



## Research article

# Current research status and emerging trends in wheat: An integrated scientometric analysis based on ploidy uncovers hidden footprints in the scientific landscape

Tilak Chandra<sup>1</sup>, Jagajjit Sahu<sup>1</sup>, Sarika Jaiswal, Mir Asif Iquebal<sup>\*</sup>, Dinesh Kumar

Division for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, New Delhi, 110012, India

## ARTICLE INFO

## Keywords:

Wheat  
 Polyploidy  
 PubMed  
 Scientometrics  
 China  
 Natural language processing  
 Topical analysis

## ABSTRACT

Wheat, a highly versatile staple crop cultivated extensively for its grains on a global scale, is poised to experience increased demand to sustain the burgeoning population, owing to its superior nutritional potential. Modern wheat, a hexaploid species, has evolved through the introgression of numerous preceding ploidies, including Einkorn, Emmer, Aegilops, and others, each possessing distinct qualitative and quantitative traits. Scientometric and topical analyses serve as effective tools to quantitatively evaluate scientific research by measuring the knowledge expressed in scientific publications and keywords. Thus, comprehending the research status regarding wheat domestication events within primary, secondary, and tertiary gene pools is paramount for enhancing wheat production. In this study, we analyze data retrieved from PubMed to elucidate the research status and identify bottlenecks across different ploidy of genomic pools of wheat. The publication trends on wheat have experienced exponential growth over the past three decades, with China emerging as a leading center for publications. In contrast to the publication frequency observed in hexaploid common wheat, scholarly output concerning Einkorn and Aegilops is approximately tenfold lesser, with emmer trailing behind at three times fewer publications. This discrepancy underscores the prioritization of expedited research initiatives targeting these species, aimed at elucidating latent biological characteristics and optimizing their breeding capabilities. Keywords such as “stress,” “GWAS,” and “gene” are prominent, reflecting the challenges posed by climatic factors on wheat production and their mitigation through molecular breeding and gene manipulation. Notably, the keyword “einkorn” highlights its potential as a donor for fine-tuning traits related to wheat adaptation processes and quality, crucial for modern wheat’s survivability under adverse climates. Conversely, higher publication rates on emmer are primarily associated with Italy, possibly due to its favorable Mediterranean climate for tetraploid wheat. Keywords like “Pasta” and “Ochratoxin, DON” are prevalent, with the former being derived from durum wheat and the latter being reported in higher amounts in durum compared to other wheat species, rendering it less suitable for consumption. Enriched keywords such as “genome” and “resistance” underscore the critical characteristics of Aegilops. Other significant keywords like “Aceria tosichella” possibly indicate multiple stages of resistance conferred by Aegilops, while the presence of the grain softness protein “puroindoline” enhances its acceptability for donation by Aegilops. Spelt, a close relative of common wheat, exhibits a research trend with thousands of annual publications and enriched keywords such as “stress” and

<sup>\*</sup> Corresponding author.

E-mail address: [ma.iquebal@icar.gov.in](mailto:ma.iquebal@icar.gov.in) (M.A. Iquebal).

<sup>1</sup> Equal contribution.

<https://doi.org/10.1016/j.heliyon.2024.e36375>

Received 22 May 2024; Received in revised form 9 August 2024; Accepted 14 August 2024

Available online 15 August 2024

2405-8440/© 2024 Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

“yield” reflect the current scientific emphasis on wheat research. Furthermore, hierarchical keywords like “bio-control” and “celiac disease” merit consideration for future research on hexaploid wheat.

## 1. Introduction

Wheat stands as the primary staple crop in temperate regions, serving both human food and livestock feed purposes [1]. As one of the world’s largest and most vital food resources, wheat necessitates continual enhancement to meet the demands of the burgeoning human population. The improvement of wheat is intricately linked with understanding its evolution and genomic diversity across its species and their relatives [2]. The evolution of the *Triticum* genus serves as a prominent model for comprehending polyploidy, a prevalent form of plant evolution. Such evolutions are primarily driven by hybridization, sequential allopolyploidization, multiple domestication processes, and recurrent formation of modern wheat from its wild relatives [3,4]. Modern wheat, as an admixture with distinct ploidy derived from diverse parents, has acquired wide adaptive plasticity, favorable yield, and extended end-use quality, positioning it as a global staple food [4,5]. Therefore, understanding species domestication and genome evolution within primary, secondary, and tertiary gene pools is imperative for enhancing wheat production and guiding further research [2]. Evaluating the underlying significance of evolutionary processes, such as the pivotal genome concept driving structural and genomic changes within polyploid and individual wheat genomes, is crucial. The three well-described ploidy levels in wheat include Einkorn ( $2n = 2x = 14$ ), representing either wild or domesticated species, Emmer representing numerous species and subspecies, with a tetraploid genome ( $4n = 4x = 28$ ), and Spelt often denoting hexaploidy (6X), inclusive of common wheat species and subspecies [6,7]. *Aegilops a*, a unique grass with numerous accepted species ( $2n = 2X = 14$ ) significantly contributes to the wheat adaptive trait evaluation [8]. Therefore, these four main keywords, along with a comprehensive assortment of keywords, could be employed to dissect the holistic scenario of wheat across distinct individual ploidy levels. While wheat consumption typically provides essential nutrients, phytochemicals, bioactive components, and dietary fibers [9], several bottlenecks are associated with wheat products, often leading to adverse reactions such as intolerances, allergies, and celiac disease [10]. Therefore, critical research is warranted into the individual ploidy levels of wheat to advance our understanding of their utility and impacts on consumer preferences. Such studies delving into wheat genetics are crucial for improving breeding strategies, exploring its evolutionary history, and developing superior varieties to address the challenges of food security and environmental sustainability. Additionally, they shed light on wheat’s historical and cultural significance, revealing insights into its agricultural history, domestication processes, and aiding breeders in developing improved varieties with enhanced yield, disease resistance, and nutritional content.

Unlocking the mysteries of current wheat ploidy-based research and trends requires the utilization of integrated scientometrics, a versatile tool aiding researchers, institutions, and policymakers in understanding the dynamics of research in wheat evolution and domestication [1]. Scientometrics serves as a pivotal tool in mapping research trends, identifying key contributors, and assessing research impact. Contemporary text mining tools, leveraging Natural Language Processing (NLP), offer profound insights into scientific attributes. Numerous resources enriched with NLP capabilities facilitate the automated annotation of scientific texts. By examining publication patterns, keyword co-occurrences, and employing Named Entity Recognition (NER), researchers can identify key milestones, emerging areas of interest, and gaps in knowledge [1,11,12]. Scientometrics measures information processes, focusing on the qualitative and quantitative aspects of science and research in wheat evolution and domestication, thereby guiding informed decisions and promoting sustainable agricultural practices [1,13]. Additionally, topical analysis offers valuable insights into the content and structure of textual data, assisting researchers in uncovering hidden patterns, trends, and themes that may inform further investigation or decision-making across various domains [1,14]. While scientometric analysis currently extends to understanding diverse scientific aspects in wheat, such as comparative research trends in wheat and barley [1], revealing basic trends in wheat research [11], unveiling trends on pasta and durum wheat [14], understanding the research status of wheat blast, a catastrophic threat [13], and analyzing the structure and trends of research across the globe [15], we hypothesize testing the evaluation of research trends by integrating scientometrics and topical analysis on the different wheat progenitors, along with the special donor *Aegilops*. Additionally, we analyze all chosen keywords representing diverse wheat ploidy levels together. Integrated scientometric and topical analysis offers a comprehensive understanding of the landscape of ploidy-based wheat research, encompassing research growth, key publication contributors, main collaboration networks, and the principal topics driving scientific inquiry. Furthermore, it uncovers hidden patterns, trends, and themes regarding keyword frequency and growth patterns, various bio-concept classes, keyword evolution, and their co-occurrence networks. These insights can guide future research directions on wheat, funding priorities, and collaborative efforts aimed at advancing knowledge in wheat genetics and breeding improvement.

