

Global Research Trends in Aquatic Stress Transcriptomics and Genomics: A Bibliometric and Science-Mapping Analysis, 2005–2025

Prasanth Kumar Guduru

prasanthzoology01@gmail.com

Sri Krishnadevaraya University

Rajasekhar Gutha

Sri Krishnadevaraya University

Research Article

Keywords: Aquatic stress, transcriptomics, genomics, zebrafish, RNA-seq, bibliometric analysis

Posted Date: October 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11128845/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Genomic model fish, including zebrafish (*Danio rerio*), medaka (*Oryzias latipes*), and Nile tilapia (*Oreochromis niloticus*), are widely used to study the transcriptomic and genomic responses to environmental and physiological stress. To identify prominent trends in this field of research, we conducted a bibliometric analysis using biblioshiny. We examined 1,126 Scopus-indexed documents published between 2005 and 2025, including 1,080 articles and 46 reviews. The search combined terms related to fish genomic models with genomic data and physiological stress. Analyses were conducted using bibliometrix in R through biblioshiny with results verified using VOSviewer. Annual publication output grew steadily from single-digit counts in the mid-2000s to 241 documents in 2025. A logistic growth model projected a peak in research output around the year 2035.2 ($R^2 = 0.903$) suggesting expansion rather than a sharp decline. Bradford's Law highlighted 13 core journals. Author productivity followed a fitted Lotka distribution with a β value of 2.18 among 5,701 authors. China led in output while its collaboration link with the United States was the strongest. Co-citation and keyword analyses uncovered two research areas: bioinformatics and RNA-seq methods and applied stress physiology and toxicology. These domains are increasingly overlapping, showing connections between stress, transcriptomic profiling, oxidative stress, toxicity, and mitochondrial dysfunction or ferroptosis. Trend-topic analysis identified neurotoxicity, cardiotoxicity, mitochondrial dysfunction, ferroptosis and microplastics as leading research topics between 2023 and 2025.

1. INTRODUCTION

Zebrafish, medaka, and Nile tilapia are some of the commonly used genomic models for studying how aquatic vertebrates respond to environmental and physiological stress at the molecular level (Kimmel et al., 1995; Westerfield, 2000; Howe et al., 2013). The development of RNA sequencing and transcriptome- and network-based analytical methods – including differential-expression frameworks and Weighted Gene Co-expression Network Analysis (WGCNA) – has made it easier to study gene expression and co-expression patterns across the transcriptome in these species (Lieschke & Currie, 2007; Kültz, 2005). Because oxidative stress is a common cellular response to both abiotic stressors (hypoxia, salinity, thermal extremes, osmotic shock, cold) and chemical or contaminant exposure, this literature substantially overlaps with ecotoxicology: many studies that describe their results in terms of "stress" are actually investigations of heavy-metal, pharmaceutical, endocrine-disruptor, or microplastic exposure that quantify oxidative-stress and antioxidant-enzyme endpoints (Hill et al., 2005). Zebrafish plays a role in this body of work, complemented by medaka for comparative salinity studies and by tilapia for research relevant to aquaculture and thermal-tolerance.

Even though this literature has grown quickly, no bibliometric study has yet mapped its intellectual structure, collaboration geography, and thematic evolution as a defined field. There is also no understanding of how much overlap exists between searches based on abiotic stress and comparative genomics terms and those focused on oxidative stress, ecotoxicology, and toxicogenomics. This overlap emerges strongly in the retrieved articles, although it was not necessarily the original focus of the

search. This gap leads to six research questions: (RQ1) What is the growth pattern of this literature and what does it tell us about the current stage of development of the field? (RQ2) Which journals publish the majority of the core literature according to Bradford's Law and how concentrated is the citation impact? (RQ3) How is authorship distributed across the field following Lotka's Law and which authors, institutions, and countries are leading the way? (RQ4) What do the collaboration networks among authors, institutions, and nations reveal about how the field is organized socially? (RQ5) What is the intellectual and conceptual structure of the field as shown through co-citation and keyword co-occurrence analyses? (RQ6) How have research topics changed over time and what defines the research front?

