Review

Exploring Bioinformatics Opportunities for Identification and Study of Medicinal Plants: A Comprehensive Review

gopal gupta *, Shan Khatoon , Rajnish singh , Shashank gupta , Devendra rai *

Posted Date: 13 May 2024

doi: 10.20944/preprints202405.0752.v1

Keywords: Bioinformation; medicinal plant; traditional medicine; medicine; biological activities

 doi:10.20944/preprints202405.0752.v1

Review

Exploring Bioinformatics Opportunities for Identification and Study of Medicinal Plants: A Comprehensive Review

Shan Ara Khatoon [1], Rajnish Singh [2], Shashank Kumar Gupta [3], Devendra Kumar Rai [4,*] and Gopal Krishna Gupta [5,*]

[1] Dept. of Zoology, Late Chandrashekhar ji Purva-Pradhanmantri Smarak Mahavidhyalaya Sevarayi, Ghazipur
[2] Dept. of Agriculture Entomology, Late Chandrashekhar ji Purva-Pradhanmantri Smarak Mahavidhyalaya Sevarayi, Ghazipur
[3] Dept. of Pharmacy, Dr. Ram Manohar Lohia College of Pharmacy, Ghaziabad, U.P.
[4] Dept. of Botany, Late Chandrashekhar ji Purva-Pradhanmantri Smarak Mahavidhyalaya Sevarayi, Ghazipur
[5] Depatment of Physics, Jannayak Chandrashekhar University, Ballia
* Correspondence: devendrarai594@gmail.com, gopal.krishna.gupta.786@gmail.com

Abstract: Medicinal plants have been a cornerstone of traditional medicine systems for centuries, offering a vast array of bioactive compounds with therapeutic potential. However, the identification, characterization, and elucidation of the pharmacological properties of these plants have often been challenging due to their complex chemical composition and diverse biological activities. Bioinformatics, with its interdisciplinary approach combining biology, computer science, and information technology, presents a powerful toolkit for the systematic study of medicinal plants. This review provides an overview of bioinformatics methodologies and resources that can facilitate the identification, phytochemical analysis, pharmacological investigation, and conservation of medicinal plants. It highlights the integration of omics technologies, computational algorithms, database resources, and molecular modeling techniques in advancing our understanding of medicinal plants' therapeutic potential and their applications in drug discovery and development.

Keywords: bioinformatics; medicinal plants; omics technologies; drug discovery; phytochemical analysis; pharmacological investigation; conservation; sustainable utilization

1. Introduction

Medicinal plants have played a pivotal role in human healthcare for millennia, serving as primary sources of remedies for various ailments across cultures and civilizations [1]. From the ancient practices of Ayurveda and Traditional Chinese Medicine to contemporary pharmacology, medicinal plants continue to offer a rich reservoir of bioactive compounds with therapeutic potential [2]. However, the comprehensive exploration and utilization of these botanical resources pose significant challenges due to their diverse chemical composition, complex biological activities, and dwindling natural habitats [3].

In recent decades, the emergence of bioinformatics has provided a transformative framework for the systematic study of medicinal plants, offering innovative tools and methodologies to unravel their complexities and harness their therapeutic benefits [4]. Bioinformatics, an interdisciplinary field at the nexus of biology, computer science, and information technology, leverages computational algorithms, high-throughput technologies, and data integration strategies to analyze biological data, elucidate molecular mechanisms, and accelerate drug discovery processes [5].

This comprehensive review aims to explore the multifaceted applications of bioinformatics in the identification and study of medicinal plants, spanning from taxonomic classification and phytochemical analysis to pharmacological investigation and conservation efforts [6]. By integrating

 doi:10.20944/preprints202405.0752.v1

omics technologies, computational modeling, and database resources, bioinformatics empowers researchers to decipher the intricate molecular networks underlying the medicinal properties of plants and expedite the discovery of novel therapeutics [7].

Through a synthesis of current research findings and technological advancements, this review elucidates the diverse bioinformatics approaches and resources available for studying medicinal plants, highlighting their potential impact on healthcare [8], drug development, and biodiversity conservation [9]. By bridging traditional knowledge with modern computational tools, bioinformatics offers a promising pathway towards unlocking the full therapeutic potential of medicinal plants and addressing global health challenges in an era of precision medicine and sustainable healthcare practices [10].

2. Bioinformatics Approaches for Medicinal Plant Identification

Bioinformatics offers a plethora of approaches for the accurate identification and classification of medicinal plants, addressing critical challenges such as species authentication, taxonomic classification, and detection of adulterants or contaminants in herbal products [11]. These approaches leverage molecular data, computational algorithms, and database resources to streamline the identification process and ensure the integrity and quality of medicinal plant materials [12]. Below are some key bioinformatics approaches used for medicinal plant identification [13]:

1. DNA Barcoding: DNA barcoding involves the amplification and sequencing of standardized DNA regions to identify and differentiate species [14]. For medicinal plants, commonly used barcoding regions include the ribosomal DNA (rDNA) internal transcribed spacer (ITS) regions, the chloroplast trnL-trnF intergenic spacer [15], and the matK gene. Bioinformatics tools such as BLAST (Basic Local Alignment Search Tool) and BOLD (Barcode of Life Data Systems) facilitate the comparison of barcode sequences against reference databases to determine species identity [16].

2. Molecular Phylogenetics: Molecular phylogenetic analysis employs computational algorithms to reconstruct evolutionary relationships among different plant taxa based on DNA sequence data [17]. Phylogenetic trees generated using methods like maximum likelihood or Bayesian inference can elucidate the evolutionary history and taxonomic classification of medicinal plants [18]. Tools such as MEGA (Molecular Evolutionary Genetics Analysis) and PhyML (Phylogenetic Maximum Likelihood) assist in phylogenetic tree construction and visualization [18].

3. Machine Learning and Pattern Recognition: Machine learning algorithms, such as support vector machines (SVM), random forests, and neural networks, can be trained on molecular data to classify medicinal plant species or detect adulterants in herbal products [19]. These algorithms learn patterns from training data and make predictions on new samples based on their molecular profiles [20]. Feature selection techniques and cross-validation methods are employed to optimize model performance and minimize overfitting [21].

4. Metabarcoding and Metagenomics: Metabarcoding involves the high-throughput sequencing of environmental DNA (eDNA) extracted from complex samples to identify the diversity of plant species present [22]. In the context of medicinal plants, metabarcoding can be applied to assess the composition of herbal products, identify contaminants or allergens, and monitor biodiversity in natural habitats [23]. Bioinformatics pipelines such as QIIME (Quantitative Insights Into Microbial Ecology) and mothur are utilized for sequence quality control, taxonomic assignment, and diversity analysis [24].

