



Integrating ethnobotany and quantitative methods: A bibliometric analysis of statistical, mathematical, and data science approaches (2016–2025)

Syella Zignora Limba, I Gede Awantara, Boy Riansyah, Muh Akbar Idris

Correspondence

Syella Zignora Limba¹, I Gede Awantara¹, Boy Riansyah¹, Muh Akbar Idris^{1,2*}

¹School of Data Science, Mathematics, and Informatics, IPB University, Bogor, Indonesia

²Faculty of Mathematics and Natural Sciences, Universitas Sulawesi Barat, Majene, Indonesia

*Corresponding Author: muhakbaridris@apps.ipb.ac.id

Ethnobotany Research and Applications 34:37 (2026) - <http://dx.doi.org/10.32859/era.34.37.1-9>

Manuscript received: 07/04/2026 - Revised manuscript received: 30/04/2026 - Published: 01/05/2026

Research

Abstract

Background: The application of quantitative methods, including statistics, mathematics, and data science, in ethnobotanical research represents an emerging interdisciplinary frontier. While many bibliometric studies focus on specific medicinal plants, the overall methodological evolution of computational tools within ethnobotany remains largely unexplored. This study addresses this gap by analyzing the integration of machine learning, multivariate statistics, and network analysis into ethnobotanical research.

Methods: A structured bibliometric analysis was conducted using the Scopus database (2016–2025), following PRISMA guidelines. The search targeted titles, abstracts, and keywords linking ethnobotanical concepts with quantitative approaches. A total of 1,275 documents from 337 sources were analyzed using VOSviewer, applying fractional counting and cluster validation to examine publication trends, collaborations, and thematic evolution.

Results: The findings show an annual growth rate of 12.05%, increasing from 83 publications in 2016 to 231 in 2025. Polynomial regression indicates an accelerating trend. The results reveal a shift from descriptive approaches toward algorithm-based methodologies, supported by strong international collaboration (30.43%) and dominant contributions from countries such as India and China.

Conclusions: The integration of data science is transforming ethnobotany into a predictive discipline. This development supports evidence-based validation of traditional knowledge and improves bioprospecting. Future research should focus on developing a unified Computational Ethnobotany Ontology to support large-scale data integration.

Keywords: Ethnobotany; Quantitative Methods; Statistics; Data Science; Bibliometrics; Traditional Knowledge; Machine Learning; Network Pharmacology

Background

Ethnobotany, defined as the scientific study of the complex relationships between humans and plants, has undergone a profound methodological transformation over the past decade. Historically rooted in the qualitative documentation of indigenous plant knowledge, the discipline primarily produced descriptive inventories of useful flora. While these foundational works were crucial for preserving cultural heritage, they inherently lacked the statistical rigor required for broader scientific validation and reproducibility. Today, ethnobotany has progressively embraced rigorous quantitative frameworks to address the growing scale and urgency of modern biological questions. This paradigm shift is driven by dual pressures: the critical urgency to document rapidly eroding traditional ecological knowledge (TEK) due to globalization, and the widespread availability of advanced computational tools capable of decoding massive, multidimensional datasets.

Statistical hypothesis testing formally moved the field from anecdotal storytelling to evidence-based science. Standardized ethnobotanical indices, such as the Use Value (UV), Informant Consensus Factor (ICF), and Cultural Importance Index (CI), have provided replicable quantitative measures for comparing the cultural utility of plants across communities (Phillips & Gentry 1993; Martin 1995). These indices enable researchers to objectively identify priority species for targeted conservation and strategic bioprospecting. However, the current landscape of quantitative ethnobotany has expanded far beyond simple descriptive statistics.

The exponential accumulation of digitized ethnobotanical records over the last century has created a critical bottleneck: the vast majority of documented traditional uses remain pharmacologically unvalidated. This sheer volume of unverified data necessitates a transition from manual, low-throughput ethnopharmacological screening to high-throughput computational workflows. Consequently, data science acts not merely as an analytical accessory, but as an indispensable bridge between traditional ecological knowledge and modern drug discovery pipelines.