2. MATERIALS AND METHODS

2.1 Data Source and Search Strategy

Bibliographic records were retrieved from Scopus (Elsevier) due to its wide interdisciplinary coverage of the genomics/stress-physiology intersection and its native compatibility with bibliometrix. The exact search query performed on the Title, Abstract, and Keyword fields was:

```
TITLE-ABS-KEY(("zebrafish" OR "Danio rerio" OR "medaka" OR "Oryzias latipes" OR "tilapia" OR "Oreochromis niloticus") AND ("transcriptomic*" OR "genomic*" OR "WGCNA" OR "Weighted Gene Co-expression Network Analysis" OR "differential expression" OR "network analysis" OR "comparative genomic*") AND ("stress*" OR "hypoxia" OR "salinity" OR "thermal*" OR "heat" OR "osmotic" OR "cold"))
```

The search strategy did not specifically include ecotoxicology, contaminant, or microplastics-related terms. Hence, the prevalence of ecotoxicology-related terms in the search outcome suggests that it emerged from the corpus rather than the search strategy. This is discussed in detail in Section 4.2, with particular emphasis on its status as an emergent property of the identified corpus. The search result was first obtained in Scopus on 12th September 2026, which initially retrieved 1,405 records. Following that, the time range was narrowed to between 2005 and 2025 within bibliometrix/biblioshiny, resulting in the final corpus of 1,126 Scopus-indexed documents (Online Resource 1 (Fig. S1)). The publication-year filter was restricted to 2005–2025, with 2025 treated as a completed publication year. The final corpus includes both research articles and 46 reviews. As such, this paper will henceforth refer to the documents identified by the search strategy as “Scopus-indexed documents” instead of “journal articles.” Every descriptive, citation, network, and thematic metric reported below is based on this final set of 1,126 documents. The search was restricted to English-language documents at the Scopus retrieval stage; the 2005–2025 publication-year filter was subsequently applied within bibliometrix/biblioshiny.

2.2 Analytical Tools and Reproducibility

Descriptive, citation, and collaboration analyses were conducted in R (v4.5.2) using bibliometrix (Aria & Cuccurullo, 2017) (package v5.5.0, biblioshiny interface). Keyword co-occurrence clustering was cross-

validated in VOSviewer v1.6.21 (van Eck & Waltman, 2010). To support reproducibility, the specific settings used were: author keywords as the unit of analysis for co-occurrence and thematic mapping, with a minimum-occurrence threshold of 5; full counting (each co-occurrence or collaboration link counted once per document, without fractionalization); association-strength normalization in VOSviewer (default), with clustering via VOSviewer's smart local moving algorithm (resolution parameter = 1.0, minimum cluster size = 1) and, in bibliometrix networks, Louvain community detection at default resolution; Bradford zones computed via bibliometrix's automatic equal-articles-per-zone partition, yielding the reported core/Zone 2/Zone 3 split; Lotka's Law exponent estimated by non-linear least squares with Kolmogorov–Smirnov goodness-of-fit testing against both the classical exponent ($\beta = 2$) and the empirically fitted exponent; the strategic (centrality–density) thematic map built from the author-keyword co-word network using bibliometrix's Walktrap clustering default; and the life-cycle and cumulative-growth curves fitted to annual and cumulative publication counts by non-linear least squares (R's nls function) using a four-parameter logistic growth model, from which the reported peak year and diffusion percentiles (50/90/99%) are model outputs rather than observed data.

Global citations (citations received from the entire Scopus database) and local citations (citations received from other documents within the retrieved corpus), together with each author's local h-index, were computed directly in bibliometrix. Country, author, and institutional collaboration networks used full counting, meaning a single multi-country or multi-author document contributes one link per collaborating pair; consequently, aggregate country or author frequencies reported below can exceed the number of documents. Network figures render node size proportional to the relevant relevance metric (documents, citations, or occurrences, depending on the analysis) and edge thickness proportional to co-occurrence or collaboration frequency; node colours denote data-driven community-detection clusters and do not carry a fixed meaning across figures. Where a figure combines two panels (Figs. 1, 4, 5, and 6), each panel was generated independently in bibliometrix or VOSviewer and combined post hoc for presentation; no data were altered in this process. An AI-based language model was used during manuscript preparation to assist with drafting, language refinement, and structural editing of the text describing the above analyses; all scientific content, interpretations, and conclusions are the authors' own and were independently verified (see Statements and Declarations).

