The Journal of Agricultural Science

cambridge.org/ags

Millet research status and prospects for alleviating food insecurity through a text-mining approach

Jagajjit Sahu, Tilak Chandra, Sarika Jaiswal, Mir Asif Iquebal and Dinesh Kumar

Division for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India

Climate Change and Agriculture Research Paper

Cite this article: Sahu J, Chandra T, Jaiswal S, Iquebal MA, Kumar D (2023). Millet research status and prospects for alleviating food insecurity through a text-mining approach. *The Journal of Agricultural Science* **161**, 633–644. https://doi.org/10.1017/S0021859623000618

Received: 1 August 2023
Revised: 2 November 2023
Accepted: 12 November 2023
First published online: 7 December 2023

Keywords:
bio-concept; international year of millet; keyword mining and evolution; natural language processing; probiotics; PubMed

Corresponding author:
Mir Asif Iquebal;
Email: ma.iquebal@icar.gov.in

Abstract

In view of the celebration of the 'International Year of Millets,' millets are popularizing sustainable agricultural output amid challenging climates and nourishing adequately as food and feed. The extent of scientific intervention is the foundation for designing, promoting and popularizing neglected crops on social platforms. Planning future directions and adaptive strategies largely require regular evaluation of research efforts to identify hotspots and research gaps, as identified in the present study by creating a robust text-mining approach that integrates scientometrics using PubMed citation data. Keyword mining reveals that India and China are the leading publication centres on millets, possibly due to their large proportion of cultivation and indigenous nature. It further reveals that the pearl millet is the predominant one, followed by foxtail and finger millet, suggesting that most research is confined to them only; however, other millets, still have a research gap in comparison. The word abiotic stress is associated with high frequency in millet research due to its adaptive nature amid climate change. Thematic representation explored the novel concept of millet's utility as a probiotic and millet bran to ensure nutrient–cereal properties based on the persistency of keywords throughput research progression; however, incurious consumption is associated with harmful ochratoxin. Bio-concept mining and knowledge graph generation divided the millet research output into four large domains, which provides a largely covered bio-concepts for millet research and co-concurrence of emerging bio-concepts to intense progress and finds an adequate literature gap to improve millet research for sustained growth and equilibrate biodiversity.

Introduction

Underutilized and neglected food crops, particularly millets, play a key role in ensuring the population's access to food and nutrition, both for humans and animals (Hassan *et al.*, 2021). These are non-commodity crops that are a part of a vast, bio-diverse community that includes thousands of domesticated, semi-domesticated, or wild species (Vetriventhan *et al.*, 2020; Saini *et al.*, 2021). These are also inadequately utilized crops that harbor beneficial plant species that researchers, breeders and politicians either completely ignore or push to the sidelines (Dayakar Rao *et al.*, 2021). While research helps to increase the productivity of millets, state policies and programs have an impact on their distribution and production (Krishnan *et al.*, 2021). Re-strategizing crop improvement and agronomic practices of the millet crops would help to identify climate-resilient varieties with improved grain attributes because the crops are primarily globally adapted (Muthamilarasan *et al.*, 2016; Vetriventhan *et al.*, 2020; Chaturvedi *et al.*, 2022). A plethora of factors associated with millet growth and development led to the development of hierarchical research-based improvement in the form of publication; however, their systemic conclusion has been awaited and would have a variable lag behind other staple crops. In response to calls for papers issued by specialized and esteemed journals in the field of crop improvement, research on millets is a response that has a strong connection to Sustainable Development Goals (SDGs) 1, 2, 3, 8 and 15 (https://www.fao.org/millets-2023/en). The current study was conducted in consideration of the SDGs to examine the research activity on millet improvements. Although research on millet crops has not significantly improved their socioeconomic status, further knowledge-based improvement management tactics could enhance millet improvement in multifaceted ways (Muthamilarasan *et al.*, 2016).

To determine the volume and growth trend of publications focusing on millet enhancement, a text-mining analysis from published abstracts is an appropriate methodology (Cooper *et al.*, 2020; Tao *et al.*, 2020; Thakur and Kumar, 2022; Adelabu and Franke, 2023). Such analysis was utilized to identify significant research themes, growth patterns across the world, active research subdomains and research institutions for upcoming financing and planning (Bakhtin *et al.*, 2020; Adelabu and Franke, 2023; Andrade Pereira and Mugnaini, 2023). The subject area of text mining is concerned with quantifying and evaluating the

© The Author(s), 2023. Published by Cambridge University Press

bibliometric analysis of scholarly publications (Andrade Pereira and Mugnaini, 2023). It reveals that the measurement of the effect of academic journals and research publications, comprehension of scientific citations and application of such measurements in management and policy contexts are all significant research concerns (Zhong *et al*., 2019; Bakhtin *et al*., 2020). The world's major cultivated staple crops, viz., wheat, rice and maize, provide about 60–70% of calories and nutrition (Luo *et al*., 2020; Palacios-Rojas *et al*., 2020; Dhaliwal *et al*., 2022). The rise in staple crop yield over the past century can be attributed to adequate knowledge of the use of cutting-edge tools and management techniques in genetic control of agronomic traits of crops (Kaur *et al*., 2021). Similar efforts can create exceptional yields from millets, which have the potential to significantly increase food production, are resilient to climate change, have a rich nutritional content, have a high capacity to reduce pest and disease infestation and are resistant to a variety of environmental factors (Muthamilarasan *et al*., 2016; Bakhtin *et al*., 2020; Vetriventhan *et al*., 2020; Hassan *et al*., 2021). A diminutive amount of information about the trajectory of millet improvement was found in a literature search utilizing the well-known database and the search engine 'PubMed'. The current study uniquely combines bibliometric analysis and a tailored text-mining workflow to comprehensively investigate the research topic. Herein, the objectives were set to enhance the understanding of millet research by developing an efficient text-mining method that integrates scientometrics and named entity recognition using PubMed citation data. The current work is primarily oriented towards offering a holistic perspective of the millet research landscape using the term 'Millet' rather than delving into individual millet species. The results were achieved by collecting data from 'PubMed' along with their integration with bibliometrics analysis, followed by keyword and bio-concept mining, KG generation and visualization. The primary goal was to establish a robust and authentic research methodology, leading us to choose PubMed as the most suitable data source. The current study aims to employ text-mining on PubMed citation data to analyse the available literature for research on millet to date to gain a better understanding of the state of knowledge in the field and identify areas where further research is requisite.

Materials and methods

Data collection

The citation data from 'PubMed' was employed for the analysis, which was collected using the search term ('millet'[Title/Abstract] AND [English] [Filter]) on February 13, 2023. The database, 'PubMed' was chosen as it is one of the largest literature databases and at the same time, it contains data for a wide range of journals. The retrieved records were saved in 'PubMed text format' as a file. The records for the year 2023 were removed as they provided a partial picture of the current year. Each of the collected records contains the title, abstract, keywords and other associated information such as authors, affiliations, year of publication and pages.

Bibliometric analysis and visualization

To investigate the research productivity and publication trends across years, countries and more, the package bibliometrix (Aria and Cuccurullo, 2017) was used. As 'PubMed' citation data does not provide citation data, citation-specific analyses were excluded. Using the bibliometrix in R, the citation data was converted to a data frame and saved into a comma-separated value (.csv) file for further analysis. The bibliometric analysis was then performed on the data frame, and information about various bibliometric indicators was obtained. To view the year-wise publication trend, a bar plot was generated. Further, the productivity of individual countries with respect to the collected data was analysed and plotted.