## 2. Materials and methods

### 2.1. Data collection and processing

Four distinct groups were initially delineated based on species names and their aliases. We chose search keywords based on three distinct ploidy levels only, i.e., (2X), (4X), and (6X), excluding any other criteria for categorization due to the confusing and controversial nature of many keywords used in wheat evolution and domestication [2]. Since wheat could represent any form of ploidy, we excluded keyword ‘wheat’ from any searches to avoid redundant data generation and ensure reproducibility. Data were

collected for each group individually using specified keywords within the Title/Abstract field on PubMed as of March 27, 2024. The language filter was restricted to English, and the search was constrained to publications up to December 31, 2023. Citation data were obtained in PubMed format for each group and stored locally. Due to the volume of records exceeding 10,000 matches (the one-time download limit on PubMed), the citation data for the Spelt group were downloaded in batches. The subsequent analysis primarily utilized R scripts, starting with the importation of citation data using the bibliometrix package [16], which were then converted into data frames. For the spelt group, comprising two files, both were imported, converted, and subsequently merged into a single dataset. Additionally, a merged dataset encompassing all four groups was prepared. Initially, data underwent preprocessing to eliminate redundant entries, and any records dated from 2024, if present, were excluded using R scripts. Subsequently, these datasets were subjected to further analysis, broadly categorized into scientometrics and topical analysis.

Summary of group-wise segregation of data and corresponding analysis workflow with PubMed search criteria. First of all we have search and analyzed all the keywords relevant to wheat species viz., Einkorn [Title/Abstract] OR Boeoticum [Title/Abstract] OR Monococcum [Title/Abstract] OR Urartu [Title/Abstract] Emmer [Title/Abstract] OR Turgidum [Title/Abstract] OR Dicocccum [Title/Abstract] OR Durum [Title/Abstract] OR Dicocccoides [Title/Abstract] OR Araraticum [Title/Abstract] OR Turanicum [Title/Abstract] OR Persicum [Title/Abstract] OR Polonicum [Title/Abstract] OR Timopheevi [Title/Abstract] Aegilops [Title/Abstract] OR Goatgrass [Title/Abstract] OR Tauschii [Title/Abstract] OR Speltoides [Title/Abstract] OR Squarrosa [Title/Abstract] Spelt [Title/Abstract] OR Zhukovskiy [Title/Abstract] OR Aestivum [Title/Abstract] OR Compactum [Title/Abstract] OR Sphaerococcum [Title/Abstract] OR Macha [Title/Abstract] first and followed by based on upgrading level of ploidy levels viz for einkorn (2X) the search keywords are Einkorn [Title/Abstract] OR Boeoticum [Title/Abstract] OR Monococcum [Title/Abstract] OR Urartu [Title/Abstract] for Emmer (4X) Emmer [Title/Abstract] OR Turgidum [Title/Abstract] OR Dicocccum [Title/Abstract] OR Durum [Title/Abstract] OR Dicocccoides [Title/Abstract] OR Araraticum [Title/Abstract] OR Turanicum [Title/Abstract] OR Persicum [Title/Abstract] OR Polonicum [Title/Abstract] OR Timopheevi [Title/Abstract] and for Aegilops (2X) the significant donor to polyploidisation, the keywords are searched Aegilops [Title/Abstract] OR Goatgrass [Title/Abstract] OR Tauschii [Title/Abstract] OR Speltoides [Title/Abstract] OR Squarrosa [Title/Abstract] and searched for spelt (6X) the search keywords are Spelt [Title/Abstract] OR Zhukovskiy [Title/Abstract] OR Aestivum [Title/Abstract] OR Compactum [Title/Abstract] OR Sphaerococcum [Title/Abstract] OR Macha [Title/Abstract].

## 2.2. Scientometric analysis

The scientometrics analysis was primarily conducted using the bibliometrix and complementary packages, aiming to offer a comprehensive overview of research productivity, as outlined by Ref. [14]. Given that PubMed citation data lack citation-specific information, analyses focused on other metrics that do not consider citation. The analysis began by examining year-wise trends in article production, followed by identifying the top contributing countries and their respective publication rates, while also assessing growth trends over time. Additionally, a network analysis of country collaborations was undertaken to discern key contributors to international research collaboration.

## 2.3. Topical analysis

The topical analysis aimed to elucidate prevalent research themes in terms of keywords and bio-concepts. Initially, a keyword cloud based on the frequency distribution of author-supplied keywords was constructed for each group using R. Subsequently, employing bibliometrix and other relevant packages, a keyword analysis was performed, culminating in the identification of the top ten keywords in terms of growth across years and the construction of a Sankey diagram illustrating keyword evolution across different epochs. Additionally, keyword co-occurrence networks were constructed and processed for each group using the igraph package [17]. These networks were exported as edge files and visualized using Gephi [18], facilitating the application of various network parameters to optimize visualization. Modularity was utilized for node coloring, while average weighted degree was employed to denote node size, highlighting dominant nodes within the network. The latter segment of the topical analysis entailed leveraging the PubTator tool [19]. Utilizing the pubmed.mineR package [20], PubTator data were retrieved in R by supplying PubMed IDs (PMIDs) to the pubtator\_function, which interfaced with the PubTator server via API services to fetch relevant data. Subsequently, these data were processed using in-house scripts to calculate the frequency of individual bio-concepts, amalgamating all alias names associated with each bio-concept. Bar plot collages were generated to visualize the top five bio-concepts in four categories: Chemical, Disease, Gene, and Species.

## 3. Results

The scientometric analysis of publications related to wheat, regardless of ploidy and other criteria, reveals a notable trend over time (Fig. 1(A and B)). Initially, from the first publication in 1887 until 1966, the frequency of publications remained in the single digits. However, from 1969 to 1989, there was an increase in publications, with numbers reaching into the double digits. Since 1990 until 2018, the frequency of publications has consistently been in the triple digits, with a sharp increment observed since 2019, albeit with slight deviations. Regarding the top ten countries contributing to wheat-related publications, China emerges as the leading publication center, with approximately 3500 publications, nearly double the number of publications from the United States, which holds the second position with around 1950 publications. Notably, both China and the USA show a progressive increase in publication trends. Apart from these major contributors, Italy, Australia, India, Germany, France, Canada, Japan, and Spain are among the top ten countries listed. Interestingly, the data reveals that the first publication on wheat did not originate from any of these top 10 countries.



**Fig. 1.** The figure is divided into two major components, **A:** Scientometrics analysis output i: No of articles per year ii: Growth rate of no of articles per year iii: Top ten countries and respective number of articles iv: Country collaboration network **B:** Topical analysis output i: Word cloud for keyword frequency ii: Top 10 keyword growth iii: Top five bio concepts under 4 classes such as Chemical, Disease, Gene, Species from PubTator analysis iv: Keyword evolution across different era v: Keyword co-occurrence network for the input keyword representing admixture of all wheat ploidy species and subspecies taken into consideration in this study.

