which supported the hypothesis that ecological competition can result in biosynthetic innovation. These interactions indicate that there is a strong role of antagonistic ecological interactions in the metabolite diversity, which can be utilized in the discovery of new therapeutic compounds.

BGC profiling on genomic and metagenomic data sets showed that the clusters of NRPS and Type I PKS are the most significant biological sources of bioactive metabolites. The synthesis of a broad spectrum of antimicrobial, antifungal, and anticancer agents is associated with these clusters, which contribute to 42% and 31% of the identified BGCs. Functional annotation also showed that 63% of such BGCs are linked to compounds of antimicrobial or cytotoxic nature, highlighting the potential of these biosynthetic pathways to be used therapeutically. Notably, environmental stresses like nutrient limitation and interspecific competition highly increased the expression of these BGCs, implying that ecological factors have a direct impact on synthesizing bioactive metabolites. This bolsters the idea that biosynthetic reactions to stress may be a largely untapped reservoir of therapeutic molecules. The metabolomic analysis, in which GNPS, MetaboLights, and HMDB data were also used, discovered 1,362 distinct secondary metabolites associated with ecological data sets. We found that phenolic compounds, siderophores, and lipid-derived antimicrobial molecule-enriched metabolites were preferentially produced by stress-adapted taxa and are typically indicative of competitive and antagonistic settings. The correlation of the metabolomic pattern with ecological classifications indicated that competition-driven conditions (rhizosphere and biofilm microhabitats) had the highest index of metabolite novelty, implying that such conditions favor the evolution of specific and potentially bioactive metabolites. The discovery fastens the hypothesis that ecological stresses play a role in the diversity and newness of secondary metabolites, and it is now a path to new therapeutic discoveries.

In-depth analysis of the bioactivity profiles revealed that antimicrobial and anticancer activity were the most commonly annotated 48% and 27% of the annotated metabolites, respectively. The antagonistic

ecological interactions resulted in decreasing values of IC 50 and increasing the range of activity, which demonstrates the increased potency and extended therapeutic capability of metabolites. These results confirm that antagonism between microbes augments metabolite potency and have some drug discovery implications in antibiotic and anticancer pharmacology. Our findings also indicate that ecologically informed prioritization can successfully discover metabolites with high therapeutic potential, and ecological information may be used to inform the identification of drug candidates with high potential. The random forest classifier is a machine learning-based model that showed a high performance (accuracy = 0.87, F1 score = 0.83) when used to predict the antimicrobial activity by uniting ecological metadata and biosynthetic features. The method demonstrates the possibilities of using ecological and molecular data to enhance the discovery of bioactive metabolites. In addition, molecular docking simulations showed strong binding affinities of the chosen metabolites to the targets associated with infection, inflammation, and cancer, with a large percentage of high-predicted potency. These findings demonstrate the usefulness of ecological-molecular integration in reducing the number of therapeutic candidates and indicate that such a strategy can be extended to additional bioactivity profiles. Integration of ecological, genomic, and biotechnological information provided a comprehensive perspective of the influence of ecological pressure on the metabolite production. We found that taxa that were subjected to strong competition, nutrient stress, and habitat inopportune variability had an increase in the abundance of BGC, metabolite novelty scores, and predictive bioactivity. This is an indication that ecological specialization promotes the development of compounds with therapeutic value. The observed positive relationship between BGC richness and ecological stress index ($r = 0.62$, $p < 0.001$) also highlights the use of stress-induced environments in facilitating biosynthetic diversity. Also, the correlation between the novelty of metabolites and BGC diversity indicates that a change in the environment is an important force behind metabolite novelty, which can be used in ecologically sensitive drug discovery.