3. RESULTS

3.1 Growth Dynamics

Annual scientific production has increased from single-digit counts in the mid-2000s to a marked acceleration from 2018 onward (Fig. 1a). Output remained moderate between 2005 and 2017 before increasing to 44 documents in 2018 and 79 in 2020. After a brief plateau at 73–80 documents during 2021–2022, annual production accelerated to 96 documents in 2023, 153 in 2024, and 241 in 2025. Table 1 summarizes the main characteristics of the analyzed corpus and the principal bibliometric indicators.

Table 1
Main information about the analyzed corpus.

Metric	Value
Timespan	2005–2025
Search date	12 September 2026
Sources (journals) analyzed	354 (Core = 13; Zone 2 = 62; Zone 3 = 279)
Documents (final corpus)	1,126 (1,080 articles + 46 reviews)
Documents (raw export, pre-filtering)	1,405
Contributing authors	5,701
Bradford multiplier (k)	4.63
Lotka's Law β coefficient (fitted)	2.18
Most productive author	Li X. (n = 42)
Most impactful author (local h-index)	Wang J. (h = 19)
Most globally cited document	Pauli A. et al., 2012, Genome Research (677 citations)
Country with highest affiliation frequency	China (1,428 country-document associations)
Life-cycle model projected peak year	2035.2 ($R^2 = 0.903$; model-dependent)

A logistic life-cycle model fitted to the annual series (Fig. 1b) produced a model-projected peak year of 2035.2 ($R^2 = 0.903$); this figure is a statistical extrapolation contingent on current growth dynamics persisting and should be interpreted as an indicator that the literature remains in an expansionary phase, not as a precise forecast of when output will peak (Section 4.1; Online Resource 1 (Fig. S2) shows the corresponding cumulative growth curve and diffusion percentiles). The shape of the curve is still informative in relative terms: the observed annual counts stay closely below the fitted logistic through 2022 and begin to approach it only in the last three years of the series, a pattern consistent with a rapidly expanding phase of the literature rather than one already approaching a clear decline. The projected peak year is model-dependent and should therefore be interpreted as a directional indication of continued expansion rather than as a precise forecast.

3.2 Citation Impact

The average number of citations per year per document (Online Resource 1 (Fig. S3)) fluctuated considerably, with local peaks around 2007, 2012, and 2019, and a decline toward 2023–2025 that reflects citation-window immaturity for recently published articles rather than declining quality. Table 2 lists the ten most globally cited documents which are dominated by foundational comparative-genomics and transcriptomics papers; the most globally cited document, Pauli et al. (2012), added 677 global citations. Table 3 lists the ten most locally cited documents — those most cited by other papers within

this corpus – dominated instead by applied transcriptome-profiling and stress-physiology studies (Long et al., 2012; Long et al., 2013). This divergence indicates that the field gains its external visibility from foundational genomics while building an increasingly distinct internal citation base. Several of the most globally cited documents – for example, comparative-genomics and cardiomyopathy-genetics papers published in Genome Research and the American Journal of Human Genetics – are only loosely connected to fish stress biology as a topic, indicating that some of the corpus's external citation impact may partly reflect the broader influence of genomics and genetics scholarship that happens to satisfy the search criteria rather than from stress-genomics-specific influence. By contrast, the locally cited documents are, without exception, empirical stress-transcriptomics studies conducted directly in zebrafish or related species, reinforcing that local citation counts are a more topic-faithful indicator of influence within this literature than global citation counts alone. This global/local divergence also has a practical implication for literature reviews conducted within the field: relying solely on globally-cited "landmark" papers to define the field's core knowledge base risks over-weighting broadly influential genomics methodology at the expense of the empirical stress-transcriptomics studies that the field's own practitioners cite most heavily among themselves.

Table 2
Top 10 most globally cited documents.