5. Database Resources and Repositories: Publicly available databases and repositories play a crucial role in medicinal plant identification by providing curated reference sequences, taxonomic information, and metadata for molecular analysis [25]. Databases such as GenBank, NCBI Taxonomy, and Medicinal Plant Genomics Resource (MPGR) house extensive genomic and transcriptomic data for a wide range of plant species [26]. These resources serve as valuable reference libraries for species authentication and phylogenetic analysis [27].

 doi:10.20944/preprints202405.0752.v1

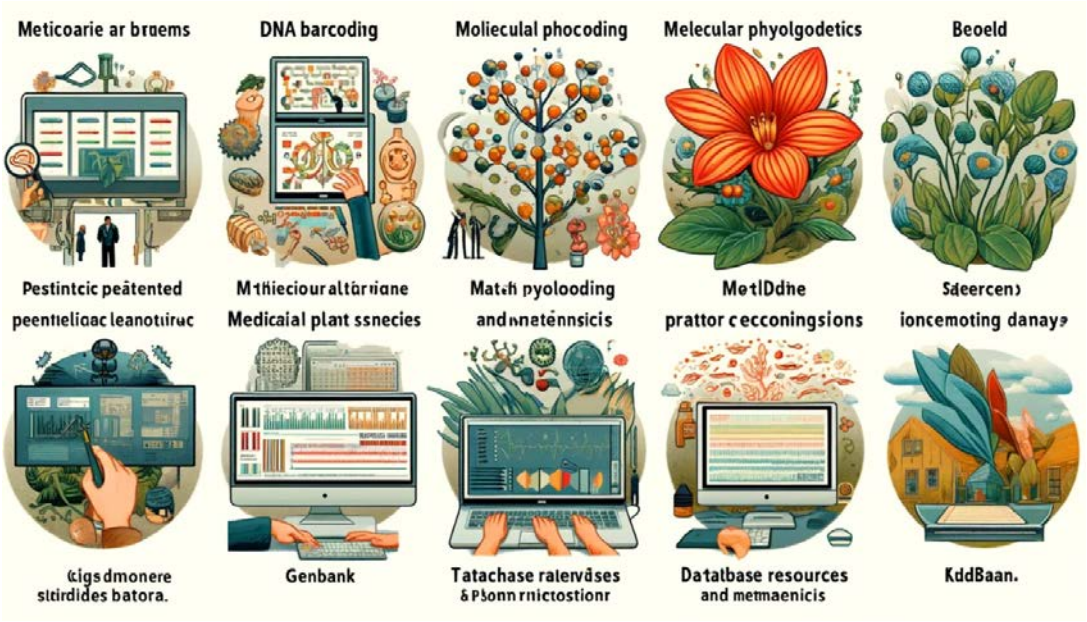


Figure 1. Bioinformatics Approaches for Medicinal Plant Identification.

3. Omics Technologies in Medicinal Plant Research

Omics technologies have revolutionized medicinal plant research by enabling comprehensive analyses of their genomes, transcriptomes, proteomes, metabolomes, and interactomes [28]. These high-throughput techniques provide insights into the molecular mechanisms underlying the biosynthesis of bioactive compounds, interactions with the environment, and therapeutic properties of medicinal plants [29]. Here are some key omics technologies and their applications in medicinal plant research:

1.Genomics: Genomic studies involve the sequencing, assembly, and annotation of the entire DNA content of medicinal plants [30]. Next-generation sequencing (NGS) platforms, such as Illumina and PacBio, facilitate the generation of whole-genome sequences, allowing researchers to explore genetic variations, gene families, and regulatory elements associated with medicinal traits. Comparative genomics analyses enable the identification of conserved pathways and candidate genes involved in the biosynthesis of secondary metabolites with pharmaceutical potential [31].

2. Transcriptomics: Transcriptomic analyses involve the high-throughput sequencing of RNA transcripts to profile gene expression patterns in different tissues, developmental stages, or environmental conditions [32]. RNA-Seq technologies provide quantitative measurements of gene expression levels and transcript isoforms, enabling the discovery of genes involved in the biosynthesis of pharmacologically active compounds [33]. Differential gene expression analysis and co-expression network analysis elucidate transcriptional regulatory networks and metabolic pathways underlying medicinal properties.

3. Proteomics: Proteomic studies focus on the large-scale identification and quantification of proteins expressed in medicinal plants [29,34]. Mass spectrometry-based proteomics techniques, such as LC-MS/MS (liquid chromatography-tandem mass spectrometry), enable the characterization of protein profiles, post-translational modifications, and protein-protein interactions [35]. Proteomics analyses uncover the abundance and activity of enzymes involved in secondary metabolism pathways, facilitating the identification of key enzymes for targeted metabolic engineering or pathway optimization.

4. Metabolomics: Metabolomic approaches aim to characterize the diverse array of small molecules (metabolites) present in medicinal plants and their metabolic profiles under different conditions [36]. Liquid chromatography (LC), gas chromatography (GC), and nuclear magnetic resonance (NMR) spectroscopy are commonly used analytical techniques for metabolite profiling and identification [37]. Metabolomics studies provide insights into the chemical composition, biosynthetic

 doi:10.20944/preprints202405.0752.v1

pathways, and metabolic regulation of bioactive compounds, guiding the selection of medicinal plant varieties and [29] cultivation practices for enhanced phytochemical yield [29].

5. Pharmacogenomics: Pharmacogenomic analyses investigate the genetic variations and polymorphisms in medicinal plant genomes that influence individual responses to herbal remedies and natural products [38]. Genome-wide association studies (GWAS) and candidate gene approaches identify genetic markers associated with pharmacokinetic and pharmacodynamic traits, drug metabolism, and adverse drug reactions [39]. Pharmacogenomic data facilitate personalized medicine approaches by guiding the selection of appropriate medicinal plant formulations and dosage regimens tailored to individual genetic profiles [40].

6. Systems Biology: Systems biology integrates omics data from genomics, transcriptomics, proteomics, and metabolomics to model and simulate the complex interactions within biological systems [40]. Network-based approaches, such as metabolic pathway analysis, gene regulatory networks, and protein interaction networks, elucidate the interconnectedness of molecular components and pathways underlying medicinal plant traits [41]. Systems biology frameworks enable the prediction of emergent properties and the identification of potential drug targets for therapeutic intervention.