Keywords-based topic analysis

To gain a broad understanding of the main themes or topics present in millet research, a keyword-based topic analysis was followed. The field 'author's keywords' was chosen as the mesh keywords were partially available for all the records. The analysis was performed to identify the top keywords and their growth, as well as attempted to provide keyword-based themes and their evolution over time following the methods of (Cobo *et al*., 2011). To further analyse the relationships between keywords, network analysis techniques were used to construct co-occurrence networks. These networks represent the relationships between keywords based on their co-occurrence in the same publications.

Bio-concept mining and KG generation

Bio-concepts were extracted with the help of the PubTator tool (Wei *et al*., 2019), which provide automated annotation of the 'PubMed' abstracts under ten entity classes, which are species, chemical, disease, genes, protein mutations, cell lines, chromosomes, protein acid changes, DNA mutations and RefSeq. The PubMed IDs were uploaded onto the PubTator's collections manager, and the annotations were downloaded. The file was saved in PubTator format on the local system. With the help of R scripts, the entity information was extracted and saved as a comma-separated value file. Following this, networks were constructed between entities based on the co-occurrences in the same abstracts. Furthermore, the top five entities were extracted for four classes, which are chemical, disease, gene and species.

Results

Basic bibliometrics and publication trend

The search for millet on PubMed retrieved a total of 3186 records which were downloaded in the PubMed format from the save option on PubMed. After converting the data into a data frame format, the 65 records, for 2023 were removed and the remaining 3121 records were considered for further analysis. Table 1 provides an overview of various bibliometrics information that delivers basic information regarding research productivity and trends. The very first record on PubMed for millet was from the year 1803.

Figure 1 shows a bar plot for the year-wise publication trends, which is very clearly a growing trend with very slight crisscrosses. The highest number of articles was in the year 2022, which was 379 in number. Not all the years are present in the graph, depicting that a total of 151 years have no records. The availability of at least one record for continuous years started in 1968.

Country-wise productivity evaluation

The analysis with respect to the productivity of individual countries as well as collaborations was done to get an understanding of

https://doi.org/10.1017/S0021859623000618 Published online by Cambridge University Press

Table 1. Basic keywords information obtained from the citation data extracted from 'PubMed' after searching the keywords 'Millet'

Parameters	Value
Timespan	1803–2022
No. of articles	3121
No. of sources (journals, books, etc.)	720
Annual growth rate %	2.75
No. of keywords	6876
Single-authored documents	146
Documents per author	0.305
Co-authors per documents	5.57
International co-authorships (%)	18.52

An inconsequential pattern of publication was observed before 1971, and there has been upward trend in the publication trends since the last few decades.

patterns of publications in millet research over time (Fig. 2). India was found to be the top country, with a total of 828 articles, followed by China and the USA with 621 and 391 articles, respectively (Fig. 2(a)). Among the top ten countries, no country had an article on millet before the year 1971 (Fig. 2(b)). Also, India had its first article in 1971, followed by Australia in 1972 and the USA in 1979. There was an increase in the publication of the top three out of the top ten countries over time. The other seven countries in the list showed a rather steady number of articles. Considering a collaboration to be a research article with more than one affiliation, India was found to have the highest number of 1984 instances (Fig. 2(c)). China and the United States came in second and third, with 1493 and 1116 participants, respectively. However, the United States has the highest total of 334 collaborations in 71 intercountry collaborations, which could be interpreted in simple terms as the United States has collaborated with 71 countries in total. India and China were found to have collaborated with 68 and 54 countries, with 328 and 251 collaborations, respectively. India is in the second position when it comes to the number of countries it has collaborated with as well as the total count of collaborations in intercountry collaborations. However, China falls to the sixth position below France and Germany with 58 and 57, respectively, in terms of the number of countries it has collaborated with. France and Germany are in the fourth and fifth places in terms of total collaborations, which are 480 and 346, and the total number of international collaborations, which are 196 and 162. Interestingly, Australia, being in the seventh place of total collaboration with 270, has collaborated with only 45 countries, however, with a total of 124 international collaborations. The collaboration network very clearly shows that the three countries – India, China and the United States– have surpassed all other countries in terms of all types of collaborations.

Keyword mining and its evolution

A total of 6876 keywords were extracted from the citation data and processed to find their distribution and co-occurrence across time. A total of 5659 unique keywords were present, which might have partial overlaps as well as lexicological similarities. The analysis, however, targeted to the most frequent keywords as they are of the utmost importance towards deriving the patterns. Figure 3, represents the distribution of the top 20 most frequent keywords across the time (Fig. 3(a)). Similarly, the most frequently occurring keywords were identified and highlighted as foxtail, pearl, stress and analysis (Fig. 3(b)). It is clearly visible that there is an intersection between the important keywords based on frequency and associations.

Thematic representations-based roadmaps

In qualitative research, a thematic analysis is used to find, analyses and report the patterns within the studied topic with the goal of examining themes for a targeted query (Cooper *et al.*, 2020). To understand the themes of the publications in millet research, the keywords helped in building themes and the evolution of these themes has been presented as a Sankey plot (Fig. 5). In millet research, the evolution of keyword themes has revealed interesting facts since the last century. The plot reveals and highlights the admixture history of millet evolution, cereal grain, millet bran and ochratoxin in prehistoric events. Following decades, there has been a shift in keyword trends and evolution that signifies the millet utility of probiotics, cover crops, ochratoxin and integration with next-generation sequencing. The latest research results have been confined to millet bran and starch, millet as a probiotic and its association with potential harmful ochratoxin.

Bio-concept mining and KG generation

Among the 3121 PubMed IDs, there were 3086 IDs for which annotations were identified using PubTator. A total of 46 338 annotations were obtained from PubTator output, however, the data was further processed to get the following table (Table 2). The table contains the total number of annotations, the number of unique annotations for individual classes and the annotations for which there was no ID. The chemical class resulted in 2185 annotations without IDs, whereas all other cases matched with a unique ID. The highest number of total annotations was in case of species class, followed by chemical, disease and gene. Figure 6 represents the top five bio-concepts for each of the six selected classes, i.e., chemical, disease, gene and species.

The co-occurrence of the bio-concepts based on their mentions in the same PubMed records was converted into a graph object, which has been plotted as a network (Fig. 7). There were a total of 3205 bio-concepts, which are connected with 41 202 edges representing co-occurrence. The average degree and average weighted degree of the network were 25.711 and 44.126, respectively, with a modularity of 0.266. The topmost nodes were mainly from the class species, with human being the top node with a degree of 2271 and a weighted degree of 8899. Among the genes, *TNF-α* was found to be the top one with a degree and a weighted degree of 73 and 125, respectively. With a degree of 361 and a weighted degree of 1163, stress is the term with the highest frequency in terms of the class disease.