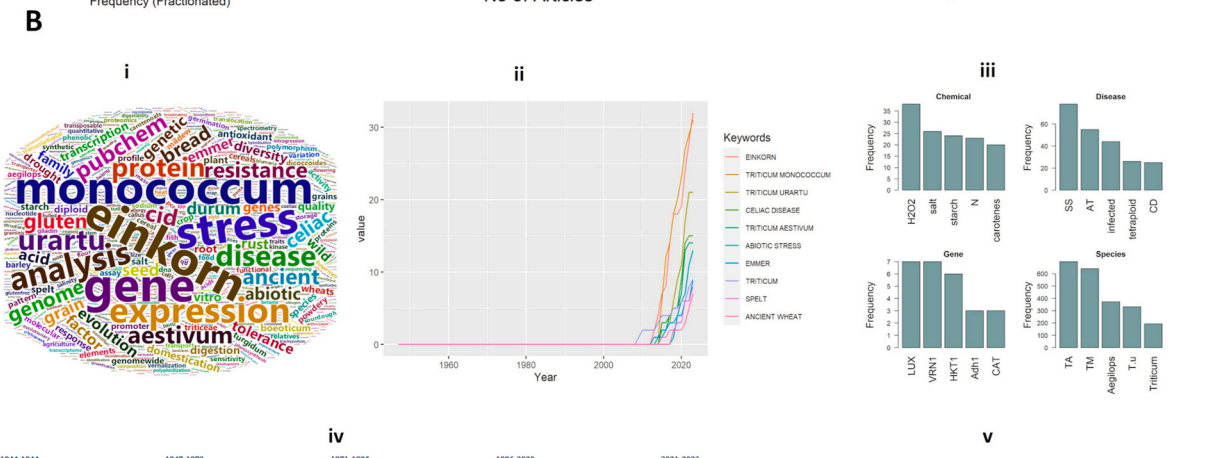
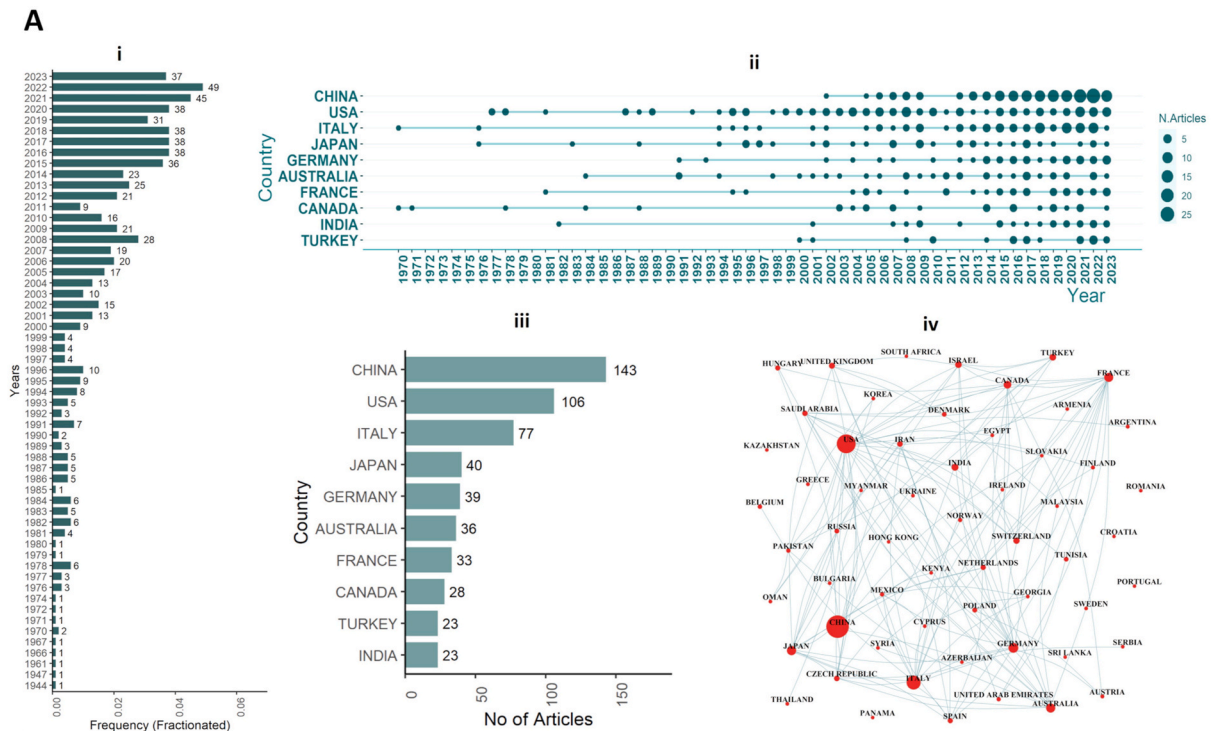
Analyzing the country collaboration network, China and the USA emerge as leading publication centers on wheat, represented by large nodes, followed by other countries with smaller extents of collaboration. A word cloud generated for keyword frequency highlights terms such as “stress” and “gene” as predominant. Additionally, a keyword growth chart indicates that “*Triticum aestivum*” is a significantly used keyword in publications. Further analysis using PubTator reveals the top five bio concepts across four classes: Chemical, Disease, Gene, and Species. For instance, under the Chemical class, keywords such as water, nitrogen, chlorophyll, sugar, and phosphorus are prevalent, while the Disease class includes keywords like stresses, infections, and toxicity. Representative keywords in the Gene class include anti-oxidative enzymes such as catalase, peroxidase, superoxide dismutase, ascorbate peroxidase, and glutathione-S-transferase, which overlap with search keywords for hexaploid species. In the Species class, “bread wheat” achieves the highest frequency against search keywords, with other related species including durum, Aegilops, and rice. A Sankey analysis illustrates the evolution of keywords across different eras, indicating a diverse set of keyword evolution across five divided eras in publication centers. Notably, keywords such as GWAS, sourdough, wheat, and genome persist hierarchically through keyword evolution. Finally, a keyword co-occurrence network reveals the association of wheat (*Triticum aestivum*) into two major clusters. The blue cluster, originating from wheat, is associated with dominant keywords such as germination, climate change, RNAseq, oxidative stress, cadmium, and gene expression. Conversely, an orange node originating from durum is linked with significant keywords like bread wheat, resistance, leaf rust, GWAS, as well as notable wheat diseases like leaf rust, stripe rust, powdery mildew, and *Fusarium* head blight. Additionally, a small green cluster represents the association of *Triticum* with Aegilops.

### 3.1. Einkorn: first and oldest single grain reveals its domestication events

Einkorn, recognized as the oldest known wheat variety, stands as humanity’s first cultivated wheat and belongs to the grass species native to western Asia. It is referred to as *Triticum boeoticum*, *Triticum monococcum*, and *Triticum urartu*, or red wild einkorn wheat, and is perceived as either distinct species or closely related to wheat. Upon exploring these three keywords representing the initial ploidy for wheat, intriguing findings emerged (Fig. 2(A and B)). The earliest publication mentioning these keywords dates back to 1944, with minimal progress observed until the early 2000s, marked by a meager number of publications in the single digits. However, from the year 2000 onwards, an exponential surge in publications on einkorn has been witnessed. Mirroring the trend observed in wheat, China leads as the primary producer, followed by the USA, Italy, Japan, Germany, Australia, France, Canada, India, with Turkey rounding out the top 10 countries, each contributing 23 publications. Analyzing the country collaboration network, China, the USA, and Italy emerge as prominent collaborators, while Australia, Japan, France, and Canada exhibit proportionally fewer collaborations. A keyword cloud highlights key terms such as “gene,” “einkorn,” “monococcum,” and “stress.” Among the top 10 keywords, the escalation of terms related to einkorn has been particularly sharp since 2010, evident in abstracts and titles. Mining data from PubTator reveals the top five bio-concepts across four classes. In the chemical class, H<sub>2</sub>O<sub>2</sub> (hydrogen peroxide) exhibits the highest frequency, followed by salt, starch, nitrogen, and carotenes. Within the Disease class, stresses hold the top position, followed by infected, tetraploid, and celiac disease. In the gene class, notable genes include LUX, VRN1, HKT1, Adh1, and CAT, with LUX and VRN1 demonstrating similar frequencies, followed by HKT1, while Adh1 and CAT are roughly half as frequent as the top genes, LUX and VRN1. Regarding species, einkorn is primarily associated with TA (*Triticum aestivum*), TM (*Triticum monococcum*), Aegilops, and TU (*Triticum urartu*), encompassing the genus *Triticum* and its various species. Furthermore, visualizations of keyword evolution across different eras depict linear Sankey diagrams, illustrating the flow of keywords proportionally across eras. Major keywords include anthocyanin biosynthesis, domestication, and earliness per se, among others relevant to einkorn. In the keyword co-occurrence network, strong associations of einkorn with keywords such as emmer are observed, while the central term “wheat” is strongly linked with celiac disease and *Triticum monococcum*.