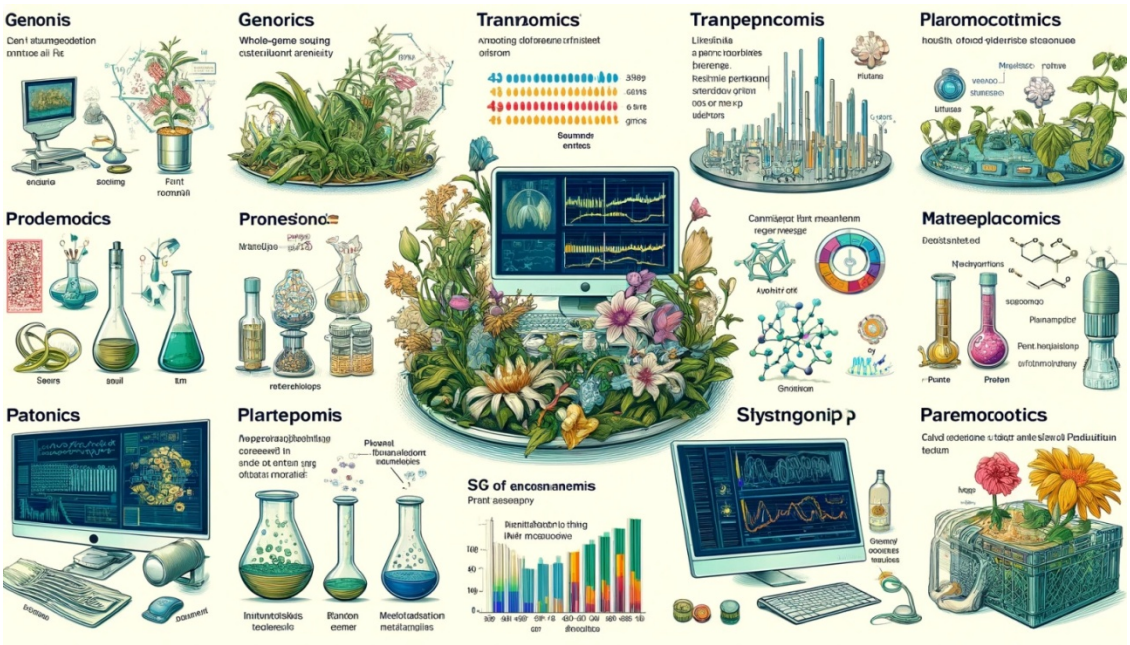


Figure 2. Omics Technologies in Medicinal Plant Research.

4. Computational Tools for Phytochemical Analysis and Drug Discovery

Computational tools play a crucial role in phytochemical analysis and drug discovery from medicinal plants by facilitating the prediction, analysis, and optimization of bioactive compounds [42]. These tools leverage molecular modeling, cheminformatics, and machine learning techniques to explore the chemical space of natural products, identify lead compounds, and optimize their pharmacological properties [43]. Here are some key computational tools used in phytochemical analysis and drug discovery:

1. Molecular Docking: Molecular docking is a computational method used to predict the binding mode and affinity of small molecules (ligands) to a target protein (receptor) [44]. Docking algorithms, such as AutoDock, DOCK, and GOLD, perform conformational sampling and scoring of ligand-receptor interactions to identify potential binding poses and binding energies [45]. Docking simulations enable the virtual screening of compound libraries against protein targets implicated in disease pathways, facilitating the identification of lead compounds with therapeutic potential.

2. Molecular Dynamics (MD) Simulations: Molecular dynamics simulations simulate the motion and interactions of atoms and molecules over time, providing insights into the dynamic behavior of protein-ligand complexes and their stability in aqueous environments [46]. MD software packages,

 doi:10.20944/preprints202405.0752.v1

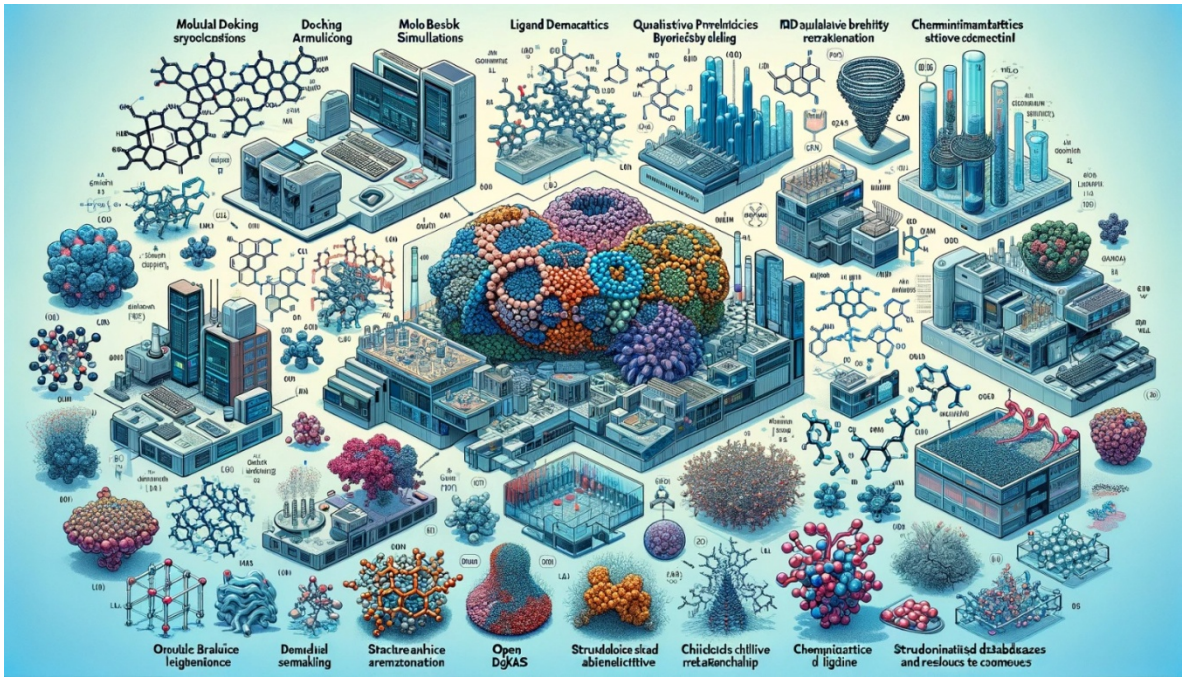
such as GROMACS, AMBER, and NAMD, solve Newton's equations of motion to predict the conformational changes and binding kinetics of ligands within protein binding sites. MD simulations refine docking poses, explore ligand-induced protein conformational changes, and assess ligand binding free energies for lead optimization [46].

3. Ligand-Based Virtual Screening: Ligand-based virtual screening methods identify structurally similar compounds to known bioactive ligands (query molecules) using molecular similarity measures or pharmacophore models [44]. Similarity search algorithms, such as Tanimoto coefficient and cosine similarity, rank compounds in chemical databases based on their structural resemblance to query molecules [47]. Pharmacophore-based screening identifies molecules that match key chemical features essential for ligand-receptor interactions. Tools like Open Babel, RDKit, and ChemMine facilitate ligand similarity searching and pharmacophore modeling for virtual screening campaigns [48].

4. Quantitative Structure-Activity Relationship (QSAR) Modeling: QSAR modeling predicts the biological activity of compounds based on their physicochemical properties and molecular descriptors. QSAR algorithms, including multiple linear regression (MLR) [49], support vector machines (SVM), and random forest (RF), correlate chemical features with experimental activity data to build predictive models. QSAR models enable the rational design of analogs with optimized pharmacological properties and the prioritization of lead compounds for experimental validation [50]. Software packages like Cheminformatics Toolkit (ChemTK) and MOE (Molecular Operating Environment) provide tools for QSAR modeling and molecular descriptor calculation.