Discussion

Over the last decade, there has been a significant shift in the research landscape towards data-driven research (Bakhtin *et al.*, 2020; Adelabu and Franke, 2023). The trend towards data-driven research is likely to continue as the amount of data generated continues to grow, and computational tools become more powerful and accessible (Bakhtin *et al.*, 2020). With the exponential growth of scientific research publications, it has become increasingly challenging to keep up with the latest developments in a

https://doi.org/10.1017/S0021859623000618 Published online by Cambridge University Press

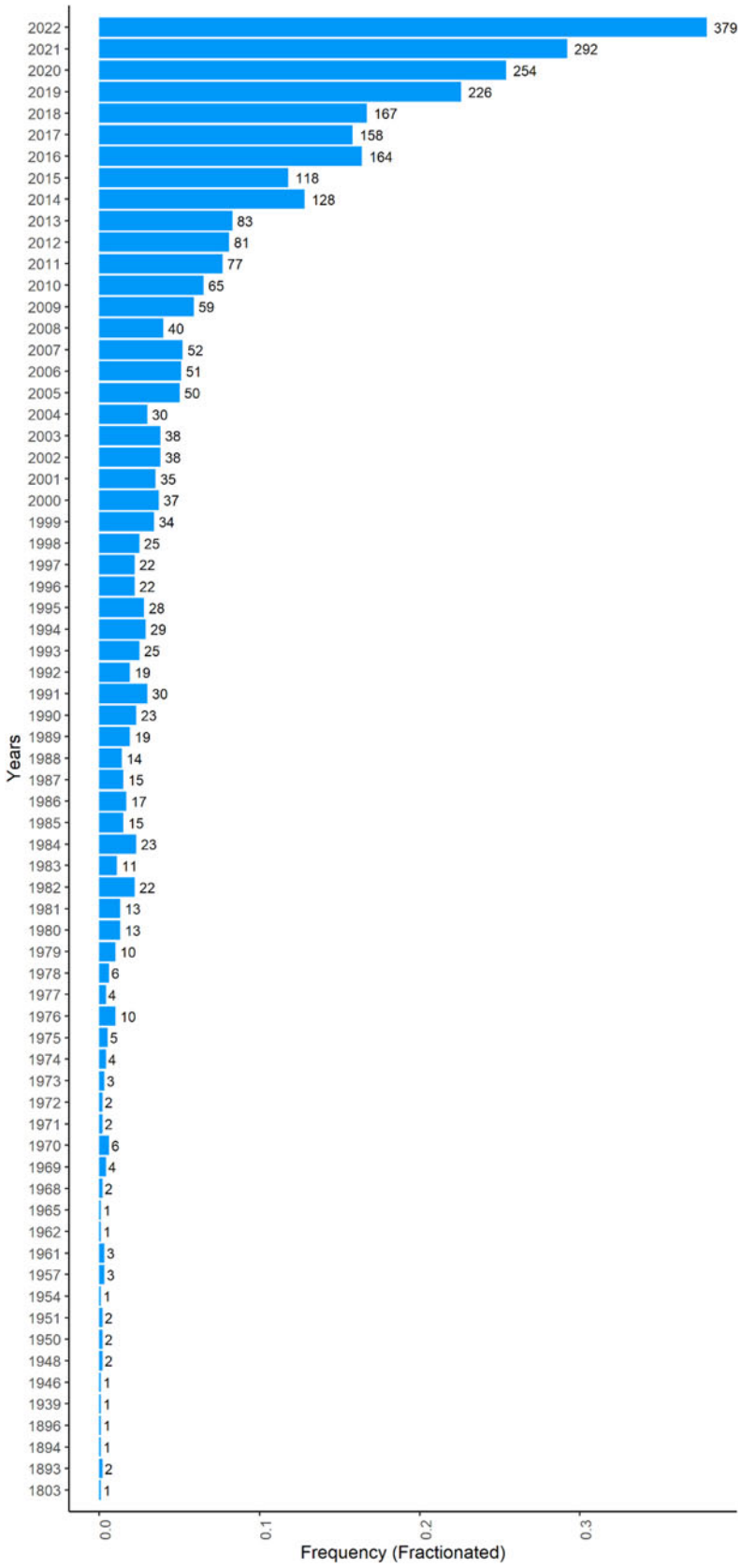

Figure 1. Year-wise publication trend on millet research to generate frequencies-based flow analysis. The frequencies were divided by thousand. Before 1971, a minimal pattern of publication was noticed, and throughout the last few decades, the trends have increased significantly.

particular field. Especially, going through such a large volume of texts manually has become very tedious. Text mining is a powerful method used to extract and analyse large volumes of text data (Bakhtin *et al.*, 2020; Cooper *et al.*, 2020; Tao *et al.*, 2020; Thakur and Kumar, 2022; Adelabu and Franke, 2023). It involves the use of natural language processing techniques and machine learning

https://doi.org/10.1017/S0021859623000618 Published online by Cambridge University Press

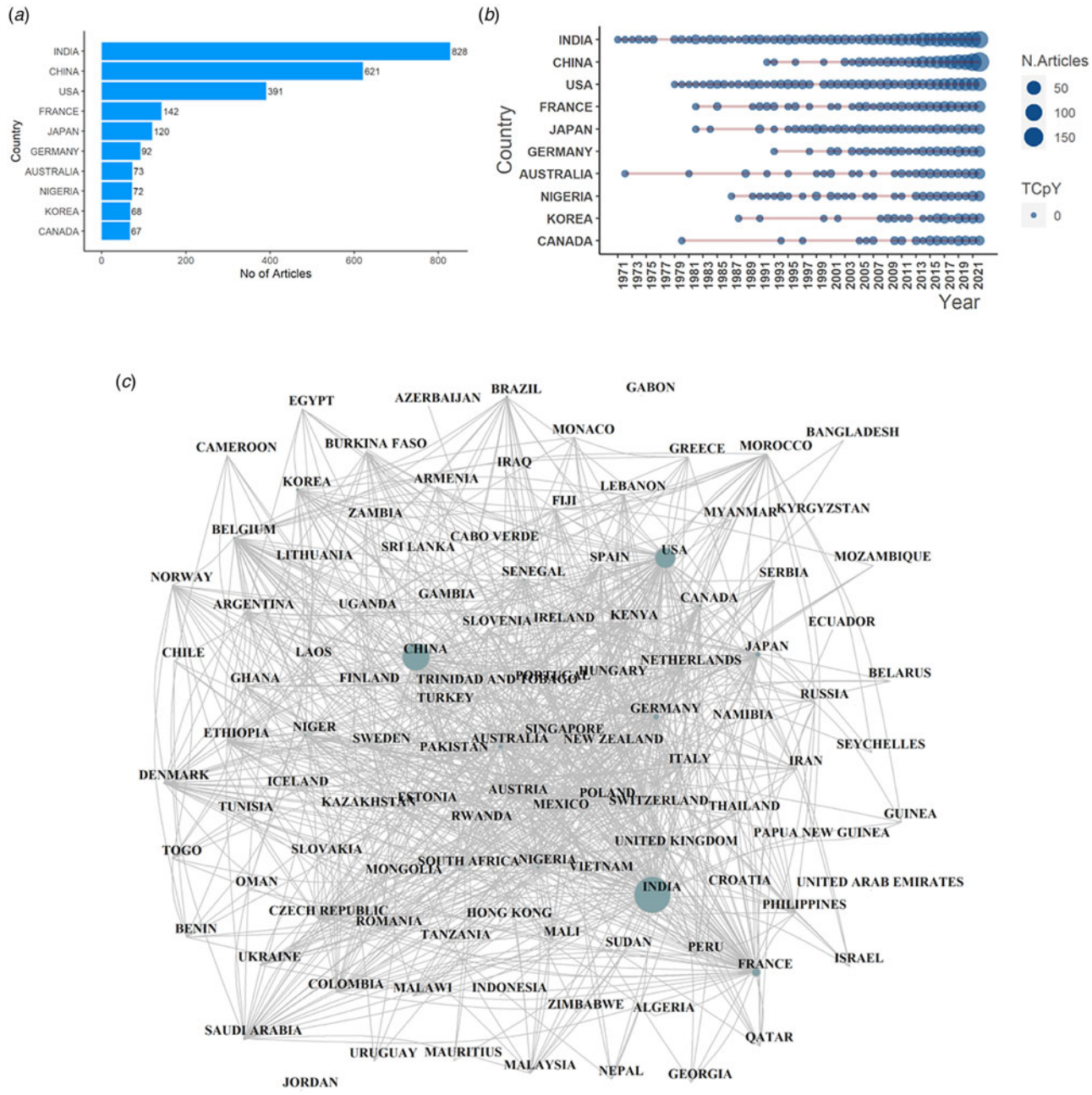

Figure 2. Country productivity with respect to the PubMed reports (a) Article frequencies against the top ten countries plotted as a bar plot (b) Publications of the same top ten countries evolved across the years. (c) A network depicting the country's collaboration as of 2022. The node size is directly proportional to the degree.

algorithms to identify patterns and relationships in scientific text data that might not be immediately apparent to humans (Adelabu and Franke, 2023). Furthermore, it facilitates the quick and efficient identification of key findings, relationships and knowledge gaps in a particular field, accelerating scientific discovery and facilitating the translation of research findings into real-world practices.