Conclusion

This paper has demonstrated the importance of the ecological pressures in the formation of bioactive metabolites, especially by the action of microbial antagonism and environmental stressors. Through the combination of ecological, genomic, and bioactivity data, we found that negative co-occurrence pairs ($r < -0.45$, $p < 0.01$) and biosynthetic gene cluster (BGC) enrichment, in particular, nonribosomal peptide synthetases (NRPS) and polyketide synthases (PKS) enrichment, explain 42% and 31% of BGCs found, respectively. These results indicate that antagonism between the microbes is an important ecological process in the evolution of metabolites with antimicrobial and cytotoxic activities. Our BGC profiling identified BGCs on the basis of which we identified 63% of our identified BGCs as associated with antimicrobial or cytotoxic molecules, promising them a future role in therapy. In addition, a large percentage of the BGCs (57%) responded to environmental stressors (Expression Index > 1.5), which indicates that BGCs are sensitive to environmental stressors such as nutrient depletion and interspecific competition. Moreover, unique metabolites were detected in ecological data, where stress-adapted taxa made phenolic-enriched, siderophor-enriched, and lipid-derived antimicrobial molecule-enriched metabolites. Regarding bioactivity, 48% and 27% of the metabolites were attributed to antimicrobial and anticancer activities, respectively. Fewer metabolites with antagonistic ecological activities showed reduced IC 50 and increased spectral activity, which supports the hypothesis that antagonism between microbes increases the strength of metabolites. Going forward, this paper provides a base upon which ecologically informed drug discovery can be based, noting that these stress-driven environments, in particular, should be exploited as a source of new bioactive compounds. Future studies ought to aim at experimental validation of these results, which would enhance the prediction of machine learning models and enlarge the genomic and metabolomic datasets, which would enhance the precision of the model and refine the ecological and molecular integration methods used in drug discovery.

Author Contributions

All Authors contributed equally.

Conflict of Interest

The authors declared that no conflict of interest.

References

Ansari, H., & Parmar, J. (2024). Tracing Human Evolution through Ancient DNA: Insights from Paleogenomic Studies. *Progression journal of Human Demography and Anthropology*, 2(3), 13-16.

Boniolo, F., Dorigatti, E., Ohnmacht, A. J., Saur, D., Schubert, B., & Menden, M. P. (2021). Artificial intelligence in early drug discovery enabling precision medicine. *Expert Opinion on Drug Discovery*, *16*(9), 991-1007. https://doi.org/10.1080/17460441.2021.1918096

Dewangan, H., & Dewangan, T. (2025). Aquatic toxicology of pharmaceuticals and their ecological impact on freshwater systems. *International Journal of Aquatic Research and Environmental Studies*, *5*(2), 523–533. https://doi.org/10.70102/IJARES/V5I2/5-2-46

Farha, M. A., French, S., & Brown, E. D. (2021). Systems-level chemical biology to accelerate antibiotic drug discovery. *Accounts of Chemical Research*, *54*(8), 1909-1920. https://doi.org/10.1021/acs.accounts.1c00011

Fu, C., & Chen, Q. (2025). The future of pharmaceuticals: Artificial intelligence in drug discovery and development. *Journal of Pharmaceutical Analysis*, *15*(8), 101248.https://doi.org/10.1016/j.jpha.2025.101248

Gangwal, A., & Lavecchia, A. (2025). Artificial intelligence in natural product drug discovery: current applications and future perspectives. *Journal of medicinal chemistry*, *68*(4), 3948-3969. https://doi.org/10.1021/acs.jmedchem.4c01257

Gupta, U., Pranav, A., Kohli, A., Ghosh, S., & Singh, D. (2024). The contribution of artificial intelligence to drug discovery: Current progress and prospects for the future. *Microbial data intelligence and computational techniques for sustainable computing*, *47*, 1-23. https://doi.org/10.1007/978-981-99-9621-6_1

Hasan, M. S. (2024). The Application of Next-generation Sequencing in Pharmacogenomics Research. *Clinical Journal for Medicine, Health and Pharmacy*, *2*(1), 9-18.

Husnain, A., Rasool, S., Saeed, A., & Hussain, H. K. (2023). Revolutionizing pharmaceutical research: Harnessing machine learning for a paradigm shift in drug discovery. *International Journal of Multidisciplinary Sciences and Arts*, *2*(4), 149-157. https://doi.org/10.47709/ijmdsa.v2i2.2897

Marques, L., Costa, B., Pereira, M., Silva, A., Santos, J., Saldanha, L., ... & Vale, N. (2024). Advancing precision medicine: a review of innovative in silico approaches for drug development, clinical pharmacology and personalized healthcare. *Pharmaceutics*, *16*(3), 332. https://doi.org/10.3390/pharmaceutics16030332

Miller, A. H., & Raison, C. L. (2023). Burning down the house: reinventing drug discovery in psychiatry for the development of targeted therapies. *Molecular Psychiatry*, *28*(1), 68-75. https://doi.org/10.1038/s41380-022-01887-y