5. Structure-Based Drug Design: Structure-based drug design (SBDD) utilizes three-dimensional structures of target proteins to design novel ligands with high affinity and selectivity [51]. SBDD approaches, such as de novo ligand design, fragment-based drug design (FBDD), and protein-ligand docking, exploit the atomic details of ligand-receptor interactions to optimize binding affinity and specificity [52]. Computational tools like Schrödinger Suite, Discovery Studio, and PyMOL enable molecular visualization, structure-based docking, and lead optimization in SBDD workflows [53].

6. Cheminformatics Databases and Resources: Cheminformatics databases and repositories provide curated collections of chemical compounds, bioactivity data, and molecular descriptors for drug discovery research. Public databases such as PubChem, ChEMBL, and ZINC offer vast libraries of natural products, synthetic compounds, and drug-like molecules for virtual screening and lead identification [54]. Cheminformatics tools like KNIME, RDKit, and ChemAxon provide workflows and libraries for chemical data processing, structure visualization, and property prediction in drug discovery projects [55].



 doi:10.20944/preprints202405.0752.v1

Figure 3. Computational Tools for Phytochemical Analysis and Drug Discovery.

5. Pharmacological Investigation and Therapeutic Applications

Pharmacological investigation and therapeutic applications of medicinal plants involve the evaluation of their biological activities, mechanisms of action, and potential therapeutic benefits for treating various diseases and health conditions [56]. Computational tools and bioinformatics approaches play a vital role in predicting, analyzing, and validating the pharmacological properties of bioactive compounds derived from medicinal plants [57]. Here are some key aspects of pharmacological investigation and therapeutic applications:

1. In Silico Screening and Prediction: In silico screening techniques, such as molecular docking, virtual screening, and pharmacophore modeling, predict the binding affinity and interaction of bioactive compounds with drug targets implicated in disease pathways [58]. Computational tools like AutoDock, Glide, and PharmMapper facilitate the identification of potential drug candidates from natural product libraries based on their structural complementarity and pharmacological profiles [59].

2. Pharmacokinetic and Toxicity Prediction: Pharmacokinetic modeling and ADME (absorption, distribution, metabolism, and excretion) prediction assess the pharmacokinetic properties and safety profiles of bioactive compounds derived from medicinal plants [60]. Computational models, such as QSAR (quantitative structure-activity relationship) and PBPK (physiologically-based pharmacokinetic) modeling, predict ADME parameters, bioavailability, and potential toxicity endpoints to guide lead optimization and candidate selection [61].

3. Network Pharmacology Analysis: Network pharmacology integrates omics data, network analysis, and systems biology approaches to elucidate the complex interactions between bioactive compounds and biological pathways underlying disease pathogenesis [62]. Network-based methods identify drug-target interactions, signaling pathways, and biological networks affected by medicinal plant-derived compounds, facilitating the discovery of multi-targeted therapies and synergistic drug combinations [63].

4. Pharmacological Validation and Experimental Studies: Pharmacological validation involves experimental studies to confirm the biological activities and therapeutic effects of medicinal plant-derived compounds in vitro and in vivo [64]. Cell-based assays, animal models, and clinical trials assess the efficacy, safety, and mechanism of action of bioactive compounds for treating specific diseases or health conditions [65]. Computational tools like GraphPad Prism, R, and SPSS analyze experimental data and evaluate the statistical significance of pharmacological outcomes.

5. Ethnopharmacology and Traditional Medicine: Ethnopharmacological studies explore the traditional uses of medicinal plants in indigenous healing practices and folk medicine systems. Bioinformatics approaches analyze ethnobotanical data, traditional knowledge, and phytochemical information to validate the pharmacological properties and therapeutic claims of medicinal plants [66]. Integrating traditional medicine with modern pharmacology enhances our understanding of the cultural significance and therapeutic potential of plant-based remedies.

6. Personalized Medicine and Precision Pharmacotherapy: Personalized medicine approaches tailor therapeutic interventions based on individual genetic variations, pharmacogenomic profiles, and clinical parameters [67]. Computational tools like pharmacogenomic databases, genotype-phenotype prediction models, and electronic health records enable the stratification of patient populations and the selection of personalized treatment regimens incorporating medicinal plant-derived compounds [68]. Precision pharmacotherapy optimizes therapeutic outcomes, minimizes adverse drug reactions, and improves patient compliance in personalized healthcare settings.

 doi:10.20944/preprints202405.0752.v1

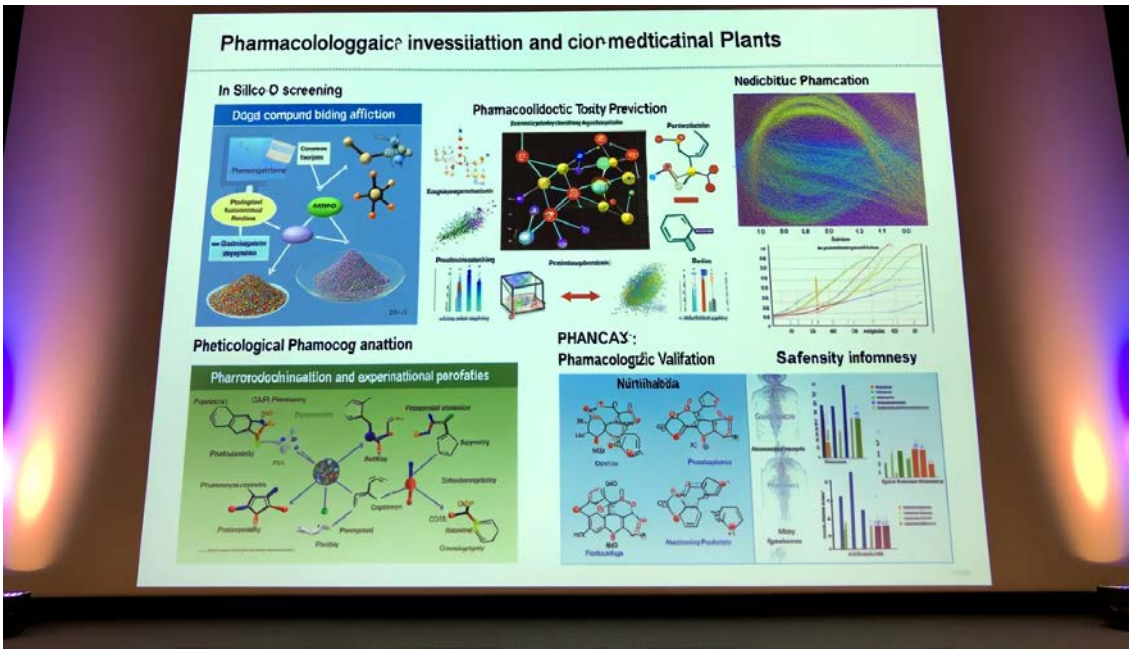


Figure 4. Pharmacological Investigation and Therapeutic Applications.