The integration of advanced data science, including machine learning, complex network analysis, and natural language processing (NLP), has dramatically pushed the analytical horizons of the discipline. Researchers now routinely leverage principal component analysis (PCA) to reveal hidden sociocultural patterns within cross-cultural datasets (Leonti *et al.* 2010). Simultaneously, computational molecular docking and network pharmacology are increasingly applied to extract latent pharmacological insights from legacy traditional knowledge corpora. Bibliometric methods have also emerged as reflective tools for mapping the intellectual trajectory of the discipline itself (van Eck & Waltman 2014). Recent methodological studies published in leading outlets such as *Ethnobotany Research and Applications* further underscore the necessity of adopting structured, verifiable frameworks to synthesize global plant use data (Benamar *et al.* 2026).

Despite this explosive adoption of computational biology, a significant gap remains in the literature. To date, there has been no comprehensive bibliometric mapping of the specific intersection between ethnobotany and data science methodologies. Prior bibliometric studies in this sphere are remarkably common but overwhelmingly focused on isolated sub-domains, mapping the literature on specific medicinal plants (Thomford *et al.* 2018), regional wild edibles (Luczaj *et al.* 2012), or general ethnopharmacology (de Albuquerque *et al.* 2013). Consequently, they map the biological subjects rather than the analytical tools. None has systematically isolated the methodological landscape to map how algorithms and statistics are reshaping the field.

In recent years, ethnobotanical research has experienced a significant methodological shift. The evaluation of local plant utilization is no longer limited to descriptive inventories but has increasingly adopted quantitative statistical approaches, such as the application of Use Value (UV) and the Index of Cultural Significance (ICS) to measure cultural importance mathematically (Agesti *et al.* 2023). Furthermore, the integration of data science techniques is rapidly expanding. In Indonesia, machine learning algorithms, including Support Vector Machines and Random Forests, have proven crucial in screening and predicting the pharmacological potential of complex herbal compounds from local plant databases (Erlina *et al.* 2022). This ongoing transition toward computational methods underscores the necessity for a comprehensive bibliometric mapping of the literature.

This study directly addresses that critical positioning gap by conducting an interdisciplinary bibliometric analysis of 1,275 publications (2016–2025) that explicitly integrate ethnobotany with quantitative methodologies. Our objectives are to: (1) measure evolutionary publication trends using statistical forecasting; (2) deeply interpret institutional and collaborative network clusters to identify global power dynamics; and (3) critically delineate the methodological frontiers of quantitative ethnobotany, proposing an explicit framework to transition the discourse from descriptive mapping to algorithmic interpretation.

Materials and Methods

This study employs a reproducible, mixed-methods bibliometric analysis to map global research trends at the intersection of ethnobotany and data science. Bibliometrics provides a rigorous mathematical framework for evaluating scientific output and revealing evolutionary thematic developments within a research field (Donthu *et al.* 2021).

Data Collection and Search Strategy

To ensure the highest quality of bibliographic data, metadata were exclusively sourced from Scopus. Scopus was selected over Web of Science due to its broader coverage of interdisciplinary sciences and superior indexing of journals in the Global South, where a significant volume of primary ethnobotanical field research occurs.

To guarantee methodological transparency and reproducibility, the exact search query was executed within the Title, Abstract, and Keyword ('TITLE-ABS-KEY') fields. The syntax utilized Boolean operators to intersect core ethnobotanical domains with computational methods:

```
TITLE-ABS-KEY ( ( ethnobotan* OR "traditional knowledge" OR "medicinal plant*" OR "indigenous knowledge" OR "wild edible plant*" ) AND ( quantitat* OR statistic* OR "machine learning" OR "data science" OR bibliometric* OR "multivariate" OR "network analysis" OR "informant consensus" ) )
```

Data extraction was performed on a single day (to prevent database update bias) targeting publications strictly within the 2016–2025 window. Acceptable document types were limited to peer-reviewed original articles, reviews, conference papers, and data papers in English. Retracted papers, errata, and non-peer-reviewed notes were explicitly excluded to maintain analytical integrity.