While evaluating potential data sources, including Web of Science and Scopus, it was found that one of the key components of the current research work, bio-concept mining using the PubTator tool, is exclusive to PubMed. Hence, the alternative databases that would have required significant workflow alterations, potentially impacting the study's precision and scope, were excluded. PubMed, a free online database of biomedical literature, contains over 32 million citations and abstracts from various scientific journals (Lu, 2011). In addition, it was selected for its capacity to provide an authentic and all-encompassing portrayal of the real-world scenario. As the largest repository of life science literature, it encompasses a wide array of studies, aligning seamlessly with the research objectives of the present study. These factors reinforce the rationale behind the decision to utilize PubMed and enhance the scientific rigour and relevance of the current study. In recent years, PubMed abstract text mining has become increasingly popular in the biomedical research community (Chen *et al*., 2017; Sahu, 2021; Gunturkun *et al*., 2022). Also, there has been an upsurge in reports regarding keywords mining

https://doi.org/10.1017/S0021859623000618 Published online by Cambridge University Press

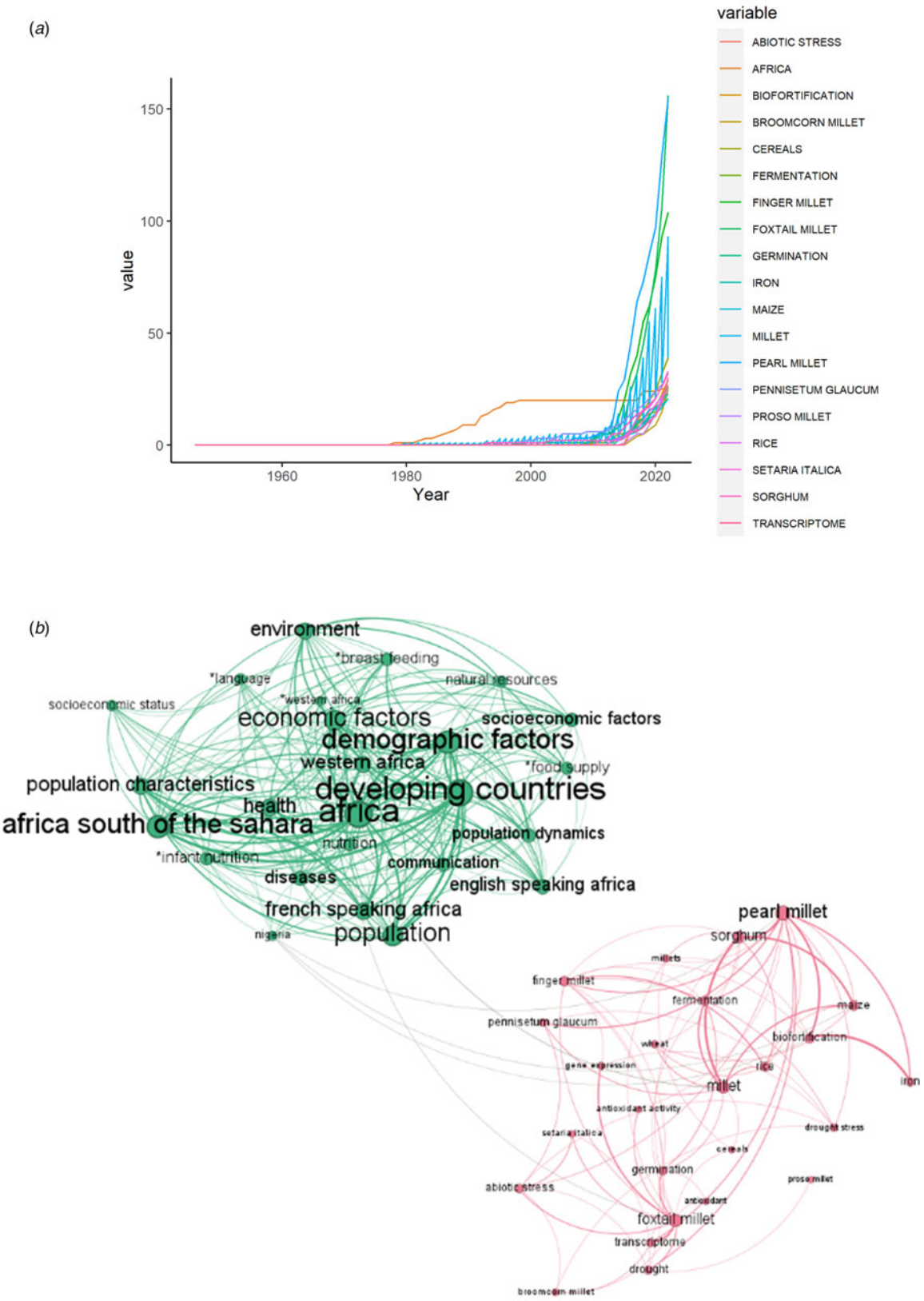

Figure 3. (a) Frequent keywords and their distribution across years. (b) Interconnected keywords based on their co-occurrence in the same articles. A timeline was also prepared for the keywords across years to have a look at the frequent keywords across individual years as well as a whole (Fig. 4).

https://doi.org/10.1017/S0021859623000618 Published online by Cambridge University Press

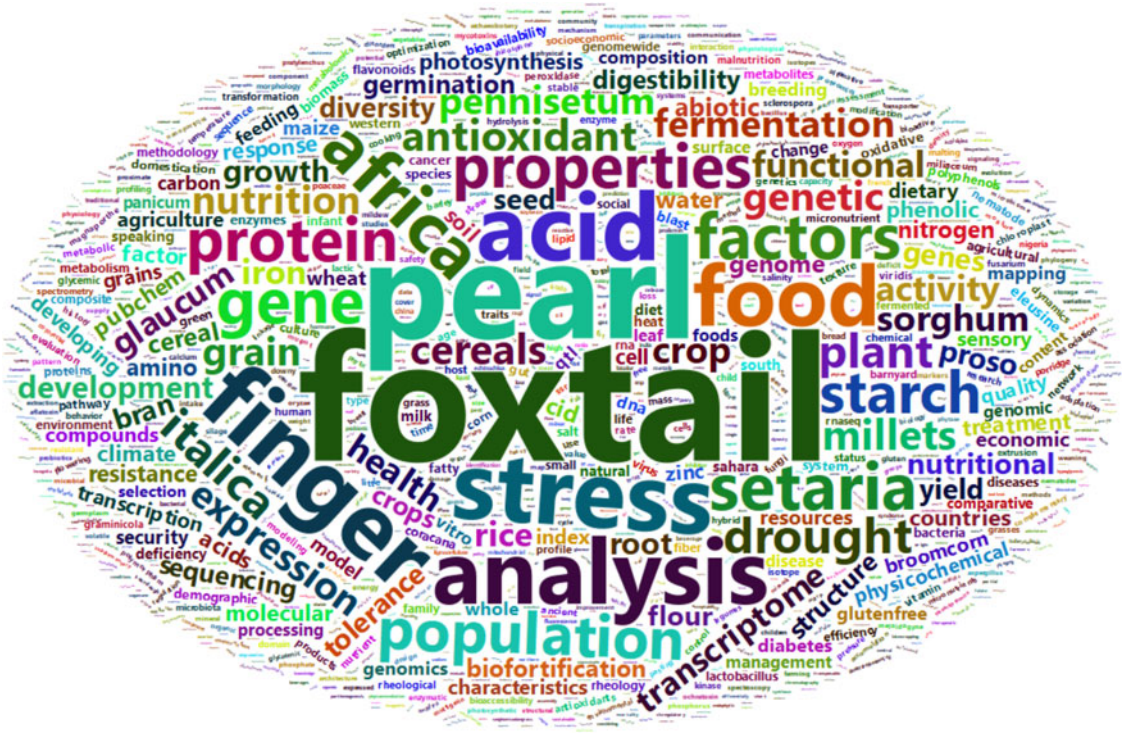

Figure 4. A word cloud representing the frequencies with the help of font size. These may contain partial keywords as only single word tokens have been considered to build the cloud.