6. Conservation and Sustainable Utilization of Medicinal Plants

Conservation and sustainable utilization of medicinal plants are critical to ensure their long-term availability for healthcare, biodiversity conservation, and socio-economic development [69]. Bioinformatics plays a pivotal role in supporting conservation efforts by providing tools and resources for biodiversity assessment, population genetics analysis, habitat modeling, and sustainable harvesting practices [70]. Here are key aspects of conservation and sustainable utilization of medicinal plants with a focus on bioinformatics:

1. Biodiversity Assessment: Bioinformatics tools enable the assessment of medicinal plant diversity at various spatial and temporal scales [71]. DNA barcoding, metabarcoding, and genomic sequencing techniques identify species richness, genetic variation, and population structure in natural plant populations. Computational algorithms, such as population genetic software packages (e.g., STRUCTURE, ADMIXTURE), analyze molecular data to estimate genetic diversity indices, gene flow patterns, and evolutionary relationships among plant populations [72]. Biodiversity hotspots and priority conservation areas are identified based on genomic data and ecological niche modeling approaches.

2. Population Genetics Analysis: Population genetics studies elucidate the genetic structure, adaptive traits, and demographic history of medicinal plant populations [73]. Bioinformatics pipelines, such as PHYLIP, BEAST, and GenomicRanges, analyze genetic variation, linkage disequilibrium, and demographic processes using genomic data [74]. Population genomics approaches integrate molecular markers, environmental data, and spatial analysis techniques to assess the impact of habitat fragmentation, climate change, and human activities on plant population dynamics [75]. Conservation strategies, such as ex situ germplasm conservation and in situ habitat restoration, are informed by population genetics data to preserve genetic diversity and adaptive potential in medicinal plant populations [76].

3. Habitat Modeling and Conservation Planning: Bioinformatics-based habitat modeling techniques predict the distribution, habitat suitability, and climate niche of medicinal plant species under current and future environmental scenarios [77]. Species distribution models (SDMs), ecological niche models (ENMs), and MaxEnt algorithms integrate environmental variables, species occurrence data, and remote sensing data to identify suitable habitats and conservation corridors for endangered plant species [78]. Conservation planning tools, such as Marxan and Zonation, prioritize conservation areas and design protected areas networks based on habitat suitability and connectivity analyses.

 doi:10.20944/preprints202405.0752.v1

4. Sustainable Harvesting Practices: Bioinformatics supports the sustainable utilization of medicinal plants by monitoring and regulating harvesting practices to ensure their ecological and socio-economic sustainability [79]. Remote sensing technologies, geographic information systems (GIS), and spatial databases track changes in land cover, habitat fragmentation, and land use patterns in medicinal plant habitats [80]. Decision support systems, such as HarvestMap and MapViewer, facilitate participatory mapping of resource use areas, stakeholder engagement, and adaptive management of wild harvests. Bioeconomic modeling approaches integrate ecological, economic, and social factors to optimize harvest quotas, harvesting seasons, and conservation incentives for sustainable resource management [81].

5. Traditional Knowledge and Indigenous Practices: Bioinformatics integrates traditional ecological knowledge (TEK) and indigenous practices into conservation strategies for medicinal plants [82]. Ethnobotanical databases, community-based monitoring programs, and participatory mapping initiatives document traditional uses, harvesting techniques, and cultural significance of medicinal plants among indigenous communities [83]. Bioinformatics tools, such as Ethnobotanical Information System (Ethnobiology), digitize and analyze ethnographic data to inform conservation policies, intellectual property rights, and benefit-sharing agreements that respect indigenous rights and promote biocultural diversity conservation [83].

Figure 5. Conservation and Sustainable Utilization of Medicinal Plants.

7. Challenges and Future Directions

Challenges and future directions in the field of medicinal plant research encompass a range of interdisciplinary and multifaceted issues, from scientific and technological hurdles to ethical, regulatory, and socio-economic considerations. Addressing these challenges and advancing research in key areas will be critical for realizing the full potential of medicinal plants in healthcare, drug discovery, and biodiversity conservation. Here are some challenges and future directions:

1. Integration of Multi-Omics Data: Integrating data from genomics, transcriptomics, proteomics, metabolomics, and other omics disciplines poses challenges in data integration, standardization, and interpretation [84]. Developing robust bioinformatics pipelines and computational frameworks for multi-omics data analysis will be essential for unraveling the complex molecular networks underlying the therapeutic properties of medicinal plants.

2. Validation and Reproducibility: Ensuring the reproducibility and validity of research findings in medicinal plant research remains a challenge due to variability in plant materials, experimental

 doi:10.20944/preprints202405.0752.v1

conditions, and methodological approaches [85]. Standardizing experimental protocols, adopting transparent reporting guidelines, and promoting data sharing practices will enhance the reliability and credibility of research outcomes.

3. Access to Genetic Resources and Benefit-Sharing: Access to genetic resources and traditional knowledge associated with medicinal plants raises ethical and legal challenges related to intellectual property rights, benefit-sharing agreements, and bioprospecting regulations. Establishing equitable access and benefit-sharing mechanisms, in accordance with international conventions such as the Nagoya Protocol, will promote fair and mutually beneficial partnerships between researchers, local communities, and indigenous peoples [86].

4. Sustainable Sourcing and Conservation: Overharvesting, habitat destruction, and climate change threaten the sustainability of medicinal plant resources and biodiversity conservation. Implementing sustainable sourcing practices, habitat restoration initiatives, and community-based conservation projects will mitigate the ecological impact of wild harvesting and ensure the long-term availability of medicinal plants for future generations [69].

5. Quality Control and Standardization: Ensuring the quality, safety, and efficacy of herbal products derived from medicinal plants requires robust quality control measures, standardization protocols, and regulatory frameworks [87]. Developing analytical methods, reference standards, and pharmacopoeial guidelines for assessing the identity, purity, and potency of herbal medicines will enhance consumer confidence and regulatory compliance in the herbal products industry.

6. Interdisciplinary Collaboration and Capacity Building: Fostering interdisciplinary collaboration and capacity building initiatives is essential for advancing research in medicinal plant science and bioinformatics [88]. Strengthening partnerships between academia, industry, government agencies, and civil society organizations will promote knowledge exchange, technology transfer, and innovation in drug discovery, conservation biology, and sustainable development [89].

7. Public Awareness and Education: Raising public awareness about the value of medicinal plants, traditional medicine systems, and the importance of biodiversity conservation is crucial for promoting sustainable practices and responsible use of natural resources [90]. Education campaigns, outreach programs, and community engagement activities can empower stakeholders to make informed decisions and support initiatives that prioritize environmental stewardship and cultural preservation [91].