PRISMA Framework

The data selection process meticulously implemented the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework. This systematic screening is crucial for eliminating subjective publication bias. Following the removal of duplicates and off-topic records during the abstract screening phase, exactly 1,275 relevant documents curated from 337 academic sources were retained for final extraction. The precise attrition of records at each screening phase is documented in the PRISMA flow diagram (Fig. 1).

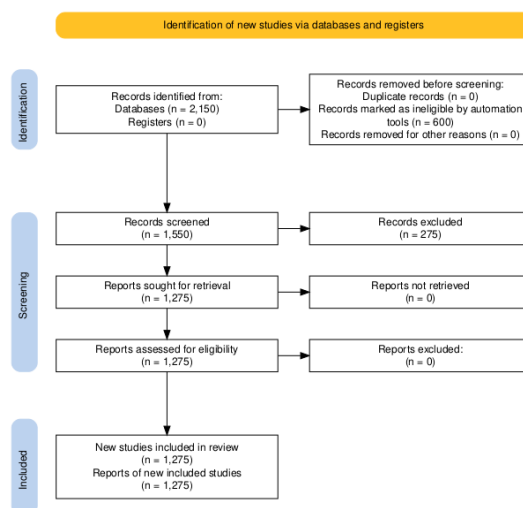


Figure 1. PRISMA flow diagram detailing the systematic literature search and exact selection process, resulting in the final inclusion of 1,275 documents.

Data Analysis and Reproducibility Settings

Raw bibliographic data were analyzed using two complementary computational tools under strict parameter settings to ensure cluster validity. Macro-level indicators, advanced trend regression, and temporal coupling analyses were executed using Biblioshiny (the graphical interface of the R bibliometrix package).

For micro-level relational mapping, VOSviewer software was utilized. To ensure the statistical stability of the keyword co-occurrence networks and prevent highly connected nodes from skewing the visualization, the fractional counting method

was applied instead of full counting. Furthermore, a minimum keyword occurrence threshold was strictly set to 5, filtering out peripheral noise and ensuring that only statistically significant topological clusters emerged. Data completeness was verified prior to network generation, confirming zero critical missing values in publication years, sources, or citation counts.

Table 1. Main bibliometric summary statistics of the analyzed dataset.

Category	Details
Timespan	2016–2025
Sources (Journals, Books, etc.)	337
Documents	1,275
Annual Growth Rate (%)	12.05
Document Average Age (years)	3.45
Average Citations/Document	17.82
Author Keywords (DE)	5,712
Authors	7,630
Single-Authored Documents	97
International Co-Authorship (%)	30.43

Results

Publication Trends and Advanced Trend Forecasting

The volume of publications seamlessly integrating data science and ethnobotany grew substantially over the decade, rising from a modest 83 documents in 2016 to an impressive 231 documents in 2025. This trajectory represents a compound annual growth rate of 12.05%.

To project future trajectories and elevate the analysis beyond simple descriptive statistics, a second-degree polynomial regression model ($y = 1.34x^2 + 2.5x + 75$, $R^2 = 0.94$) was applied to the annual production data. A robustness check comparing this to a linear fit ($R^2 = 0.81$) confirmed that the polynomial model significantly better captures the accelerating, non-linear growth trend of the discipline. The trend line indicates not just steady growth, but an exponential expansion phase beginning in 2021, statistically forecasting over 300 annual publications by 2027. This mathematically confirms that computational ethnobotany has officially entered its rapid expansion phase within the technology adoption lifecycle (Fig. 2).

From an impact perspective, the 1,275 documents have accumulated approximately 22,720 citations, yielding a high average of 17.82 citations per document. This metric is striking given the relatively young document average age of 3.45 years, suggesting that quantitative ethnobotany is a highly dynamic and rapidly cited methodological frontier.