as well as KG generation by text mining due to the advancement in computational approaches (Hickman *et al.*, 2022). Keywords analysis involves the quantitative analysis of scientific publications and their impact, whereas KG generation provides a more scientific view considering the bio-concepts captured through named entity recognition (Chen *et al.*, 2020; Adelabu and Franke, 2023).

Further, Fig. 1 suggests that an increasing trend has been observed in scientific research since 2010, and the highest number of articles was published in the year 2022. This trend is duly supported by the increased awareness of the healthfulness of miracle grains and the associated enhanced awareness among scientific society (Muthamilarasan *et al.*, 2016; Gowri and Shivakumar, 2020; Vetriventhan *et al.*, 2020). With a total of 828 articles, India was revealed to be the top country, followed by China and the United States with 621 and 391 articles, respectively (Fig. 2(a)). India is a top cultivator of millet with production per country (FAOSTAT, 2021), followed by Niger and China among the leading millet producers (Chandra *et al.*, 2021). Scientific progress is paradoxically related to millet cultivation and its production systems, but it is important to be aware of scientific improvement to enhance its productivity and yield potential. The Fig. 2(b), depicts the top ten countries that did not have an article on millet before the year 1971, as opposed to Fig. 1, where a few countries were found that had negligible but significant publications before 1971. The possible cause behind such a paradox is that there were no publications from top-selected countries before 1971. Similar trends were observed for research collaboration and partner countries involved in millet research, with a central hub depicting a high-frequency node that largely represents India, followed by China and the United States (Fig. 2(c)). The United States is the country with the highest number of international collaborations (334) with countries, among which China is the country with the highest number of collaborations (58), highest collaborations. India, in the second place, represents a total of about 328 international collaborations with 68 other countries, with a maximum frequency of (53) with the United States.

Keyword mining is the method of searching for a list of key terms or phrases that apply to a particular type of study to identify relevance. Top twenty keywords were selected in the archived records, the frequency of distribution across years was calculated (Fig. 3(a)), and found that from late 2015 onwards they exhibited peak responses for all category variables. It clearly suggests a trending role in millet publications since the last decade (Parthasarathy and Basavaraj, 2015; Naresh *et al.*, 2023). A sharp peak for the keyword 'millet', followed by 'finger millet' and 'foxtail millet,' was found since 2015, suggesting a high rate of publication related to these millets; however, the highly frequent role of the keyword 'abiotic stress' associated with millet research since 1980 suggests that millets are essential to sustain climate perturbation under minimal resources (Shivhare and Lata, 2016; Nadeem *et al.*, 2020). Other keywords represent a minor class of words largely associated with millets and other associated cereal research crops. Furthermore, interconnected keywords and their co-occurrences in the same articles clearly demonstrate that four keywords comprising 'millet,' 'Foxtail millet,' 'Finger millet,' and 'Pearl millet' are largely interconnected within the article text with high frequencies. It demonstrates an intense rate of research has been going on with these millets for food (Shankaramurthy and Somannavar, 2019), feed (Rao *et al.*, 2004) and accessible extrusion processing (Kharat *et al.*, 2019). A possible extension for keyword mining in millet research reflects a ten-year map with a timeline exhibiting the most common terms for each year (Fig. 4). A keyword cloud for all the years

https://doi.org/10.1017/S0021859623000618 Published online by Cambridge University Press

https://doi.org/10.1017/S0021859623000618 Published online by Cambridge University Press

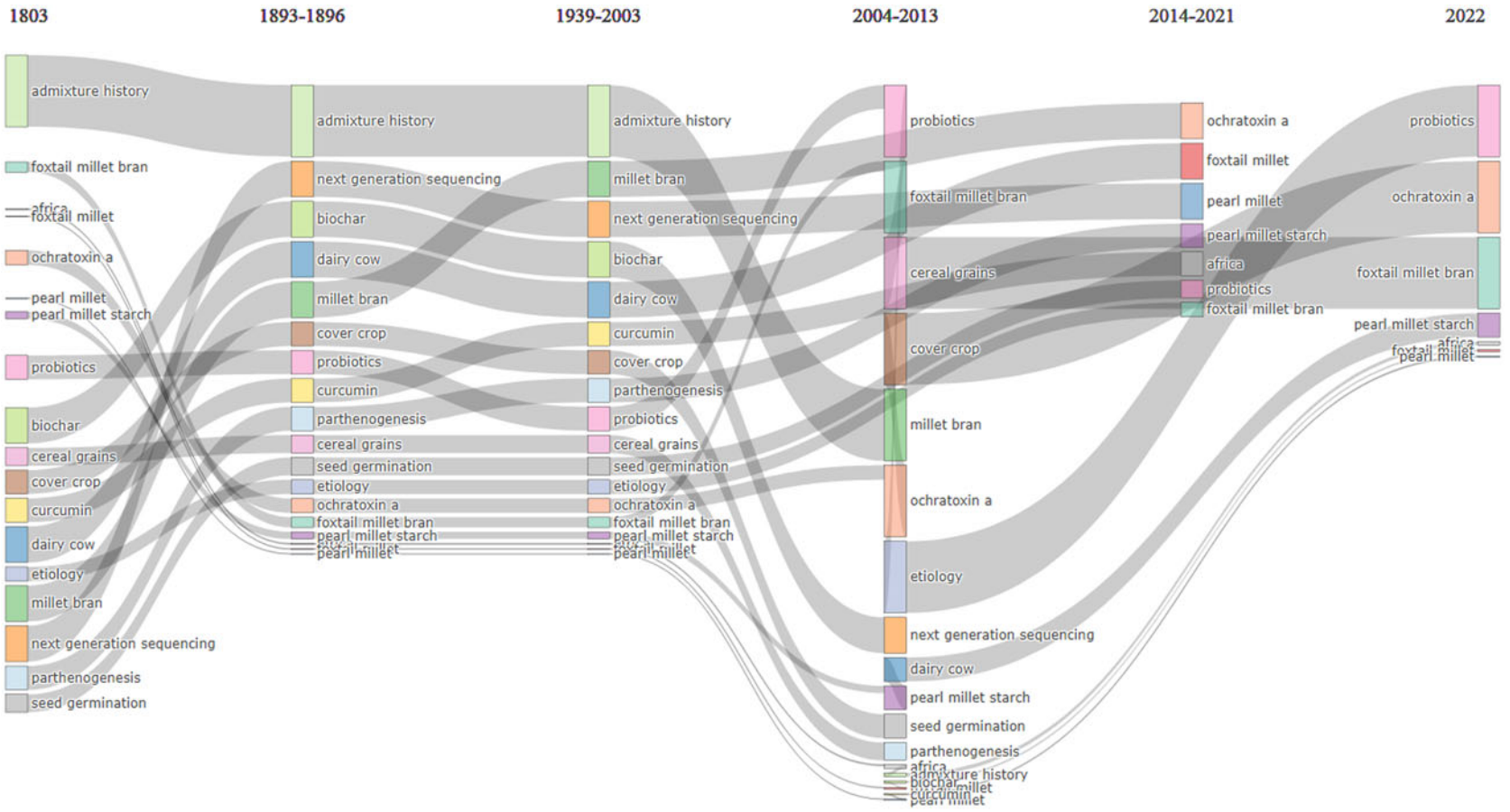

Figure 5. A Sankey plot to depict the evolution of keyword-based themes across different decades. The clusters are of different colours, but individual clusters are of the same colour across all six year ranges. The names of the clusters are basically the most frequent keywords in that cluster.