Rank	Document	Global Citations
1	Pauli A. et al., 2012, Genome Research	677
2	Kari G. et al., 2007, Clinical Pharmacology & Therapeutics	421
3	Hill J.T. et al., 2013, Genome Research	414
4	Nakatani Y. et al., 2007, Genome Research	405
5	Huang S. et al., 2008, Genome Research	347
6	Nielsen C.B. et al., 2007, RNA	331
7	Simões F.C. et al., 2020, Nature Communications	328
8	Qiao R. et al., 2019, Science of the Total Environment	327
9	Norton N. et al., 2011, American Journal of Human Genetics	319
10	Scott G.R. et al., 2012, PNAS	296

Table 3
Top 10 most locally cited documents (citations received from within the analyzed corpus).

Rank	Document	Local Citations
1	Long Y. et al., 2012, PLOS ONE	16
2	Long Y. et al., 2013, BMC Genomics	10
3	Clark K.J. et al., 2011, Reviews in the Neurosciences	9
4	Aanes H. et al., 2011, Genome Research	8
5	Lai K.P. et al., 2015, BMC Genomics	8
6	Vergauwen L. et al., 2010, Comp. Biochem. Physiol. A	7
7	Li C.G. et al., 2014, Comp. Biochem. Physiol. B	7
8	Zhou W. et al., 2021, Journal of Hazardous Materials	6
9	Zhao Y. et al., 2015, Genetics and Molecular Research	6
10	Wang Q. et al., 2014, PLOS ONE	6

3.3 Bradford's Law and Core Journals

The source distribution closely followed Bradford's Law (Bradford, 1934) (Fig. 2; $R^2 = 0.9935$; Bradford multiplier $k = 4.63$; Kolmogorov–Smirnov $p = 0.255$). Of 354 total sources, a core of 13 journals accounted for as many documents as the subsequent 62 Zone 2 journals combined and the remaining 279 Zone 3 journals combined. Table 4 lists the ten most relevant sources: Science of the Total Environment, Ecotoxicology and Environmental Safety, and Aquatic Toxicology led the ranking, followed by Journal of Hazardous Materials and BMC Genomics. This composition – environmental-toxicology together with genomics journals focussed on methodology— results directly from the "stress*" wildcard capturing oxidative-stress endpoints reported in both abiotic- and contaminant-exposure studies (Section 4.2), and positions ecotoxicology/toxicogenomics as an important emergent component of a corpus originally defined around physiological stress terms rather than toxicological ones.

Table 4
Top 10 most relevant sources by number of documents.

Rank	Source	N. of Documents
1	Science of the Total Environment	46
2	Ecotoxicology and Environmental Safety	44
3	Aquatic Toxicology	41
4	Journal of Hazardous Materials	34
5	BMC Genomics	32
6	Environmental Pollution	29
7	Fish and Shellfish Immunology	29
8	PLOS ONE	26
9	Genome Research	25
10	International Journal of Molecular Sciences	25

The close match to Bradford's model ($R^2 = 0.9935$) also has a practical implication for literature discovery: The 13 core journals make up roughly one-third of the corpus, while comprehensive coverage requires monitoring the substantially larger Zone 2 and Zone 3 tail. The sharp decline from Core to Zone 3 – 279 journals contributing the same volume as 13 – indicates a highly scattered periphery in which individual journals publish stress-genomics work only occasionally, consistent with a field whose methodology (transcriptomics, network analysis) is portable across many disciplinary venues rather than confined to a small set of specialist outlets.

3.4 Author Productivity and Impact

Author productivity approximated Lotka's Law (Lotka, 1926) (Fig. 3), with a fitted exponent $\beta = 2.18$ and constant $C = 0.3069$ ($R^2 = 0.9129$) across the 5,701 authors contributing to the corpus. The fitted curve differed significantly from the classical inverse-square distribution (KS $p = 0.00141$ against $\beta = 2$) but not from the empirical data (KS $p = 0.0794$), indicating a productivity distribution somewhat more concentrated than the classical Lotka pattern: approximately 82% of contributing authors published only a single document in the corpus.