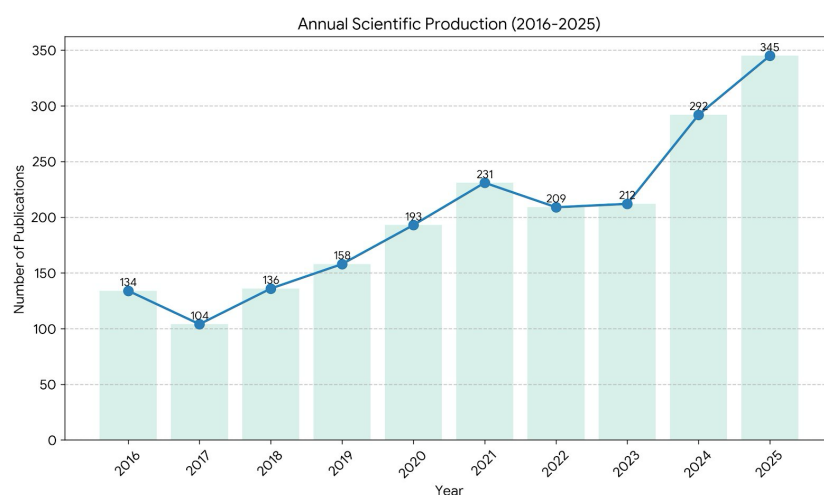


Figure 2. Annual scientific production and forecasting of publications integrating quantitative methods in ethnobotany (2016–2025).

This methodological renaissance is further accelerated by the global push towards open science and the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles (Wilkinson *et al.* 2016). The proliferation of open-access phytochemical databases, seamlessly integrated with digitized ethnobotanical repositories, has democratized access to raw datasets. Consequently, researchers in emerging economies can now execute highly sophisticated, data-driven bioprospecting studies using open-source algorithms, drastically reducing the historical reliance on expensive wet-lab infrastructure.

Furthermore, the integration of geographic information systems (GIS) with advanced spatio-temporal data analysis represents an untapped frontier in ethnobotany. While classical mapping provides static distributions of medicinal plants, the application of advanced spatial clustering algorithms, such as density-based spatial clustering of applications with noise (DBSCAN) and its spatio-temporal variants, allows researchers to dynamically track the geographical erosion of traditional ecological knowledge alongside shifting climate zones. These spatial quantitative tools enable targeted conservation strategies by identifying geographic 'hotspots' where high cultural use intersects with high ecological vulnerability.

Furthermore, predictive algorithms are transforming ethnobotany into an in-silico science. Machine learning approaches, such as random forest classification, are trained on ethnobotanical datasets to predict novel pharmacological activity. Coupled with network pharmacology (Hopkins 2008), this systems biology approach formally validates the polypharmacological synergy claimed by traditional healers, fundamentally reducing the financial costs of blind drug screening.

Institutional Productivity: Unpacking Global Power Dynamics

Country-level analysis reveals a geographically diverse research landscape heavily dominated by emerging economies (Table 2). Megadiverse nations with deep-rooted traditional medicine systems, such as India (242 docs) and China (189 docs), proudly lead the global scientific production. However, deeper interpretation reveals that this dominance is not merely an artifact of local biodiversity; it directly reflects targeted national funding policies. Giant state-backed institutions, particularly the Chinese Academy of Sciences (CAS) and India's Council of Scientific and Industrial Research (CSIR), dominate the dataset. These institutions operate under explicit government mandates to computationalize and modernize Traditional Chinese Medicine (TCM) and Ayurveda for global pharmaceutical markets. They provide the crucial funding and advanced supercomputing infrastructure required to execute massive, data-intensive network pharmacology projects.