### 3.2. Emmer, a wheat with substantial societal impacts

To glean further insights from our scientometric analysis, we expanded our study to encompass varying levels of ploidy, leading to intriguing observations (Fig. 3(A and B)). According to our analysis, the earliest publication containing related keywords dates back to 1925 from Canada, a country among the top 10 contributors. The relative growth rate of the number of articles per year suggests that Italy leads as the primary publication center for emmer, exhibiting a contrasting trend compared to other tested keywords. China, USA, and Spain follow Italy closely, with 401, 387, and 242 publications, respectively. The country collaboration network also underscores Italy, China, and the USA as leading collaborators, evident from the density of nodes. A word cloud for keyword frequency distinctly highlights “durum” and “triticum” as dominant keywords compared to others profiled. Examining the growth of the top 10 keywords reveals a sharp peak for “durum wheat” in comparison to others, with additional keywords such as “pasta,” “abiotic stress,” “GWAS,” and “QTL” routinely associated with the keyword search for durum wheat. The top five bio-concepts across four classes—Chemical, Disease, Gene, and Species—reveal notable patterns. In the Chemical class, water exhibits the highest frequency, followed by nitrogen, while sugar, cadmium (Cd), and deoxynivalenol (DON) share approximately half the frequency of water and nitrogen. Within the



(caption on next page)

**Fig. 2.** The figure is divided into two major components, **A:** Scientometrics analysis output i: No of articles per year ii: Growth rate of no of articles per year iii: Top ten countries and respective number of articles iv: Country collaboration network **B:** Topical analysis output i: Word cloud for keyword frequency ii: Top 10 keyword growth iii: Top five bio concepts under 4 classes such as Chemical, Disease, Gene, Species from pubtator analysis iv: Keyword evolution across different era v: Keyword co-occurrence network for the input keyword representing einkorn (2X) wheat ploidy species and subspecies.

Disease class, keywords such as Hd, stresses, infected, toxicity, and celiac disease are prevalent. Among the Gene class, keywords such as Pasta, CAT, PPO, SSR, and IL-1beta stand out, with a pattern of decreasing frequency, respectively. Under the Species classification, relatedness is closely associated with DW (durum wheat), BW (bread wheat), Triticum, Aegilops, and (TD) *Triticum dicoccoides*. A Sankey scheme facilitates the visualization of keyword evolution across different eras, highlighting promising keywords such as bunium persicum, cadmium, ochratoxin, and CRISPR. A large number of criss-crosses appear throughout the keyword evolutions during the third episode, with peculiar keywords emerging in relation to emmer, including bunium persicum, cadmium, ochratoxin, and pasta. Lastly, the keyword co-occurrence network presents a comprehensive network largely interlinked with other relevant keywords, such as durum wheat with QTL, GWAS, gene expression, abiotic stress, yield, landraces, genetic diversity, salinity, and climate change. Additionally, it is notably integrated with keywords related to resistance, leaf rust, celiac disease, and *Fusarium* head blight (FHB). The association between *Triticum aestivum* and other keywords signifies their interconnectedness in the context of domestication.

### 3.3. *Aegilops*: a crucial resources of biotic resistance genes reservoir

The *Aegilops* species have made significant contributions to wheat breeding by incorporating biotic resistance genes and improving traits such as yield and abiotic tolerance for drought and heat [8,44]. The keyword search focusing on the genus *Aegilops* has unveiled surprising insights, as it has played a pivotal role in wheat evolution (Fig. 4(A and B)). In this study, we specifically selected *Aegilops tauschii*, among many known species, because its D genome diploid species is commonly utilized in both artificial and natural selection processes. Other species pose greater challenges due to lower meiotic recombination frequencies and extensive translocation/introgression lines [44]. The first publication related to *Aegilops* emerged in 1887. It took nearly a century for the publication count to reach double digits, but progress has accelerated, reaching triple digits in recent years. China leads in publications related to *Aegilops*, with nearly double the number of publications compared to follower countries such as the USA, Japan, Germany, Australia, Israel, India, Spain, France, and Russia, all among the top ten contributors. Analyzing the country collaboration network reveals China and the USA as strong nodes, collaborating with the largest number of representative countries, while Germany, Japan, and Australia exhibit smaller nodes, indicating relatively less collaboration with other countries. A word cloud analysis of keyword frequency highlights “Triticum,” “tauschii,” “genome,” and “resistance” as prominent keywords, reflecting hallmark attributes for *Aegilops*. Furthermore, keyword growth analysis indicates a successive increase in keywords such as “abiotic stress” and “*Aegilops*” during the period 2010–2020. Top five bio-concepts across four classes—Chemical, Disease, Gene, and Species—reveal noteworthy patterns. In the Chemical class, “nitrogen” and “water” are the top keywords, with “sodium dodecyl sulfate,” “abscisic acid,” and “carbon” also prominent. In the Disease class, “*Aceria tosichella*” and various stresses exhibit the highest frequencies, while “infected,” “tetraploid,” and “polyploidy” are relatively less frequent. The Gene class frequencies reflect “grain softness protein” and “puroindoline” as the most frequent keywords, compared to others like “glutathione-S-transferase” and “low-molecular-weight glutenin subunits.” For the Species class, “*Aegilops*” and “bread wheat” are the top frequencies, while “Triticum,” “durum wheat,” and “barley” are less frequent. Keyword evolution across different eras reveals zig-zag patterns in keyword frequencies throughout the publication eras. Major keywords such as “cytoplasmic male sterility,” “disease resistance,” “admixture,” “substitution line,” “allelochemicals,” “grassland restoration,” and “anthocyanin biosynthesis” are prevalent across various epochs. Keywords like “cytoplasmic male sterility” and “disease resistance” persist throughout the developmental era of *Aegilops* literature evolution. Keyword co-occurrence network analysis reveals two distinct networks, one focusing on *Aegilops* speciation and cellular functioning. The genus *Aegilops* is directly linked to wheat via a broad line, followed by a narrow line connecting barley and triticales, with *Aegilops tauschii* and hexaploid common wheat in between. Additionally, it is connected to key keywords such as “synthetic hexaploid wheat,” “genome evolution,” and “introgression.” On the other side, *Aegilops* is directly linked to keywords associated with cellular mechanisms such as “autophagy,” “ATG8,” “cellulose synthase,” and “cytomixis.”