Table 2. Distribution statistics for bio-concepts across the ten classes' coverage reveals keyword cell line, chemical, chromosome, disease, DNA mutation, gene, protein acid change, protein mutation, RefSeq and species

Bio-concept classes	Total no. of bio-concepts	No. of unique bio-concept annotations	No. of bio-concepts without ID
Cell line	125	37	0
Chemical	14 591	1157	2185
Chromosome	36	10	0
Disease	3881	515	0
DNA mutation	2	2	0
Gene	919	284	0
Protein acid change	1	1	0
Protein mutation	21	19	0
RefSeq	1	1	0
Species	26 761	1180	0

combined is also present, with a similar trend to the pattern of keyword evolution during millet research (Rao *et al.*, 2004; Kharat *et al.*, 2019; Shankaramurthy and Somannavar, 2019).

To expand the understanding of the evolutionary trajectory of keywords, it was found that three keywords, namely 'Africa', 'Foxtail millet' and 'Pearl millet' were consistent throughout the hierarchy (Fig. 5). The higher frequency of these keywords is largely associated with major research carried out on these millets and domestication events (Vetriventhan *et al.*, 2020). Recent trends in keyword frequency like 'Probiotics' clearly suggest that millet could be the best replacement not only to alleviate hunger but also for bio-fortified foods made from millet, which can be considered potential sources of probiotics with significant health advantages (Di Stefano *et al.*, 2017; Budhwar *et al.*, 2020). Further, foxtail millet bran is largely associated with the production of fatty acids and antioxidants in millet-based secondary foods (Zhu *et al.*, 2018) and has anti-carcinogenic potential (Shan *et al.*, 2014). To ensure hygiene in food and feed, detection and proper eradication are necessitated to avoid them. Interestingly, the current analysis is not confined to the advantages of consuming millet but also includes several disadvantages likely imposed on health. The prevalence of ochratoxin in millet has widely increased over the last two decades (Fig. 5). Contaminated millet consumption can have severe implications due to the presence of ochratoxin, an abundant mycological contaminating agent (Makun *et al.*, 2013; Bui-Klimke and Wu, 2015). Proper assurance is necessary before preparation for millet-based foods (Zhang *et al.*, 2017).

Data gathered during scientific inquiries is frequently used to produce knowledge (Rossanez *et al.*, 2020). An enormous amount of data is produced by an expanding body of scientific research across diverse situations, and new information must be extracted from this data with the aid of computation (Wei *et al.*, 2019; Rossanez *et al.*, 2020). For millet research, the top four bio-concepts for each of the four classes, including chemical, disease, gene and species, were chosen (Fig. 6). Under the class of

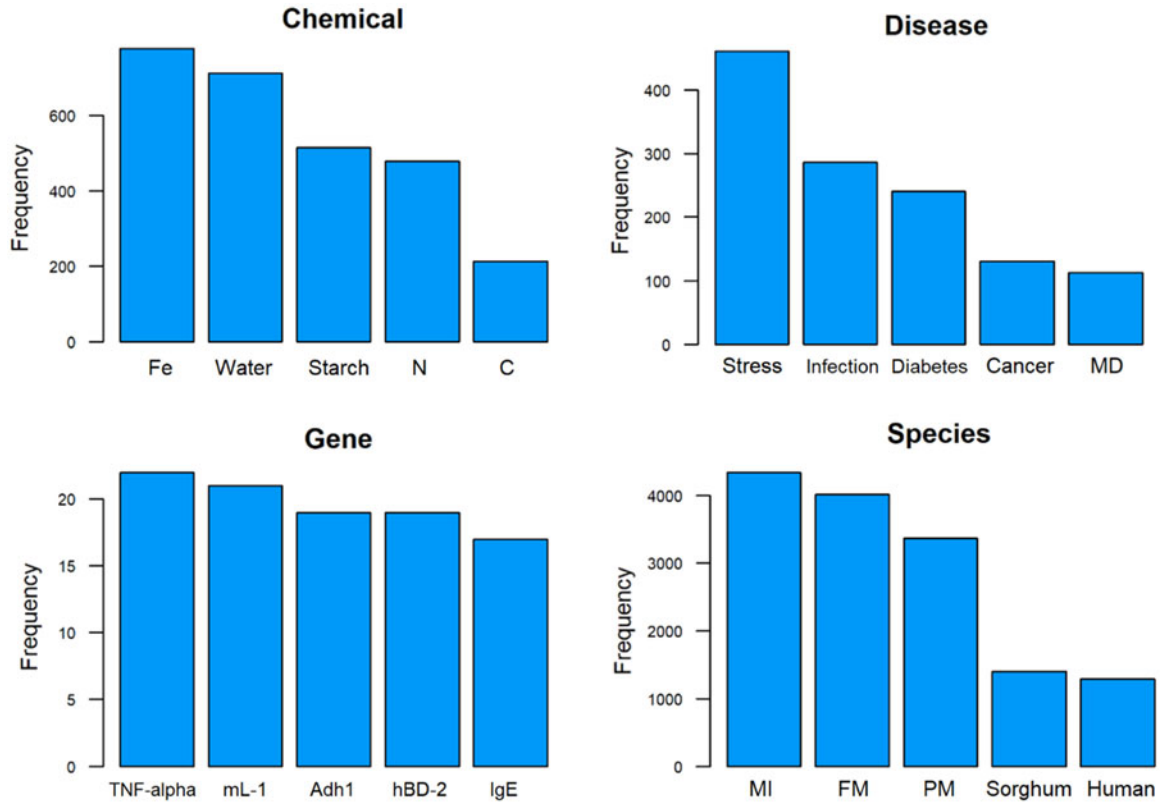

Figure 6. The top most frequent bio-concepts for the four important classes, which are chemical, disease, gene and species. The major compounds comes under chemical bio concept are iron, water, starch carbon and nitrogen. Similarly, under disease bio concept, the keyword stress, infection diabetes cancer and myotonic dystrophy (MD). For frequency-based gene bio concept, TNF-α, ml-1, Adh1, hBD-2 and IgE, respectively. The frequency of bio concept on species level provides the keyword millet (MI), foxtail millet (FM), pearl millet (PM), sorghum and human.

https://doi.org/10.1017/S0021859623000618 Published online by Cambridge University Press

Figure 7. A network representation of co-occurring top bio-concepts. Bio-concepts present in a single PubMed record are connected with edges to form the network, and then the sub-graph was created by considering only the top five bio-concepts falling under four classes, which are chemical, disease, gene and species.

chemicals, iron, water and starch are prevalent, as millets are rich sources of these identifiers (Vetriventhan *et al.*, 2020). Among broad classes of disease, the keywords stress, infection, diabetes and cancer are frequent, mostly interlinked with millets (Rotela *et al.*, 2021; Vetriventhan *et al.*, 2020). Among gene groups, a downgrading trend of terminology for *TNF-α, adh1, mL1, hBD2,* and IgE is widespread, and their interrelation with activation of the immune system induced by the anti-inflammatory and antioxidant properties of bioactive compounds from millets is largely reported (Budhwar *et al.*, 2020; Liu *et al.*, 2021; He *et al.*, 2022). The largely covered broad class of species includes millets, foxtail millet and pearl millet, which proportionally gain weight over sorghum and humans (Fig. 6). Following the further extension, a network representation of co-occurring bio-concepts revealed similar trends in network association among these words (Fig. 7). Millets represent a central node interconnected with foxtail millet, sorghum and pearl millet with a wide arrow and with other gene networks through narrow ones (Budhwar *et al.*, 2020; Liu *et al.*, 2021), while myotonic dystrophy could be associated with unknown nutrient compounds from millets (Malik *et al.*, 2022).