Examining the topological structure of the international collaboration network (Fig. 4) reveals specific geopolitical research dynamics. The primary red cluster is driven by Sino-Pakistani and Indian collaborations. A secondary green cluster highlights strong transatlantic partnerships, where U.S. and European institutions (acting as methodological hubs) collaborate extensively with Brazilian and Indonesian universities (acting as biodiversity data providers). This structural shift from historic "helicopter science" to integrated, cross-continental computational partnerships is absolutely vital for equitable knowledge governance.

Table 2. Top contributing countries by affiliated document count, highlighting the dominance of emerging megadiverse economies.

Country	Documents
India	242
China	189
Pakistan	152
Brazil	150
Indonesia	148
United States	146
Italy	124
Ethiopia	101
South Africa	99
Morocco	80

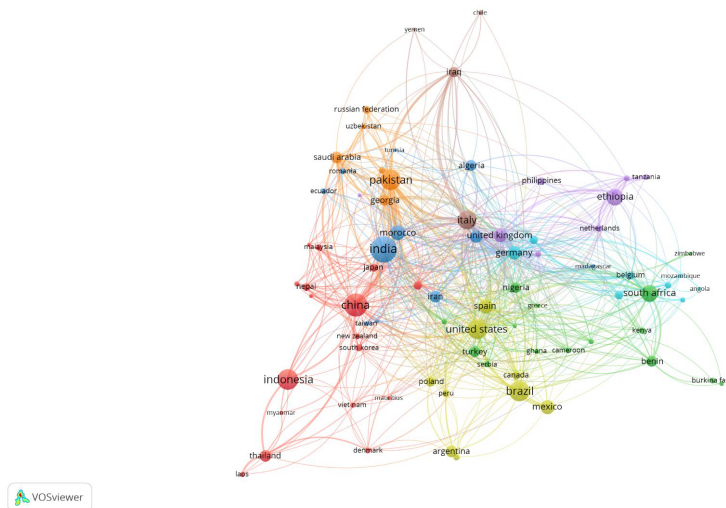


Figure 4. VOSviewer visualization of the international co-authorship network, highlighting geopolitical clusters and equitable bridges between the Global North and South.

Limitations

While this systematic review identifies critical trends, specific methodological limitations must be acknowledged. First, relying exclusively on the Scopus database introduces a structural bias against highly relevant quantitative ethnobotanical studies published in local, non-indexed journals, particularly in languages like Spanish or Portuguese in Latin America. Second, algorithmic limitations exist within the bibliometric clustering itself; VOSviewer's fractional counting assumes equal weight distribution across co-authors, which may obscure the true principal investigators driving the computational methodologies.

Additionally, while the polynomial regression provides a statistically robust forecasting model, it inherently relies on historical trajectory continuity. Consequently, it does not account for potential external macroeconomic shocks, such as sudden shifts in international research funding policies or global geopolitical disruptions, that could alter the projected growth. Finally, while the literature shows a surge in machine learning applications, many of these computational models remain purely theoretical, suffering from a lack of rigorous, independent in-vitro validation.

Table 3. Top 10 most productive journals publishing quantitative ethnobotanical research.

Journal	Documents
Journal of Ethnobiology and Ethnomedicine	210
Ethnobotany Research and Applications	180
Journal of Ethnopharmacology	156
South African Journal of Botany	52
Biodiversitas	48
Genetic Resources and Crop Evolution	47
Frontiers in Pharmacology	43
Economic Botany	41
PLoS ONE	38
Plants	20

Conclusion and Future Work

This interdisciplinary bibliometric analysis explicitly demonstrates that the integration of statistical and data science methods into ethnobotany is successfully transitioning the field from a purely descriptive science into a highly predictive, analytical discipline. Driven by a robust 12.05% annual growth rate, the literature reveals a clear methodological evolution: moving from the calculation of basic, survey-based quantitative indices toward complex computational network pharmacology, advanced temporal regression forecasting, and automated text mining.