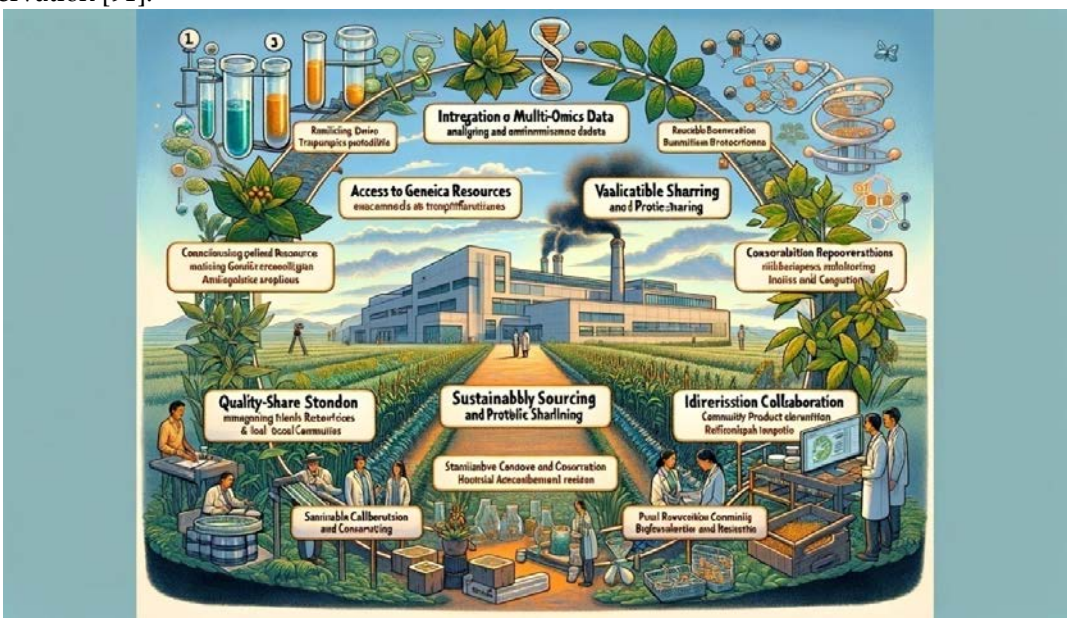


Figure 6. Challenges and Future Directions.

8. Conclusion

Bioinformatics offers unprecedented opportunities for the identification, characterization, and study of medicinal plants, revolutionizing traditional approaches to drug discovery and

 doi:10.20944/preprints202405.0752.v1

development. By leveraging omics technologies, computational tools, and database resources, researchers can unravel the complex biological properties of medicinal plants and harness their therapeutic potential for addressing global health challenges. Continued interdisciplinary collaboration and investment in bioinformatics research are essential for unlocking the full promise of medicinal plants in modern medicine and sustainable healthcare.

References

1. Srivastava, A.K., *Significance of medicinal plants in human life*, in *Synthesis of Medicinal Agents from Plants*. 2018, Elsevier. p. 1-24.
2. Ramawat, K., S. Dass, and M. Mathur, *The chemical diversity of bioactive molecules and therapeutic potential of medicinal plants.* Herbal drugs: ethnomedicine to modern medicine, 2009: p. 7-32.
3. Rashid, S., et al., *Phytomedicines: Diversity, extraction, and conservation strategies*, in *Phytomedicine*. 2021, Elsevier. p. 1-33.
4. Ogunjobi, T.T., et al., *Bioinformatics Applications in Chronic Diseases: A Comprehensive Review of Genomic, Transcriptomics, Proteomic, Metabolomics, and Machine Learning Approaches.* Medinformatics, 2024.
5. Vodovotz, Y., et al., *Solving immunology?* Trends in immunology, 2017. **38**(2): p. 116-127.
6. Chassagne, F., et al., *A systematic review of plants with antibacterial activities: A taxonomic and phylogenetic perspective.* Frontiers in pharmacology, 2021. **11**: p. 586548.
7. Gayathiri, E., et al., *Computational approaches for modeling and structural design of biological systems: A comprehensive review.* Progress in Biophysics and Molecular Biology, 2023.
8. Romano, J.D. and N.P. Tatonetti, *Informatics and computational methods in natural product drug discovery: a review and perspectives.* Frontiers in genetics, 2019. **10**: p. 442506.
9. Balick, M.J. *Ethnobotany, drug development and biodiversity conservation—exploring the linkages*. in *Ciba Foundation Symposium 185-Ethnobotany and the Search for New Drugs: Ethnobotany and the Search for New Drugs: Ciba Foundation Symposium 185*. 2007. Wiley Online Library.
10. Raza, K. and N. Dey, *Translational bioinformatics in healthcare and medicine*. 2021: Academic Press.
11. Abraham, E.J. and J.J. Kellogg, *Chemometric-guided approaches for profiling and authenticating botanical materials.* Frontiers in Nutrition, 2021. **8**: p. 780228.
12. Zhang, Y. and Y. Wang, *Recent trends of machine learning applied to multi-source data of medicinal plants.* Journal of Pharmaceutical Analysis, 2023.
13. Ma, X., et al., *Bioinformatics-assisted, integrated omics studies on medicinal plants.* Briefings in bioinformatics, 2020. **21**(6): p. 1857-1874.
14. von Cräutlein, M., et al., *DNA barcoding: a tool for improved taxon identification and detection of species diversity.* Biodiversity and conservation, 2011. **20**: p. 373-389.
15. Khan, S.A., et al., *ITS2: an ideal DNA barcode for the arid medicinal plant Rhazya stricta.* Pharmaceutical Medicine, 2019. **33**: p. 53-61.
16. Ali, M.A., et al., *The changing epitome of species identification–DNA barcoding.* Saudi journal of biological sciences, 2014. **21**(3): p. 204-231.
17. Uncu, A.O., et al., *A primer to molecular phylogenetic analysis in plants.* Critical reviews in plant sciences, 2015. **34**(4): p. 454-468.
18. Day, P.D., et al., *Evolutionary relationships in the medicinally important genus Fritillaria L.(Liliaceae).* Molecular Phylogenetics and Evolution, 2014. **80**: p. 11-19.
19. Dev, S.A., et al., *Quantification of adulteration in traded ayurvedic raw drugs employing machine learning approaches with DNA barcode database.* 3 Biotech, 2021. **11**: p. 1-16.
20. Moler, E., M. Chow, and I. Mian, *Analysis of molecular profile data using generative and discriminative methods.* Physiological genomics, 2000. **4**(2): p. 109-126.
21. El-Hasnony, I.M., et al., *Improved feature selection model for big data analytics.* IEEE Access, 2020. **8**: p. 66989-67004.
22. Deiner, K., et al., *Environmental DNA metabarcoding: Transforming how we survey animal and plant communities.* Molecular ecology, 2017. **26**(21): p. 5872-5895.
23. Raclariu-Manolică, A.C. and H.J. de Boer, *DNA barcoding and metabarcoding for quality control of botanicals and derived herbal products*, in *Evidence-Based Validation of Herbal Medicine*. 2022, Elsevier. p. 223-238.
24. López-García, A., et al., *Comparison of Mothur and QIIME for the analysis of rumen microbiota composition based on 16S rRNA amplicon sequences.* Frontiers in Microbiology, 2018. **9**: p. 417268.
25. Sorokina, M. and C. Steinbeck, *Review on natural products databases: where to find data in 2020.* Journal of cheminformatics, 2020. **12**(1): p. 20.