### 3.4. *Spelt*: widely modern cultivated species

Based on the ploidy level, hexaploid wheat is commonly cultivated due to its wide adaptability in modern agriculture. Our scientometric analysis of hexaploid wheat revealed that the first publication on this topic dates back to 1894, with scientific progress accelerating notably by 1970 (Fig. 5(A and B)). Since then, progress has been steady, with thousands of publications emerging annually on hexaploid wheat, focusing on keywords associated with hexaploidy. China emerges as the primary publication center, followed by the USA, Australia, India, Germany, France, Canada, Japan, Pakistan, and Italy. Remarkably, China alone publishes approximately twice as many publications as the USA, its closest counterpart in publication output. Australia holds the distinction of producing the first publication among the top ten countries. The collaborative network among countries demonstrates widespread interest in hexaploid wheat research, with China and the USA leading in publication output, depicted by their larger and darker nodes. Analysis of keyword frequency through word clouds highlights “stress” as the most prominent keyword, followed by gene, plant, resistance, grain,





**Fig. 3.** The figure is divided into two major components, **A:** Scientometrics analysis output i: No of articles per year ii: Growth rate of no of articles per year iii: Top ten countries and respective number of articles iv: Country collaboration network **B:** Topical analysis output i: Word cloud for keyword frequency ii: Top 10 keyword growth iii: Top five bio concepts under 4 classes such as Chemical, Disease, Gene, Species from pubtator analysis iv: Keyword evolution across different era v: Keyword co-occurrence network for the input keyword representing emmer (4X) ploidy species and subspecies.

and yield. The top ten keywords in terms of growth include *Triticum aestivum*, drought, abiotic stress, gene expression, grain yield, and GWAS. In the realm of bio concepts, under the chemical class, water and nitrogen emerge as high-frequency keywords, while chlorophyll, sugar, and phosphorus are relatively less frequent. In the disease class, keywords such as stresses, severity, infection, toxicity, and cell death are prominent. Under the gene class, antioxidative genes such as catalase, peroxidase, superoxide dismutase, ascorbate peroxidase, and glutathione-s-transferase are prevalent. In the species class, *Triticum aestivum* dominates in frequency, followed by barley, rice, maize, and *aegilops*. Keyword evolution across different eras reveals significant fluctuations in important keywords such as cadmium, wheat, antioxidant activity, bioactive compounds, bio-control, celiac disease, and septoria tritici blotch. The keyword “bio-control” is predominantly associated with fungal disease management, while “bioactive compounds” is linked with essential oils. Keyword co-occurrence networks highlight wheat’s associations with two main domains: inherent adaptability and disease and their mechanistic insights. In conclusion, our analysis summarizes the diverse keywords related to hexaploid wheat and their co-occurrence patterns throughout the abstracts and titles of publications.

#### 4. Discussion

The scientometric analysis of wheat-related publications reveals a distinct trend: single-digit outputs from 1887, progressing to double digits between 1969 and 1989, followed by triple digits from 1990 to 2018, with a notable surge since 2019, albeit with slight fluctuations (Fig. 1(A and B)). Remarkably, the publication trends on wheat have remained unaffected by external factors such as the negative impacts of COVID-19 on price, production, and trades [21]. China has emerged as the foremost publication hub, exhibiting a progressively increasing trend over the past two decades due to the adoption of dominant technologies and management practices [22]. Furthermore, funding and infrastructure for Research & Development have experienced exponential growth, particularly in terms of business expenditure, surpassing government expenditure [23]. China’s pivotal role as the largest publication center on wheat, its collaborations span a wide range across the globe, indicating future priorities for collaboration focused on the development of climate-resilient cultivars using cutting-edge technology [24]. Analysis of word clouds underscores “stress” and “gene” as prominent and self-explanatory keywords. In the realm of bio-concepts, water, nitrogen, chlorophyll, sugar, and phosphorus emerge as the top five in the chemical class, shaping the majority of wheat’s vegetative and reproductive tissues [25]. Keywords such as stresses, infections, and toxicity are prevalent, while antioxidative enzymes like catalase, peroxidase, and glutathione-s-transferase are widely represented due to wheat’s extensive associations with multiple stresses [26]. Persistent keywords such as GWAS, sourdough, and genome persist throughout hierarchical evolution, facilitating the evaluation of complex traits related to biotic, abiotic, and quality aspects of wheat through advanced technologies [27,28]. Additionally, associated dominant keywords such as germination, climate change, RNAseq, oxidative stress, cadmium, and gene expression are frequently linked with wheat throughout its evolution and domestication.

##### 4.1. Einkorn: first and oldest single grain reveals its domestication events

Einkorn, hailed as the oldest known wheat variety, stands as mankind’s initial encounter with wheat and is a grass species native to western Asia [29]. Despite its superior nutritional qualities, progress in Einkorn-related publications has remained sluggish since the year 2000, with only single-digit publications recorded [30]. However, following the recognition of its exceptional nutritional attributes, the scientific community has begun to explore this underappreciated wheat species, catalyzing an exponential surge in publications on Einkorn, with China emerging as the leading country in this domain once again (Fig. 2(A and B)). Key terms highlighted in the keyword cloud include “gene”, “einkorn”, “monococum”, and “stresses”, as these keywords frequently recur and are intricately linked with Einkorn’s evolution and adaptation processes [29]. In the realm of bio-concepts, hydrogen peroxide and stresses emerge as the most frequent in the chemical class, possibly correlated with Einkorn’s high levels of reactive oxygen species, rendering it more resilient to stress compared to bread wheat [30,31]. Notably, “celiac disease” stands out as a prominent keyword, often associated with wheat consumption and underscoring the unsuitability of Einkorn for individuals with celiac disease [30]. Another significant keyword is “genes”, highlighting the pivotal role genes play in Einkorn’s life cycle and development. Among these, the gene LUX has garnered attention due to its deletion observed in an early-flowering Einkorn mutant [32], while *VRN1* maintains the vegetative phase upon deletion [33]. Keyword evolution across different eras, such as anthocyanin biosynthesis, domestication, and earliness, correlates closely with Einkorn’s evolutionary trajectory. Anthocyanin biosynthesis, for instance, may be linked with the acquisition of aleurone color through chromosomal substitutions from diploid Einkorn [34], while the earliness locus in diploid Einkorn, mapped to regulate flowering time independently of vernalization and photoperiod, is crucial for fine-tuning wheat adaptation to diverse environments [29,35].



**Fig. 4.** The figure is divided into two major components, **A:** Scientometrics analysis output i: No of articles per year ii: Growth rate of no of articles per year iii: Top ten countries and respective number of articles iv: Country collaboration network **B:** Topical analysis output i: Word cloud for keyword frequency ii: Top 10 keyword growth iii: Top five bio concepts under 4 classes such as Chemical, Disease, Gene, Species from pubtator analysis iv: Keyword evolution across different era v: Keyword co-occurrence network for the input keyword representing *Aegilops* (2X), a special genome donor to wheat.

#### 4.2. *Emmer, a wheat with substantial societal impacts*

Upon expanding our study to encompass various ploidy levels, our scientometric analysis has unveiled intriguing insights. The earliest publication containing related keywords dates back to 1925, originating from Canada, which is among the top-ten countries in this domain. Italy emerges as the leading publication center for emmer, with China, the USA, and Spain closely following in terms of publication extent (Fig. 3(A and B)). This trend may be attributed to emmer being predominantly grown in Mediterranean countries such as Turkey, Italy, Greece, and parts of the Middle East, owing to their favorable climates and soil conditions [36]. In the country collaboration network, Italy, China, and the USA stand out as leading collaborators, with “durum” and “triticum” dominating keywords in the word cloud. Durum is frequently associated with keywords such as pasta, abiotic stress, GWAS, and QTL, elucidated through various research endeavors [28,37,38]. Under bio-concepts, water and nitrogen display the highest chemical frequency, while sugar, cadmium, and deoxynivalenol exhibit roughly half the frequency compared to water and nitrogen. The former keywords are highly relevant to discrepant nitrogen uptake, use efficiency, and high ionic water imbalance [39,40], whereas the latter highlighted keywords correlate with higher susceptibility to stresses and *Fusarium* head blight [28]. Notably, higher levels of deoxynivalenol contamination have been reported in durum compared to rye, triticale, and bread wheat [41]. The prevalence of cadmium stress in durum wheat can be attributed to a combination of environmental factors, the crop’s growth characteristics, and its limited tolerance to cadmium contamination [42]. The Sankey scheme reveals promising keywords such as cadmium, ochratoxin, and CRISPR, with a criss-cross pattern appearing during the third episode, signifying prominent keywords that underscore the impact of mycotoxins on the processing of wheat for human consumption [43]. Furthermore, the keyword co-occurrence network suggests a comprehensive interlinked network with keywords such as durum wheat, gene expression, abiotic stress, yield, landraces, genetic diversity, salinity, climate change, resistance, leaf rust, celiac disease, and *Fusarium* head blight [28,36,37].