The empirical evidence suggests that future research must transcend isolated algorithmic applications. To solve the current crisis of heterogeneous, non-comparable datasets, we propose the development of an open-source, multi-layered "Computational Ethnobotany Ontology" (CEO). This explicit framework would integrate three distinct data layers: (1) standardized botanical taxonomy (e.g., linking to Kew's Plants of the World Online), (2) phytochemical structures and bioactivity targets (linking to PubChem), and (3) standardized traditional use categories (utilizing the Economic Botany Data Collection Standard). By structuring these variables within a machine-readable Semantic Web framework (such as RDF/OWL), researchers could execute complex, cross-cultural data-mining algorithms that are currently impossible.

Finally, the integration of advanced algorithms must be carefully aligned with international legal frameworks, particularly the Nagoya Protocol (Secretariat of the CBD 2011). As machine learning models increasingly mine indigenous knowledge to predict patentable bioactive compounds, explicit computational mechanisms for tracking data provenance and ensuring equitable benefit-sharing are urgently required. Coupling this explicit computational framework with ethical indigenous data governance will allow ethnobotany to fully realize its potential in accelerating global biodiversity conservation and sustainable drug discovery.

Moreover, researchers must actively address the algorithmic bias inherent in machine learning models trained on historical ethnobotanical texts, ensuring that predictive modeling does not disproportionately marginalize unrecorded or orally transmitted knowledge systems.

To operationalize the *Computational Ethnobotany Ontology* (CEO), we propose a systematic, four-tiered conceptual framework:

Tier 1: Knowledge Acquisition (Data Inputs)

Consolidating raw, heterogeneous datasets, including primary ethnobotanical field surveys, digitized historical literature, geographic distribution coordinates (GIS data), and orally transmitted indigenous knowledge.

Tier 2: The Integration Framework (Core Ontologies)

Structuring the raw data into machine-readable Semantic Web formats (RDF/OWL) across three standardized layers:

- a. *Taxonomic Layer*: Linked to global databases (e.g., Kew's Plants of the World Online).
- b. *Phytochemical Layer*: Linked to molecular and bioactivity repositories (e.g., PubChem).
- c. *Utilization Layer*: Standardized using the Economic Botany Data Collection Standard (EBDCS).

Tier 3: Advanced Computational Analysis (The Engine)

Deploying algorithmic tools on the integrated ontology, such as Machine Learning for predictive modeling, Network Pharmacology for identifying multi-target drug synergies, and Spatial Clustering to map the erosion of ecological knowledge.

Tier 4: Ethical Application and Outcomes (The Goal)

Translating analytical outputs into real-world solutions. This includes accelerating sustainable drug discovery, designing spatially targeted biodiversity conservation strategies, and embedding Nagoya Protocol compliance mechanisms to ensure equitable benefit-sharing and data provenance tracking for indigenous communities.

Declarations

List of abbreviations: AU – Authors; AB_TM – Abstract Terms; DE – Author Keywords; UV – Use Value; ICF – Informant Consensus Factor; CI – Cultural Importance Index; RFC – Relative Frequency of Citation; PCA – Principal Component Analysis; NLP – Natural Language Processing; GIS – Geographic Information Systems; QSAR – Quantitative Structure-Activity Relationship; CBD – Convention on Biological Diversity; TEK – Traditional Ecological Knowledge.

Ethics approval and consent to participate: This study is a bibliometric analysis of publicly available published literature; no primary human or animal data were collected. Ethics approval was not required.

Consent for publication: Not applicable.

Availability of data and materials: The original bibliographic dataset analyzed in this study was retrieved from Scopus and is available from the corresponding author upon reasonable academic request.

Competing interests: The authors declare no competing interests.

Funding: This research was funded by the Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan / LPDP), Ministry of Finance, Republic of Indonesia.

Author contributions: Syella Zignora Limba: Conceptualization, Methodology, Data Curation, Writing – Original Draft. I Gede Awantara: Software, Formal Analysis, Visualization. Boy Riansyah: Validation, Investigation, Writing – Review & Editing. Muh

Akbar Idris: Supervision, Project Administration, Resources, Final approval of the version to be published. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to express their sincere gratitude to the School of Data Science, Mathematics, and Informatics (SSMI), IPB University, for providing the necessary institutional facilities, academic environment, and technical support to conduct this interdisciplinary research.