 doi:10.20944/preprints202405.0752.v1

26. Zahoor, I., et al., *Bioinformatics and medicinal plant research: current scenario.* Essentials of Bioinformatics, Volume III: In Silico Life Sciences: Agriculture, 2019: p. 141-157.
27. Wong, L.L., et al., *DNA barcoding of catfish: species authentication and phylogenetic assessment.* PloS one, 2011. **6**(3): p. e17812.
28. Dai, X. and L. Shen, *Advances and trends in omics technology development.* Frontiers in Medicine, 2022. **9**: p. 911861.
29. Zhang, W., et al., *Integration of high-throughput omics technologies in medicinal plant research: The new era of natural drug discovery.* Frontiers in Plant Science, 2023. **14**: p. 1073848.
30. Zhang, Y., et al., *Assembly and annotation of a draft genome of the medicinal plant Polygonum cuspidatum.* Frontiers in Plant Science, 2019. **10**: p. 1274.
31. Duan, Y., et al., *Genome sequencing of Inonotus obliquus reveals insights into candidate genes involved in secondary metabolite biosynthesis.* BMC genomics, 2022. **23**(1): p. 314.
32. Wolf, J.B., *Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial.* Molecular ecology resources, 2013. **13**(4): p. 559-572.
33. Wang, F., et al., *Transcriptome analysis of salicylic acid treatment in Rehmannia glutinosa hairy roots using RNA-seq technique for identification of genes involved in acteoside biosynthesis.* Frontiers in Plant Science, 2017. **8**: p. 787.
34. Jain, R., et al., *Proteomics Approaches in Medicinal Plant Research and Pharmacological Studies*, in *Plant Proteomics*. CRC Press. p. 197-225.
35. Neagu, A.-N., et al., *Applications of tandem mass spectrometry (MS/MS) in protein analysis for biomedical research.* Molecules, 2022. **27**(8): p. 2411.
36. Scossa, F., et al., *The integration of metabolomics and next-generation sequencing data to elucidate the pathways of natural product metabolism in medicinal plants.* Planta medica, 2018. **84**(12/13): p. 855-873.
37. Gathungu, R.M., et al., *Identification of metabolites from liquid chromatography–coulometric array detection profiling: gas chromatography–mass spectrometry and refractionation provide essential information orthogonal to LC–MS/microNMR.* Analytical biochemistry, 2014. **454**: p. 23-32.
38. Thomford, N.E., et al., *Pharmacogenomics implications of using herbal medicinal plants on African populations in health transition.* Pharmaceuticals, 2015. **8**(3): p. 637-663.
39. v. Wintzingerode, F., U.B. Göbel, and E. Stackebrandt, *Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis.* FEMS microbiology reviews, 1997. **21**(3): p. 213-229.
40. Babar, M.M., N. us Sahar Sadaf Zaidi, and A.G. Kazi, *Plant Pharmacogenomics: From Drug Discovery to Personalized Ethnomedicine.* PlantOmics: The Omics of Plant Science, 2015: p. 699-730.
41. Mukherjee, P.K., S. Banerjee, and A. Kar, *Molecular combination networks in medicinal plants: understanding synergy by network pharmacology in Indian traditional medicine.* Phytochemistry Reviews, 2021. **20**(4): p. 693-703.
42. Egbuna, C., M. Rudrapal, and H. Tijjani, *Phytochemistry, computational tools, and databases in drug discovery*. 2022: Elsevier.
43. Moshawih, S., et al., *Synergy between machine learning and natural products cheminformatics: Application to the lead discovery of anthraquinone derivatives.* Chemical Biology & Drug Design, 2022. **100**(2): p. 185-217.
44. Guedes, I.A., C.S. de Magalhães, and L.E. Dardenne, *Receptor–ligand molecular docking.* Biophysical reviews, 2014. **6**: p. 75-87.
45. Naqvi, A.A., et al., *Advancements in docking and molecular dynamics simulations towards ligand-receptor interactions and structure-function relationships.* Current topics in medicinal chemistry, 2018. **18**(20): p. 1755-1768.
46. Shukla, R. and T. Tripathi, *Molecular dynamics simulation of protein and protein–ligand complexes.* Computer-aided drug design, 2020: p. 133-161.
47. Willett, P., *The calculation of molecular structural similarity: principles and practice.* Molecular informatics, 2014. **33**(6-7): p. 403-413.
48. Horvath, D., *Pharmacophore-based virtual screening.* Chemoinformatics and computational chemical biology, 2011: p. 261-298.
49. Kobayashi, Y. and K. Yoshida, *Development of QSAR models for prediction of fish bioconcentration factors using physicochemical properties and molecular descriptors with machine learning algorithms.* Ecological Informatics, 2021. **63**: p. 101285.
50. Patel, H.M., et al., *Quantitative structure–activity relationship (QSAR) studies as strategic approach in drug discovery.* Medicinal chemistry research, 2014. **23**: p. 4991-5007.
51. Wang, X., et al., *Structure-based drug design strategies and challenges.* Current Topics in Medicinal Chemistry, 2018. **18**(12): p. 998-1006.