Table 5 lists the ten most productive authors, led by Li X. (42 documents), Zhang Y. (41), and Li Y. and Wang Y. (39 each). Author impact, measured by local h-index (also in Table 5), produces a partially different ranking: Wang J. attains the highest h-index (19) despite ranking fifth in raw productivity, ahead of the most prolific author, Li X. ($h = 17$). Local citation impact, rather than total output volume, therefore distinguishes contributors with greater influence within the retrieved corpus. All ten of the most productive authors carry Chinese surnames, consistent with the country-level concentration reported in Section 3.5, and several surnames (Wang, Li, Zhang) recur across the top-ten list, which – combined

with the surname-and-initial disambiguation limitation discussed in Section 4.5 – means the productivity figures reported here are best interpreted as approximate rather than exact for any single named individual.

Table 5
Top 10 most productive and most impactful authors.

Rank	Author	N. of Documents	Local h-index
1	Li X.	42	17
2	Zhang Y.	41	15
3	Li Y.	39	15
4	Wang Y.	39	17
5	Wang J.	35	19
6	Wang X.	32	14
7	Zhang J.	31	—
8	Chen L.	30	16
9	Li J.	30	—
10	Liu Y.	29	15
Note: "—" indicates that the author did not appear in the top-10 local h-index ranking.			

3.5 Geographic Distribution and Collaboration

China had the highest country-affiliation frequency in the corpus – 1,428 country-document associations under full counting, a figure that exceeds the 1,126 total documents because collaborative articles with authors based in more than one country are attributed to each contributing country. The United States ranked second, with a second tier comprising Canada, Brazil, India, several European countries, and Australia (Fig. 4a). The international collaboration network (Fig. 4b; an alternative geographic rendering appears in Online Resource 1 (Fig. S13)) identifies China–USA as the single strongest collaboration tie, with China additionally anchoring a regional East Asian/Southeast Asian cluster (Hong Kong, Korea, Japan, Thailand) and the USA anchoring a broad transatlantic/Western cluster (United Kingdom, Germany, Sweden, Canada, and others). Smaller clusters link Australia, Brazil, South Africa, and other emerging-economy nations, while several Middle Eastern, South Asian, and African countries remain peripheral to the network core.

The disproportionate width of the China–USA tie relative to all other links in Fig. 4b indicates that international collaboration is strongly concentrated around the China–USA link, while most other bilateral ties are substantially weaker. This pattern is common in emerging, geographically concentrated scientific fields, where a small number of high-output countries anchor most cross-border co-authorship

while mid- and lower-output countries collaborate primarily within their own regional or economic blocs (as seen in the distinct Western-European/Commonwealth, Southern-Hemisphere/emerging-economy, and Middle Eastern/South Asian clusters in Fig. 4b) rather than directly with the dominant pair.

3.6 Institutional and Author Collaboration

Institutional collaboration (Online Resource 1 (Fig. S4)) resolves into several largely nationally bounded clusters rather than one integrated global network, anchored by City University of Hong Kong, the Institute of Hydrobiology (Chinese Academy of Sciences), and Shanghai Ocean University, with only thin cross-cluster bridging ties. A smaller number of institutions — notably City University of Hong Kong, which carries the thickest single tie in the network to the University of Edinburgh — do bridge otherwise separate regional clusters, indicating that international institutional collaboration, while limited in volume, is not entirely absent but concentrated through a small number of consistently well-connected "bridge" institutions rather than distributed broadly across the field. The author collaboration network (Online Resource 1 (Fig. S5)) similarly resolves into dense "invisible college" clusters, the largest anchored by Wang-surnamed authors (Wang J., Wang Y., Wang X., Wang Z.) and a second by Zhang Y., Li X., Li J., and Li Y.; because Scopus author disambiguation relies partly on surname-and-initial conventions, some apparent fragmentation may reflect unresolved author-identity ambiguity rather than genuinely separate individuals (Section 4.5). Both networks share a common structural feature: dense, well-connected clusters internally, but comparatively few edges bridging clusters, whether those clusters are defined by institution or by co-authorship group. This structural correspondence between institutional and individual-author levels suggests the fragmentation observed in Section 3.5 is not an artifact of any single level of analysis but a consistent feature of how this field's researchers organize and collaborate — predominantly within, rather than across, national and institutional boundaries.