#### 4.3. *Aegilops: a crucial resources of biotic resistance genes reservoir*

For *Aegilops*, a unique grass species that contributes a significant portion of the gene pool to the wheat genome [44], notable trends emerge from our research analysis. The first publication, dating back to 1887, took approximately a century to reach double digits, but recent trends have witnessed a rapid rise, with publications reaching three digits in a short span (Fig. 4(A and B)). China leads in publication production, nearly doubling the output compared to follower countries. In the country collaboration network among the top ten publication centers, China and the USA emerge as the strongest nodes, while collaborations involving Germany, Japan, and Australia are comparatively smaller. These publication trends and relative collaborations align with previous findings [13,22]. Analysis of keyword frequency in a word cloud highlights “Triticum,” “tauschaii,” “genome,” and “resistance” as key characteristics of *Aegilops* [8,45]. In the top five bio-concepts across four classes, nitrogen and water emerge prominently in the chemical class, supported by existing literature [46]. In the realm of disease classification, notable keywords encompass *Aceria tosichella* and diverse stress factors, potentially indicating varying degrees of resistance offered by *Aegilops* accessions against a globally significant cereal pest and various other stressors [47]. Within the gene class, keywords such as “grain softness protein” and “puroindoline” further strengthen the likelihood that *T. tauschii* serves as the donor of puroindoline genes to modern wheat [48]. Keyword evolution in *Aegilops* literature reflects a zig-zag motion of frequencies across different eras. Key keywords include cytoplasmic male sterility, disease resistance, admixture, substitution line, allelochemicals, grassland restoration, and anthocyanin biosynthesis [49,50]. The keyword co-occurrence network reveals two distinct networks: one connected with synthetic hexaploid wheat, genome evolution, and introgression, while the other links *Aegilops* to cellular mechanisms such as autophagy, ATG8, cellulose synthase, and cytomixis [51–53].

#### 4.4. *Spelt: a neighbour of common wheat*

Hexaploid wheat, renowned for its wide adaptability, stands as a staple in modern agriculture. Scientific inquiry into hexaploid wheat commenced in 1894, with notable acceleration observed by 1970 (Fig. 5(A and B)). Since then, the annual publication output on hexaploid wheat has surged into the thousands [1]. China leads the publication and collaboration networks in this domain, followed by the USA, Australia, India, and Germany, with China’s publication output approximately doubling that of the USA [1,11]. The word cloud illustrates “stress” as a prominent keyword, followed by genes, plants, resistance, grain, and yields, which aptly reflect the current scientific focus on wheat research [11]. Within the top five bio concepts across four classes, water and nitrogen emerge as the most frequently used chemical terms [46]. The disease class is characterized by keywords such as stresses, severity, infection, toxicity, and cell death. The majority of antioxidative genes, including catalase, peroxidase, superoxide dismutase, ascorbate peroxidase, and glutathione-s-transferase, are found within the gene class. In the species class, *Triticum aestivum*, barley, rice, maize, and *Aegilops* signify promising avenues for agricultural sustainability [54]. Keyword evolution across different eras highlights significant changes, with relevant keywords including cadmium, wheat, antioxidant activity, bioactive compounds, bio-control, celiac disease, septoria tritici blotch, fungal disease management, essential oils, and wheat’s association with inherent adaptability and disease mechanistic



**Fig. 5.** The figure is divided into two major components, **A:** Scientometrics analysis output i: No of articles per year ii: Growth rate of no of articles per year iii: Top ten countries and respective number of articles iv: Country collaboration network **B:** Topical analysis output i: Word cloud for keyword frequency ii: Top 10 keyword growth iii: Top five bio concepts under 4 classes such as Chemical, Disease, Gene, Species from pubtator analysis iv: Keyword evolution across different era v: Keyword co-occurrence network for the input keyword representing spelt, including modern wheat belonging to hexaploidy (6X) species and subspecies.

insights [9]. These keywords underscore the multifaceted nature of wheat research, spanning from agricultural sustainability to disease management and nutritional aspects.

## 5. Conclusion

The integration of text mining along with scientometric mapping approaches represents a quantitative method utilized to analyze the structure and evolution of research activities within specific scientific fields or disciplines. Our investigation is focused on understanding the primary, secondary, and tertiary gene pools, along with their interrelationships, based on data retrieved from PubMed. This aims to comprehend the research structure and trends in wheat domestication processes. Over the past three decades, there has been an exponential surge in publications on wheat, with China emerging as the leading publication center for all search keywords except “Emmer,” which represents durum wheat. Keywords such as “stress,” “GWAS,” and “gene” are prominent, reflecting the focus on addressing climatic challenges through molecular breeding and gene manipulation. In the case of the diploid progenitor einkorn, limited progress is evident compared to other publication trends. However, numerous genes from einkorn have contributed to modern wheat, as revealed by bio-concept classes. These keywords indicate einkorn’s potential as a donor of genes for fine-tuning traits related to wheat adaptation and quality, crucial for survivability under adverse climates. Emmer, primarily associated with Italy, benefits from the favorable Mediterranean climate for tetraploid wheat. Enriched keywords such as “Pasta” and “Ochratoxin, DON” highlight the production of pasta from durum wheat and the reported higher levels of ochratoxin in durum compared to other cultivated wheat species, making it less suitable for consumption. Keywords like “genome” and “resistance” are pivotal characteristics of Aegilops. Additionally, “*Aceria tosichella*” appears prominently, possibly due to multiple stages of resistance exhibited by Aegilops. The presence of the grain softness protein “puroindoline” strengthens its acceptability, donated by Aegilops. Spelt, a close relative of common wheat, demonstrates a significant research trend, with thousands of annual publications on hexaploid wheat. Enriched keywords such as “stress” and “yield” reflect current scientific demands in wheat research. Furthermore, hierarchical keywords like “bio-control” and “celiac disease” warrant consideration for future research on hexaploid wheat. In conclusion, this analysis offers valuable insights into the genetic, evolutionary, and practical aspects of wheat biology and agriculture.

## Data availability

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

## Financial support

The authors declare that no external funds, grants or other support was received for conducting this study.

## CRediT authorship contribution statement

**Tilak Chandra:** Writing – review & editing, Writing – original draft, Visualization, Conceptualization. **Jagajjit Sahu:** Software, Resources, Methodology, Formal analysis, Data curation. **Sarika Jaiswal:** Project administration, Methodology, Investigation. **Mir Asif Iqbal:** Project administration, Methodology, Investigation. **Dinesh Kumar:** Project administration, Methodology, Investigation.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We express our gratitude to the Indian Council of Agricultural Research (ICAR), Ministry of Agriculture and Farmers’ Welfare, Government of India, for the provision of infrastructure and other essential facilities at ICAR-Indian Agricultural Statistics Research Institute (ICAR-IASRI), New Delhi, India, established under the National Agricultural Innovation Project, funded by the ICAR-New Delhi.