 doi:10.20944/preprints202405.0752.v1

52. Lounnas, V., et al., *Current progress in structure-based rational drug design marks a new mindset in drug discovery.* Computational and structural biotechnology journal, 2013. **5**(6): p. e201302011.
53. Cox, P.B. and R. Gupta, *Contemporary computational applications and tools in drug discovery.* ACS Medicinal Chemistry Letters, 2022. **13**(7): p. 1016-1029.
54. Gozalbes, R. and A. Pineda-Lucena, *Small molecule databases and chemical descriptors useful in chemoinformatics: an overview.* Combinatorial chemistry & high throughput screening, 2011. **14**(6): p. 548-558.
55. Mazanetz, M.P., C.H. Goode, and E.I. Chudyk, *Ligand-and structure-based drug design and optimization using KNIME.* Current medicinal chemistry, 2020. **27**(38): p. 6458-6479.
56. Anand, U., et al., *A comprehensive review on medicinal plants as antimicrobial therapeutics: potential avenues of biocompatible drug discovery.* Metabolites, 2019. **9**(11): p. 258.
57. Mehmood, M.A., U. Sehar, and N. Ahmad, *Use of bioinformatics tools in different spheres of life sciences.* Journal of Data Mining in Genomics & Proteomics, 2014. **5**(2): p. 1.
58. Olğaç, A., I.E. Orhan, and E. Banoglu, *The potential role of in silico approaches to identify novel bioactive molecules from natural resources.* Future Medicinal Chemistry, 2017. **9**(14): p. 1665-1686.
59. Patel, J.R., et al., *A review on computational software tools for drug design and discovery.* Indo Global Journal of Pharmaceutical Sciences, 2022. **12**: p. 53-81.
60. Durán-Iturbide, N.A., B.I. Díaz-Eufracio, and J.L. Medina-Franco, *In silico ADME/Tox profiling of natural products: A focus on BIOFACQUIM.* ACS omega, 2020. **5**(26): p. 16076-16084.
61. Sucharitha, P., et al., *Absorption, distribution, metabolism, excretion, and toxicity assessment of drugs using computational tools*, in *Computational approaches for novel therapeutic and diagnostic designing to mitigate SARS-CoV-2 infection*. 2022, Elsevier. p. 335-355.
62. Leung, E.L., et al., *Network-based drug discovery by integrating systems biology and computational technologies.* Briefings in bioinformatics, 2013. **14**(4): p. 491-505.
63. Kibble, M., et al., *Network pharmacology applications to map the unexplored target space and therapeutic potential of natural products.* Natural product reports, 2015. **32**(8): p. 1249-1266.
64. Yikna, B.B. and A.S. Yehualashet, *Medicinal plant extracts evaluated in vitro and in vivo for antidiabetic activities in Ethiopia: bases for future clinical trials and related investigations.* Evidence-Based Complementary and Alternative Medicine, 2021. **2021**.
65. Pratap, K., et al., *A comprehensive review on natural bioactive compounds and probiotics as potential therapeutics in food allergy treatment.* Frontiers in immunology, 2020. **11**: p. 518111.
66. Asiamah, I., et al., *Applications of molecular docking in natural products-based drug discovery.* Scientific African, 2023. **20**: p. e01593.
67. Zhou, S.-F., et al., *Clinical pharmacogenetics and potential application in personalized medicine.* Current drug metabolism, 2008. **9**(8): p. 738-784.
68. Shah, I.M., et al., *Computational and pharmacogenomic resources*, in *Pharmacogenomics*. 2023, Elsevier. p. 345-362.
69. Mir, T.A., et al., *Medicinal plant resources: threat to its biodiversity and conservation strategies.* Medicinal and Aromatic Plants: Healthcare and Industrial Applications, 2021: p. 717-739.
70. Karaca, M. and A.G. Ince, *Conservation of biodiversity and genetic resources for sustainable agriculture.* Innovations in sustainable agriculture, 2019: p. 363-410.
71. Sharma, V. and I.N. Sarkar, *Bioinformatics opportunities for identification and study of medicinal plants.* Briefings in bioinformatics, 2013. **14**(2): p. 238-250.
72. Elbrecht, V., et al., *Estimating intraspecific genetic diversity from community DNA metabarcoding data.* PeerJ, 2018. **6**: p. e4644.
73. Dang, H., et al., *Population evolution, genetic diversity and structure of the medicinal legume, glycyrrhiza uralensis and the effects of geographical distribution on leaves nutrient elements and photosynthesis.* Frontiers in plant science, 2022. **12**: p. 708709.
74. Porubsky, D., et al., *Haplotype-resolved inversion landscape reveals hotspots of mutational recurrence associated with genomic disorders.* bioRxiv, 2021: p. 2021.12. 20.472354.
75. Hoffmann, A., et al., *A framework for incorporating evolutionary genomics into biodiversity conservation and management.* Climate Change Responses, 2015. **2**: p. 1-24.
76. Hawkes, J.G., N. Maxted, and B.V. Ford-Lloyd, *The ex situ conservation of plant genetic resources*. 2012: Springer Science & Business Media.
77. Fulekar, M., *Environmental biotechnology*. 2010: CRC Press.
78. Roy, S., et al., *Use of species distribution models to study habitat suitability for sustainable management and conservation in the Indian subcontinent: A decade's retrospective.* Frontiers in Sustainable Resource Management, 2022. **1**: p. 1031646.

 doi:10.20944/preprints202405.0752.v1

79. Kumar, R. and P. Saikia, *Forest resources of Jharkhand, Eastern India: socio-economic and bio-ecological perspectives.* Socio-economic and Eco-biological Dimensions in Resource use and Conservation: Strategies for Sustainability, 2020: p. 61-101.
80. Ramachandran, R.M., et al., *Long-term land use and land cover changes (1920–2015) in Eastern Ghats, India: Pattern of dynamics and challenges in plant species conservation.* Ecological Indicators, 2018. **85**: p. 21-36.
81. Clark, C.W., *Mathematical bioeconomics: the mathematics of conservation*. Vol. 91. 2010: John Wiley & Sons.
82. O'Neill, A.R., et al., *Integrating ethnobiological knowledge into biodiversity conservation in the Eastern Himalayas.* Journal of ethnobiology and ethnomedicine, 2017. **13**: p. 1-14.
83. Shukla, S. and J. Gardner, *Local knowledge in community-based approaches to medicinal plant conservation: lessons from India.* Journal of Ethnobiology and Ethnomedicine, 2006. **2**: p. 1-9.
84. Misra, B.B., et al., *Integrated omics: tools, advances and future approaches.* Journal of molecular endocrinology, 2019. **62**(1): p. R21-R45.
85. Pferschy-Wenzig, E.-M. and R. Bauer, *The relevance of pharmacognosy in pharmacological research on herbal medicinal products.* Epilepsy & behavior, 2015. **52**: p. 344-362.
86. Robinson, D.F., *Biodiversity, access and benefit-sharing: global case studies*. 2014: Routledge.
87. Govindaraghavan, S. and N.J. Sucher, *Quality assessment of medicinal herbs and their extracts: Criteria and prerequisites for consistent safety and efficacy of herbal medicines.* Epilepsy & Behavior, 2015. **52**: p. 363-371.
88. Council, N.R., et al., *Convergence: Facilitating transdisciplinary integration of life sciences, physical sciences, engineering, and beyond*. 2014: National Academies Press.
89. Kueffer, C., et al., *Enabling effective problem-oriented research for sustainable development.* Ecology and Society, 2012. **17**(4).
90. Hamilton, A.C., *Medicinal plants, conservation and livelihoods.* Biodiversity & Conservation, 2004. **13**: p. 1477-1517.
91. Jacobson, S.K., M.D. McDuff, and M.C. Monroe, *Conservation education and outreach techniques*. 2015: Oxford University Press.