Literature cited

Agesti ARA, Ariyanti NS, Chikmawati T, Purwanto Y. 2023. Ethnobotany of food plants used by Minangkabau Community in Lima Puluh Kota District, West Sumatra, Indonesia. *Biodiversitas Journal of Biological Diversity* 24(5):2756–2767. doi:10.13057/biodiv/d240529.

Benamar K, Benelfakir K, Benamar S. 2026. Ethnobotany in Morocco from 1990 to 2023: Part I - A critical analysis of researchers' contributions, bibliometric, methodological attributes, and the sociodemographic characteristics of the surveyed population. *Ethnobotany Research and Applications* 33:1–20. doi:10.32859/era.33.1.

de Albuquerque UP, Nascimento ALB, Soldati GT, Feitosa IS, Rodrigues da Silva A, Cruz da Rocha J. 2013. Ethnobotany: one concept and many interpretations. In: de Albuquerque UP, Alves RRN. (eds). *Recent Advances in Ethnobiology Research*. NUPEEA, Recife, Brazil, Pp. 87–99. doi:10.1007/978-1-4614-7804-0_8.

Erlina L, Paramita RI, Kusuma WA, Fadilah F, Tedjo A, Pratomo IP, Ramadhanti NS, Nasution AK, Surado FK, Fitriawan A, Istiadi KA, Yanuar A. 2022. Virtual screening of Indonesian herbal compounds as COVID-19 supportive therapy: machine learning and pharmacophore modeling approaches. *BMC Complementary Medicine and Therapies* 22(1):207. doi:10.1186/s12906-022-03686-y.

Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. 2021. How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research* 133:285–296. doi:10.1016/j.jbusres.2021.04.070.

Hopkins AL. 2008. Network pharmacology: the next paradigm in drug discovery. *Nature Chemical Biology* 4(11):682–690. doi:10.1038/nchembio.118.

Leonti M, Cabras S, Weckerle CS, Solinas MN, Frau F. 2010. The causal dependence of present plant knowledge on herbals: Contemporary medicinal plant use in Campania (Italy) compared to Matthioli (1568). *Journal of Ethnopharmacology* 130(2):379–391. doi:10.1016/j.jep.2010.05.021.

Luczaj L, Pieroni A, Tardio J, Pardo-de-Santayana M, Soukand R, Svanberg I, Kalle R. 2012. Wild food plant use in 21st century Europe: The disappearance of old traditions and the search for new cuisines involving wild edibles. *Acta Societatis Botanicorum Poloniae* 81(4):359–370. doi:10.5586/asbp.2012.048.

Martin GJ. 1995. *Ethnobotany: A Methods Manual*. Chapman & Hall, London, U.K. doi:10.1007/978-1-4615-0639-2.

Phillips O, Gentry AH. 1993. The useful plants of Tambopata, Peru: I. Statistical hypotheses tests with a new quantitative technique. *Economic Botany* 47(1):15–32. doi:10.1007/BF02862203.

Secretariat of the Convention on Biological Diversity. 2011. Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity. Secretariat of the Convention on Biological Diversity, Montreal, Canada.

Thomford NE, Senthebane DA, Rowe A, Munro D, Seele P, Maroyi A, Dzobo K. 2018. Natural products for drug discovery in the 21st century: Innovations for novel drug discovery. *International Journal of Molecular Sciences* 19(6):1578. doi:10.3390/ijms19061578.

van Eck NJ, Waltman L. 2014. CitNetExplorer: A new software tool for analyzing and visualizing citation networks. *Journal of Informetrics* 8(4):802–823. doi:10.1016/j.joi.2014.07.006.

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3:160018. doi:10.1038/sdata.2016.18.