## Abbreviations

DON	Deoxynivalenol
NLP	Natural Language Processing
GWAS	Genome wide association studies
QTL	Quantitative trait loci
CRISPR	Clustered regularly interspaced short palindromic repeats
VRN1	Vearnilization 1
USA	United States of America
FHB	<i>Fusarium</i> head blight
TD	<i>Triticum dicoccoides</i>
Cd	Cadmium
TA	<i>Triticum aestivum</i>
TM	Triticum monococcum
HKT1	High-affinity K <sup>+</sup> transporter 1
Adh1	Alcohol dehydrogenase I
CAT	Catalase
TU	Triticum urartu
H <sub>2</sub> O <sub>2</sub>	Hydrogen peroxide;
PMID	PubMed identity
W	Water
N	Nitrogen
API	Application programming interface
CHL	Chlorophyll
P	Phosphorus
SS	Stresses
POD	Peroxidase
SOD	Superoxide dismutase
APX	Ascorbate peroxidase
GST	Glutathione-s-transferase
BW	Bread wheat
DW	Durum wheat
COX2	Cyclooxygenase 2
TD	Triticum turgidum
PPO	Polyphenol oxidase
C	Carbon
ABA	Abscisic acid
AT	Aceria tosichella
GSP	Grain softness protein
Pina; b	Puroindoline a, b
SSR	Simple sequence repeat

## References

- [1] P. Giraldo, E. Benavente, F. Manzano-Agugliaro, E. Gimenez, Worldwide research trends on wheat and barley: a bibliometric comparative analysis, *Agronomy* 9 (7) (2019 Jul 3) 352.
- [2] P. Gustafson, O. Raskina, X. Ma, E. Nevo, Wheat Evolution, Domestication, and Improvement. *Wheat Science and Trade*, vol. 29, 2009 May, p. 3, 0.
- [3] J.H. Peng, D. Sun, E. Nevo, Domestication evolution, genetics and genomics in wheat, *Mol. Breed.* 28 (2011 Oct) 281–301.
- [4] S. Rahman, S. Islam, Z. Yu, M. She, E. Nevo, W. Ma, Current progress in understanding and recovering the wheat genes lost in evolution and domestication, *Int. J. Mol. Sci.* 21 (16) (2020 Aug 14) 5836.
- [5] F. Pourazari, G. Vico, P. Ehsanzadeh, M. Weih, Contrasting growth pattern and nitrogen economy in ancient and modern wheat varieties, *Can. J. Plant Sci.* 95 (5) (2015 Sep) 851–860.
- [6] Stallknecht GF, Gilbertson KM, Ranney JE. Alternative Wheat Cereals as Food Grains: Einkorn, Emmer, Spelt, Kamut, and Triticale.
- [7] Y. Şahin, F.P. Karakas, Chemical Composition of Einkorn (*Triticum monococcum* ssp. *monococcum*), Emmer (*Triticum dicoccum*), and Spelt (*Triticum spelta*), *In Ancient Wheat*. 30 (2022 Aug) 119–145. Cham: Springer International Publishing.
- [8] K. Gaurav, S. Arora, P. Silva, J. Sánchez-Martín, R. Horsnell, L. Gao, G.S. Brar, V. Widrig, Raupp W. John, N. Singh, S. Wu, Population genomic analysis of *Aegilops tauschii* identifies targets for bread wheat improvement, *Nat. Biotechnol.* 40 (3) (2022 Mar) 422–431.
- [9] E.S.M. Abdel-Aal, I. Rabalski, Bioactive compounds and their antioxidant capacity in selected primitive and modern wheat species, *Open Agric. J.* 2 (1) (2008).
- [10] H. Wieser, P. Koehler, K.A. Scherf, The two faces of wheat, *Front. Nutr.* 7 (2020 Oct 21) 517313.
- [11] B.Z. Yuan, J. Sun, Research trends and status of wheat (*Triticum aestivum* L.) based on the Essential Science Indicators during 2010–2020: a bibliometric analysis, *Cereal Res. Commun.* (2021) 1–2.
- [12] J. Sahu, T. Chandra, S. Jaiswal, M.A. Iqbal, D. Kumar, Millet research status and prospects for alleviating food insecurity through a text-mining approach, *J. Agric. Sci.* 161 (5) (2023 Oct) 633–644.