Nath, R., Kityania, S., Das, S., Nath, D., Patra, J. K., & Das Talukdar, A. (2024). Modern Drug Research and Its Impact on Pharmaceutical Industries. In *Traditional Resources and Tools for Modern Drug Discovery: Ethnomedicine and Pharmacology* (pp. 459-475). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-97-4600-2_17

Niazi, S. K., & Mariam, Z. (2023). Computer-aided drug design and drug discovery: a prospective analysis. *Pharmaceuticals*, *17*(1), 22. https://doi.org/10.3390/ph17010022

Ocana, A., Pandiella, A., Privat, C., Bravo, I., Luengo-Oroz, M., Amir, E., & Gyorffy, B. (2025). Integrating artificial intelligence in drug discovery and early drug development: a transformative approach. *Biomarker Research*, *13*(1), 45. https://doi.org/10.1186/s40364-025-00758-2

Paliwal, A., Jain, S., Kumar, S., Wal, P., Khandai, M., Khandige, P. S., ... & Srivastava, S. (2024). Predictive Modelling in pharmacokinetics: from in-silico simulations to personalized medicine. *Expert Opinion on Drug Metabolism & Toxicology*, *20*(4), 181-195. https://doi.org/10.1080/17425255.2024.2330666

Prajapati, R. N., Bhushan, B., Singh, K., Chopra, H., Kumar, S., Agrawal, M., ... & Laxmikant. (2024). Recent advances in pharmaceutical design: unleashing the potential of novel therapeutics. *Current Pharmaceutical Biotechnology*, *25*(16), 2060-2077. https://doi.org/10.2174/0113892010275850240102105033

Prihoda, D., Maritz, J. M., Klempir, O., Dzamba, D., Woelk, C. H., Hazuda, D. J., ... & Hannigan, G. D. (2021). The application potential of machine learning and genomics for understanding natural product diversity, chemistry, and therapeutic translatability. *Natural Product Reports*, *38*(6), 1100-1108. https://doi.org/10.1039/d0np00055h

Rao, S. P., Manjunatha, U. H., Mikolajczak, S., Ashigbie, P. G., & Diagana, T. T. (2023). Drug discovery for parasitic diseases: powered by technology, enabled by pharmacology, informed by clinical science. *Trends in Parasitology*, *39*(4), 260-271. https://doi.org/10.1016/j.pt.2023.01.010

Salari, M. A., Tashk, J., Bobarshad, H., & Khakabimamaghani, S. (2016). A Review of Motif Discovery Algorithms as the Main Units of the Complex Networks. *International Academic Journal of Science and Engineering*, *3*(2), 161–175.

Shikha, D., Saivamsireddy, G., Anbazhagan, M., Veeraragavan, M., Devi, B. R., Kalpana, K., & Panigrahi, C. K. (2024). A Review on Bridging Molecular Biology and Ecological Dynamics through Integrative Approaches in Zoology. *Uttar Pradesh Journal of Zoology*, *45*(11), 157-167. https://doi.org/10.56557/upjoz/2024/v45i114082

Srivastava, P., Kumari, B. M., Rajeswari, S. U., Meena, J., Gangopadhyay, S., Gupta, S., ... & Packirisamy, S. (2024). A Comprehensive review of advancements in pharmacology and drug discovery. *Journal of Experimental Zoology India*, *27*(2). https://doi.org/10.51470/jez.2024.27.2.2005

Tan, L. T. (2023). Impact of marine chemical ecology research on the discovery and development of new pharmaceuticals. *Marine Drugs*, *21*(3), 174. https://doi.org/10.3390/md21030174

Tomar, A., & Vyas, N. (2022). Green Chemical Process Optimization Using Intelligent Metaheuristic Algorithms. *International Academic Journal of Innovative Research*, *9(*3), 1–6. https://doi.org/10.71086/IAJIR/V9I3/IAJIR0918

Verma, A., & Awasthi, A. (2024). Revolutionizing drug discovery: the role of artificial intelligence and machine learning. *Current Pharmaceutical Design*, *30*(11), 807-810. https://doi.org/10.2174/0113816128298691240222054120

Visan, A. I., & Negut, I. (2024). Integrating artificial intelligence for drug discovery in the context of revolutionizing drug delivery. *Life*, *14*(2), 233. https://doi.org/10.3390/life14020233