All things included, the current analysis provides valuable insights into past research, changing trends in millet research, the current state of research and future research directions towards agriculture sustainability. It will decipher the first baseline data on the subject for comparisons in the future and to help policymakers develop strategies for increasing production of underutilized crops to achieve the sustainable goal as intended for the International Year of Millets (IYoM). The current in-depth, abstract keyword-based study of millet improvement contributes to a better understanding of the millet crop's agronomic features and management techniques. Furthermore, such studies will surely be helpful for breeders, farmers, policymakers and the industrial context in the collection and evaluation of data and research activities on the trajectory of improvement of millet crops, which is further helpful in gaining a thorough mainstreaming of research action in millets and is essential for formulating future protective and adaptive policies. The current work also highlights the need for additional research on mainstreaming the genetic variety improvements of other millets by knocking down their conservation status. Additionally, there was a lack of effective international research collaboration, particularly among nations in the African and Australian regions, in millet research domains. Further, national and international food security organizations should support and encourage more researchers to conduct more studies and compilations of data on the millets. More research projects that are conducted in collaboration with national and international initiatives are being done on landraces, genetic diversity, conservation and climate change adaptation. It enhanced the production of millet-based value-added products to leverage through marketing and entrepreneurship and underwent quality assurance and toxicity appraisal. Moreover, appropriate research on millets provides a tremendous opportunity to understand disease, diagnosis and treatments for a plethora of diseases and to prove the concept of 'Nutri-cereals.' Thus, these results will raise farmer's awareness, income, food and nutritional security, especially under the dearth of natural resources and certainly improve the effectiveness of millet research programs worldwide to uncover miracle grains for marvellous healthfulness.

Conclusion

The current investigation represents the first comprehensive exploration of the overlooked and underutilized realm of millet enhancement. The distinctive approach in the current research integrates bibliometric analysis and a tailored text-mining workflow, utilizing PubMed for its unparalleled advantages in granularity and comprehensive coverage of millet research. The study reveals crucial research themes, key figures and literature gaps, offering a robust baseline for millet enhancement studies. The findings underscore the imperative for intensified international

https://doi.org/10.1017/S0021859623000618 Published online by Cambridge University Press

collaboration, particularly in African and Australian regions, emphasizing the need for increased support from food security organizations.

Data availability. The authors confirm that the data supporting the findings of the current study are available within the article or its supplementary materials.

Acknowledgements. We are thankful to the Indian Council of Agricultural Research, Ministry of Agriculture and Farmers' Welfare, Govt. of India for infrastructure setup and other facilities at ICAR-IASRI, New Delhi, India created under National Agricultural Innovation Project, funded by World Bank at ICAR-IASRI, New Delhi. Financial support from DBT and CABin Grant of ICAR-IASRI, New Delhi is also thankfully acknowledged.

Author contributions. JS, MAI and DK conceived theme of the study. JS curated and analysed the data. JS, TC drafted the manuscript; MAI and DK reviewed and edited the manuscript; all authors read and approved the final manuscript and contributed to the article and approved the submitted version.

Funding statement. The authors declare that no external funds, grants or other support was received for conducting the current study.

Competing interests. The authors declare no conflict of interest.

Ethical standards. Not applicable.

References

Adelabu DB and Franke AC (2023) Research status of seed improvement in underutilized crops: prospects for enhancing food security. *The Journal of Agricultural Science* **161**, 398–411.

Andrade Pereira F and Mugnaini R (2023) Mapping the use of google scholar in evaluative bibliometric or scientometric studies: a bibliometric review. *Quantitative Science Studies* **4**, 233–245.

Aria M and Cuccurullo C (2017) bibliometrix: an R-tool for comprehensive science mapping analysis. *Journal of Informetrics* **11**, 959–975.

Bakhtin P, Khabirova E, Kuzminov I and Thurner T (2020) The future of food production–a text-mining approach. *Technology Analysis & Strategic Management* **32**, 516–528.

Budhwar S, Sethi K and Chakraborty M (2020) Efficacy of germination and probiotic fermentation on underutilized cereal and millet grains. *Food Production, Processing and Nutrition* **2**, 1–17.

Bui-Klimke TR and Wu F (2015) Ochratoxin A and human health risk: a review of the evidence. *Critical Reviews in food Science and Nutrition* **55**, 1860–1869.

Chandra AK, Chandora R, Sood S and Malhotra N (2021) Global production, demand, and supply. In Singh M and Sood S (eds), *Millets and pseudo Cereals*. United kingdom: Woodhead Publishing, pp. 7–18.

Chaturvedi P, Govindaraj M, Govindan V and Weckwerth W (2022) Sorghum and pearl millet as climate resilient crops for food and nutrition security. *Frontiers in Plant Science* **13**, 851970.

Chen L, Friedman C and Finkelstein J (2017) Automated metabolic phenotyping of cytochrome polymorphisms using pubmed abstract mining. In AMIA Annual Symposium Proceedings (Vol. 2017, p. 535). American Medical Informatics Association.

Chen Q, Lee K, Yan S, Kim S, Wei CH and Lu Z (2020) BioConceptVec: creating and evaluating literature-based biomedical concept embeddings on a large scale. *PLoS computational biology* **16**, e1007617.

Cobo MJ, Lopez-Herrera AG, Herrera-Viedma E and Herrera F (2011) An approach for detecting, quantifying, and visualizing the evolution of a research field: a practical application to the fuzzy sets theory field. *Journal of informetrics* **5**, 146–166.

Cooper MW, Brown ME, Niles MT and ElQadi MM (2020) Text mining the food security literature reveals substantial spatial bias and thematic broadening over time. *Global Food Security* **26**, 100392.

Dayakar Rao B, Bhat V, Niranjan T, Sujatha M and Tonapi VA (2021) Demand creation measures and value chain model on millets in India. In Kumar A, Tripathi MK, Joshi D and Kumar V (eds), *Millets and Millet Technology*. Singapore: Springer, pp. 381–411.

Dhaliwal SS, Sharma V, Shukla AK, Verma V, Kaur M, Shivay YS and Hossain A (2022) Biofortification – A frontier novel approach to enrich micronutrients in field crops to encounter the nutritional security. *Molecules* **27**, 1340.

Di Stefano E, White J, Seney S, Hekmat S, McDowell T, Sumarah M and Reid G (2017) A novel millet-based probiotic fermented food for the developing world. *Nutrients* **9**, 529.

FAOSTAT (2021) Food and Agriculture Organisation of the United Nations Statistical Database; Statistical Division; FAO: Rome, Italy. Available online http://www.fao.org/statistics/en/

Gowri MU and Shivakumar KM (2020) Millet scenario in India. *Economic Affairs* **65**, 363–370.