- [13] V.S. Marchioro, G. Benin, D. Meira, C. Meier, T. Olivoto, L.A. Klein, L.G. Woyann, M. Toebe, A.H. Bozi, A scientometric view of wheat blast: the new catastrophic threat to wheat worldwide, *J. Plant Pathol.* 105 (1) (2023 Feb) 121–128.
- [14] C. Cecchini, P. Menesatti, F. Antonucci, C. Costa, Trends in research on durum wheat and pasta, a bibliometric mapping approach, *Cereal Chem.* 97 (3) (2020) 581–588.
- [15] A. Blanco, Structure and trends of worldwide research on durum wheat by bibliographic mapping, *Int. J. Plant Biol.* 15 (1) (2024 Feb 17) 132–160.
- [16] M. Aria, C. Cuccurullo, bibliometrix: an R-tool for comprehensive science mapping analysis, *J. Informetr.* 11 (4) (2017 Nov 1) 959–975.
- [17] G. Csardi, T. Nepusz, The igraph software, *Complex Syst.* 1695 (2006) 1–9.
- [18] M. Bastian, S. Heymann, M. Jacomy, Gephi: an open source software for exploring and manipulating networks, *Proc. Int. AAAI Conf. Web Soc. Media* 3 (1) (2009 Mar 19) 361–362.
- [19] C.H. Wei, A. Allot, R. Leaman, Z. Lu, PubTator central: automated concept annotation for biomedical full text articles, *Nucleic Acids Res.* 47 (W1) (2019 Jul 2) W587–W593.
- [20] J. Rani, A.R. Shah, S. Ramachandran, pubmed. mineR: an R package with text-mining algorithms to analyse PubMed abstracts, *J. Biosci.* 40 (2015 Oct) 671–682.
- [21] C. Ozden, M. Bulut, B. Şen, Covid-19 and food security: impact on wheat, *Turk. J. Agric.-Food Sci. Technol.* 10 (4) (2022 Apr 30) 549–554.
- [22] Z. Xu, Z. Yu, J. Zhao, Theory and application for the promotion of wheat production in China: past, present and future, *J. Sci. Food Agric.* 93 (10) (2013 Aug 15) 2339–2350.
- [23] P. Zhou, L. Leydesdorff, The emergence of China as a leading nation in science, *Res. Pol.* 35 (1) (2006 Feb 1) 83–104.
- [24] Z. He, X. Xia, Y. Zhang, Y. Zhang, Y. Xiao, X. Chen, S. Li, Y. Hao, A. Rasheed, Z. Xin, Q. Zhuang, China-CIMMYT collaboration enhances wheat improvement in China, *Front. Agric. Sci. Eng.* 6 (3) (2019 Sep 15) 233–239.
- [25] Z. Sramková, E. Gregová, E. Šturdík, Chemical composition and nutritional quality of wheat grain, *Acta Chim. Slovaca* 2 (1) (2009) 115–138.
- [26] A. Hossain, M. Skalicky, M. Brestic, S. Maitra, M. Ashraf Alam, M.A. Syed, J. Hossain, S. Sarkar, S. Saha, P. Bhadra, T. Shankar, Consequences and mitigation strategies of abiotic stresses in wheat (*Triticum aestivum* L.) under the changing climate, *Agronomy* 11 (2) (2021 Jan 28) 241.
- [27] D.K. Saini, Y. Chopra, J. Singh, K.S. Sandhu, A. Kumar, S. Bazzar, P. Srivastava, Comprehensive evaluation of mapping complex traits in wheat using genome-wide association studies, *Mol. Breed.* 42 (2022 Jan) 1–52.
- [28] Z.A. Mir, T. Chandra, A. Saharan, N. Budhlakoti, D.C. Mishra, M.S. Saharan, R.R. Mir, A.K. Singh, S. Sharma, V.K. Vikas, S. Kumar, Recent advances on genome-wide association studies (GWAS) and genomic selection (GS); prospects for Fusarium head blight research in Durum wheat, *Mol. Biol. Rep.* 50 (4) (2023 Apr) 3885–3901.
- [29] M. Zaharieva, P. Monneveux, Cultivated einkorn wheat (*Triticum monococcum* L. subsp. *monococcum*): the long life of a founder crop of agriculture, *Genet. Resour. Crop Evol.* 61 (2014 Mar) 677–706.
- [30] A. Hidalgo, A. Brandolini, Nutritional properties of einkorn wheat (*Triticum monococcum* L.), *J. Sci. Food Agric.* 94 (4) (2014 Mar 15) 601–612.
- [31] A. Serfling, S.E. Templar, P. Winter, F. Ordon, Microscopic and molecular characterization of the prehaustorial resistance against wheat leaf rust (*Puccinia triticina*) in Einkorn (*Triticum monococcum*), *Front. Plant Sci.* 7 (2016 Nov 9) 1668.
- [32] A. Nishiura, S. Kitagawa, M. Matsumura, Y. Kazama, T. Abe, N. Mizuno, S. Nasuda, K. Murai, An early-flowering einkorn wheat mutant with deletions of PHYTOCLOCK 1/LUX ARRHYTHMO and VERNALIZATION 2 exhibits a high level of VERNALIZATION 1 expression induced by vernalization, *J. Plant Physiol.* 222 (2018 Mar 1) 28–38.
- [33] N. Shitsukawa, C. Ikari, S. Shimada, S. Kitagawa, K. Sakamoto, H. Saito, H. Ryuto, N. Fukunishi, T. Abe, S. Takumi, S. Nasuda, The einkorn wheat (*Triticum monococcum*) mutant, maintained vegetative phase, is caused by a deletion in the VRN1 gene, *Gene Genet. Syst.* 82 (2) (2007) 167–170.
- [34] F.J. Zeller, M.C. Cermenio, T.E. Miller, Cytological analysis on the distribution and origin of the alien chromosome pair conferring blue aleurone color in several European common wheat (*Triticum aestivum* L.) strains, *Theor. Appl. Genet.* 81 (1991 Apr) 551–558.
- [35] S. Lewis, M.E. Faricelli, M.L. Appendino, M. Valárik, J. Dubcovsky, The chromosome region including the earliness per se locus Eps-Am1 affects the duration of early developmental phases and spikelet number in diploid wheat, *J. Exp. Bot.* 59 (13) (2008 Oct 1) 3595–3607.
- [36] M. Zaharieva, N.G. Ayana, A.A. Hakimi, S.C. Misra, P. Monneveux, Cultivated emmer wheat (*Triticum dicoccon* Schrank), an old crop with promising future: a review, *Genet. Resour. Crop Evol.* 57 (2010 Aug) 937–962.
- [37] L.K. Kucek, E. Dyck, J. Russell, L. Clark, J. Hamelman, S. Burns-Leader, S. Senders, J. Jones, D. Benschler, M. Davis, G. Roth, Evaluation of wheat and emmer varieties for artisanal baking, pasta making, and sensory quality, *J. Cereal. Sci.* 74 (2017 Mar 1) 19–27.
- [38] A. Sheibanirad, A. Mirlohi, R. Mohammadi, P. Ehsanzadeh, B.E. Sayed-Tabatabaei, Cytogenetic and crossability studies in hulled wheat collected from Central Zagros in Iran, *Plant Systemat. Evol.* 300 (2014 Oct) 1895–1901.
- [39] Z. Abdehpour, P. Ehsanzadeh, Concurrence of ionic homeostasis alteration and dry mass sustainment in emmer wheats exposed to saline water: implications for tackling irrigation water salinity, *Plant Soil* 440 (2019 Jul 14) 427–441.
- [40] S. Fatholah, P. Ehsanzadeh, H. Karimmojeni, Ancient and improved wheats are discrepant in nitrogen uptake, remobilization, and use efficiency yet comparable in nitrogen assimilating enzymes capabilities, *Field Crops Res.* 249 (2020 Apr 1) 107761.
- [41] D.S. Gaikpa, B. Lieberherr, H.P. Maurer, C.F. Longin, T. Miedaner, Comparison of rye, triticale, durum wheat and bread wheat genotypes for Fusarium head blight resistance and deoxynivalenol contamination, *Plant Breed.* 139 (2) (2020 Apr) 251–262.
- [42] M. Vergine, A. Aprile, E. Sabella, A. Genga, M. Siciliano, P. Rampino, M.S. Lenucci, A. Luvisi, L.D. Bellis, Cadmium concentration in grains of durum wheat (*Triticum turgidum* L. subsp. *durum*), *J. Agric. Food Chem.* 65 (30) (2017 Aug 2) 6240–6246.
- [43] S. Schaarschmidt, C. Fahl-Hassek, The fate of mycotoxins during the processing of wheat for human consumption, *Compr. Rev. Food Sci. Food Saf.* 17 (3) (2018 May) 556–593.
- [44] A. Schneider, I. Molnár, M. Molnár-Láng, Utilisation of *Aegilops* (goatgrass) species to widen the genetic diversity of cultivated wheat, *Euphytica* 163 (2008 Sep) 1–9.
- [45] M. Kishii, An update of recent use of *Aegilops* species in wheat breeding, *Front. Plant Sci.* 10 (2019 May 9) 452618.
- [46] R. Sanchez-Bragado, M.D. Serret, J.L. Araus, The nitrogen contribution of different plant parts to wheat grains: exploring genotype, water, and nitrogen effects, *Front. Plant Sci.* 7 (2017 Jan 9) 215227.
- [47] S.G. Carrera, H. Davis, L. Aguirre-Rojas, M. Murugan, C.M. Smith, Multiple categories of resistance to wheat curl mite (Acari: eriophyidae) expressed in accessions of *Aegilops tauschii*, *J. Econ. Entomol.* 105 (6) (2012 Dec 1) 2180–2186.
- [48] M.F. Gautier, P. Cosson, A. Guirao, R. Alary, P. Joudrier, Puroindoline genes are highly conserved in diploid ancestor wheats and related species but absent in tetraploid *Triticum* species, *Plant Sci.* 153 (1) (2000 Apr 14) 81–91.
- [49] A.A. Elbashir, Y.S. Gorafi, I.S. Tahir, J.S. Kim, H. Tsujimoto, Wheat multiple synthetic derivatives: a new source for heat stress tolerance adaptive traits, *Breed Sci.* 67 (3) (2017) 248–256.
- [50] M. Garg, H. Tsujimoto, R.K. Gupta, A. Kumar, N. Kaur, R. Kumar, V. Chunduri, N.K. Sharma, M. Chawla, S. Sharma, J.K. Munday, Chromosome specific substitution lines of *Aegilops geniculata* alter parameters of bread making quality of wheat, *PLoS One* 11 (10) (2016 Oct 18) e0162350.
- [51] W. Yue, H. Zhang, X. Sun, N. Su, Q. Zhao, Z. Yan, S. Weining, H. Yue, The landscape of autophagy-related (ATG) genes and functional characterization of TaVAMP727 to autophagy in wheat, *Int. J. Mol. Sci.* 23 (2) (2022 Jan 14) 891.
- [52] Z.S. Peng, J. Yang, G.C. Zheng, Cytomixis in pollen mother cells of new synthetic hexaploid amphidiploid (*Aegilops tauschii* × *Triticum turgidum*), *Cytologia* 68 (4) (2003) 335–340.
- [53] M. Feldman, A.A. Levy, Genome evolution due to allopolyploidization in wheat, *Genetics* 192 (3) (2012 Nov 1) 763–774.
- [54] D. Soto-Gómez, P. Pérez-Rodríguez, Sustainable agriculture through perennial grains: wheat, rice, maize, and other species. A review, *Agric. Ecosyst. Environ.* 325 (2022 Feb 28) 107747.