Gunturkun MH, Flashner E, Wang T, Mulligan MK, Williams RW, Prins P and Chen H (2022) GeneCup: mining PubMed and GWAS catalog for gene–keyword relationships. *G3* **12**, jkac059.

Hassan ZM, Sebola NA and Mabelebele M (2021) The nutritional use of millet grain for food and feed: a review. *Agriculture & Food Security* **10**, 1–14.

He R, Liu M, Zou Z, Wang M, Wang Z, Ju X and Hao G (2022) Anti-inflammatory activity of peptides derived from millet bran in vitro and in vivo. *Food & Function* **13**, 1881–1889.

Hickman L, Thapa S, Tay L, Cao M and Srinivasan P (2022) Text preprocessing for text mining in organizational research: review and recommendations. *Organizational Research Methods* **25**, 114–146.

Kaur B, Sandhu KS, Kamal R, Kaur K, Singh J, Roder MS and Muqaddasi QH (2021) Omics for the improvement of abiotic, biotic, and agronomic traits in major cereal crops: applications, challenges, and prospects. *Plants* **10**, 1989.

Kharat S, Medina-Meza IG, Kowalski RJ, Hosamani A, Ramachandra CT, Hiregoudar S and Ganjyal GM (2019) Extrusion processing characteristics of whole grain flours of select major millets (foxtail, finger, and pearl). *Food and Bioproducts Processing* **114**, 60–71.

Krishnan P, Praharaj CS, Kantharajan G, Bhoomaiah D, Sekar I, Soam SK and Rao CS (2021) A scientometric analysis of research on pulses in India during 2000–2017. *Agricultural Research* **11**, 565–578.

Liu F, Shan S, Li H, Shi J, Hao R, Yang R and Li Z (2021) Millet shell polyphenols prevent atherosclerosis by protecting the gut barrier and remodeling the gut microbiota in ApoE−/− mice. *Food & Function* **12**, 7298–7309.

Lu Z (2011) PubMed and beyond: a survey of web tools for searching biomedical literature. *Database* **2011**, baq036.

Luo Y, Zhang Z, Li Z, Chen Y, Zhang L, Cao J and Tao F (2020) Identifying the spatiotemporal changes of annual harvesting areas for three staple crops in China by integrating multi-data sources. *Environmental Research Letters* **15**, 074003.

Makun HA, Adeniran AL, Mailafiya SC, Ayanda IS, Mudashiru AT, Ojukwu UJ and Salihu DA (2013) Natural occurrence of ochratoxin A in some marketed Nigerian foods. *Food Control* **31**, 566–571.

Malik HZ, Sharma G, Moreno C, Parcha SP and Parcha S (2022) A medley of malnutrition and myotonic dystrophy: twice unlucky. *Cureus* **14**, e21180.

Muthamilarasan M, Dhaka A, Yadav R and Prasad M (2016) Exploration of millet models for developing nutrient rich graminaceous crops. *Plant Science* **242**, 89–97.

Nadeem F, Ahmad Z, Ul Hassan M, Wang R, Diao X and Li X (2020) Adaptation of foxtail millet (Setaria italica L.) to abiotic stresses: a special perspective of responses to nitrogen and phosphate limitations. *Frontiers in Plant Science* **11**, 187.

Naresh RK, Bhatt R, Singh PK, Kumar Y, Tiwari H, Saini A and Thakur H (2023) Millet: the super food in context of climate change for combating food and water security: a review. *The Pharma Inno* **12**, 1040–1049.

Palacios-Rojas N, McCulley L, Kaeppler M, Titcomb TJ, Gunaratna NS, Lopez-Ridaura S and Tanumihardjo SA (2020) Mining maize diversity and improving its nutritional aspects within agro-food systems. *Comprehensive Reviews in Food Science and Food Safety* **19**, 1809–1834.

Parthasarathy Rao P and Basavaraj G (2015) Status and prospects of millet utilization in India and global scenario. In *Millets: Promotion for Food, Feed, Fodder, Nutritional and Environment Security, Proceedings of Global Consultation on Millets Promotion for Health & Nutritional Security*. Hyderabad: Society for Millets Research, ICAR Indian Institute of Millets Research, pp. 197–209.

https://doi.org/10.1017/S0021859623000618 Published online by Cambridge University Press

Rao SV, Raju MVLN, Reddy MR and Panda AK (2004) Replacement of yellow maize with pearl millet (*Pennisetum typhoides*), foxtail millet (*Setaria italica*) or finger millet (*Eleusine coracana*) in broiler chicken diets containing supplemental enzymes. *Asian-Australasian Journal of Animal Sciences* **17**, 836–842.

Rossanez A, Dos Reis JC, Torres RDS and de Ribaupierre H (2020) KGen: a knowledge graph generator from biomedical scientific literature. *BMC Medical Informatics and Decision Making* **20**, 1–24.

Rotela S, Borkar S and Borah A (2021) Health benefits of millets and their significance as functional food: a review. *Journal of Pharmaceutical Innovation* **10**, 158–162.

Sahu J (2021) Mining proteome research reports: a bird's eye view. *Proteomes* **9**, 29.

Saini S, Saxena S, Samtiya M, Puniya M and Dhewa T (2021) Potential of underutilized millets as nutri-cereal: an overview. *Journal of Food Science and Technology* **58**, 4465–4478.

Shan S, Li Z, Newton IP, Zhao C, Li Z and Guo M (2014) A novel protein extracted from foxtail millet bran displays anti-carcinogenic effects in human colon cancer cells. *Toxicology Letters* **227**, 129–138.

Shankaramurthy K and Somannavar M (2019) Moisture, carbohydrate, protein, fat, calcium, and zinc content in finger, foxtail, pearl, and proso millets. *Indian Journal of Health Sciences and Biomedical Research* **12**, 228–228.

Shivhare R and Lata C (2016) Selection of suitable reference genes for assessing gene expression in pearl millet under different abiotic stresses and their combinations. *Scientific Reports* **6**, 1–12.

Tao D, Yang P and Feng H (2020) Utilization of text mining as a big data analysis tool for food science and nutrition. *Comprehensive Reviews in Food Science and Food Safety* **19**, 875–894.

Thakur K and Kumar V (2022) Application of text mining techniques on scholarly research articles: methods and tools. *New Review of Academic Librarianship* **28**, 279–302.

Vetriventhan M, Azevedo VC, Upadhyaya HD, Nirmalakumari A, Kane-Potaka J, Anitha S and Tonapi VA (2020) Genetic and genomic resources, and breeding for accelerating improvement of small millets: current status and future interventions. *The Nucleus* **63**, 217–239.

Wei CH, Allot A, Leaman R and Lu Z (2019) PubTator central: automated concept annotation for biomedical full text articles. *Nucleic Acids Research* **47**, 587–593.

Zhang Y, Wang L, Shen X, Wei X, Huang X, Liu Y and Lei H (2017) Broad-specificity immunoassay for simultaneous detection of ochratoxins A, B, and C in millet and maize. *Journal of Agricultural and Food Chemistry* **65**, 4830–4838.

Zhong B, Wu H, Li H, Sepasgozar S, Luo H and He L (2019) A scientometric analysis and critical review of construction related ontology research. *Automation in Construction* **101**, 17–31.

Zhu Y, Chu J, Lu Z, Lv F, Bie X, Zhang C and Zhao H (2018) Physicochemical and functional properties of dietary fiber from foxtail millet (Setaria italic) bran. *Journal of Cereal Science* **79**, 456–461.

https://doi.org/10.1017/S0021859623000618 Published online by Cambridge University Press