3.7 Co-citation Structure

Co-citation analysis identifies two major intellectual domains within the corpus. The author co-citation network (Online Resource 1 (Fig. S6)) separates a cluster of bioinformatics/RNA-seq methodologists — including Love M.I. (DESeq2) (Love et al., 2014), Livak K.J. (qPCR quantification) (Livak & Schmittgen, 2001), Dobin A. (STAR) (Dobin et al., 2013), Langmead B. (Bowtie2) (Langmead & Salzberg, 2012), and Robinson M.D. (edgeR) (Robinson et al., 2010), and Kim D. (HISAT) (Kim et al., 2015) — from a second domain of substantive stress-physiology and toxicology researchers (Zhang Y., Wang Y., Li X., Li J., Li Y., Wang J.). The former group supplies a shared, widely cited methodological toolkit; the latter constitutes the field's own applied empirical literature. The source co-citation network (Online Resource 1 (Fig. S7)) mirrors this pattern at the journal level, and the historiographic map (Online Resource 1 (Fig. S9)) shows that RNA-seq pipeline tools consolidated in the literature during approximately 2010–2015, a period that preceded and was associated with the subsequent expansion of applied stress-genomics output documented in Section 3.1 — an association rather than a demonstrated causal mechanism. A paper-level co-citation network appears in Online Resource 1 (Fig. S8). Taken together, the author-, source-, and paper-level co-citation results consistently distinguish a compact methods-oriented component from a larger and more heterogeneous applied stress-physiology/toxicology component. This consistency

across three independent co-citation units of analysis strengthens confidence that the two-domain structure reflects a genuine organizational feature of the field rather than an artifact of any single network-construction choice.

3.8 Conceptual Structure and Keyword Co-occurrence

The author-keyword co-occurrence network (Fig. 5a) and the corroborating VOSviewer map (Fig. 5b) point to a conceptual structure with three clusters. One cluster centers on zebrafish, which appears with oxidative stress, apoptosis, DNA damage, developmental, cardiotoxicity, neurotoxicity, hepatotoxicity, mitochondrial dysfunction, and ferroptosis. A second cluster centers on transcriptome, gene expression and RNA-seq appearing with fish, tilapia, medaka, hypoxia, salinity, osmoregulation and aquaculture. A smaller cluster links liver, microarray, genomics and cadmium. Zebrafish is the largest and most central node in both maps linking all three clusters. Although WGCNA, network analysis, and differential expression are explicit search terms, none appears as a prominent node in either map. This suggests that these terms are used as words inside abstracts rather than as author keywords so the prevalence of WGCNA network analysis and differential expression in the corpus is probably not captured by keyword-based visualization alone.

A word cloud of Scopus Index Keywords (Online Resource 1 (Fig. S11)) dominated by descriptors is shown only for indexing reference. The relative sizes and connectivity of the three clusters provide additional information. The oxidative-stress/toxicology cluster and the transcriptome/physiology cluster are clearly larger and tightly connected than the smaller legacy microarray/cadmium cluster. This shows that thematic evolution and trend-topic analyses suggest a move from genomics and microarray terms to transcriptome, oxidative-stress, toxicity and pathway themes and also to network and pathway descriptions of stress responses. The strong links between zebrafish and both the physiology and toxicology clusters, rather than with only one of them, provide additional bibliometric support for the convergence described in Section 4.2. Zebrafish is not the model organism of one sub-field but the common platform where both physiological-stress and toxicological-stress research take place.

3.9 Thematic Evolution and the Central Research Narrative

The results in Sections 3.7 and 3.8 show a shift over time and in ideas, starting with physiological stress, moving through transcriptomic profiling, and then extending into oxidative stress, toxicity, and mitochondrial dysfunction/ferroptosis. The two-phase thematic evolution analysis (2005–2022 vs 2023–2025; Fig. 6a), split at a turning point in annual output, reveals that earlier themes such as osmoregulation, gene expression, hypoxia, fish, and zebrafish mostly lead into two main themes after 2023: ‘zebrafish’ and ‘transcriptome.’ Meanwhile, neurotoxicity, developmental toxicity, ferroptosis, and mitochondrial dysfunction appear as areas of focus with relatively little direct connection to the earlier research topics. These connections reflect how keywords tend to appear together or follow one another in the published literature, not biological cause-and-effect relationships in any single organism or experiment. For example, the flow from "hypoxia" to "stress", in Fig. 6a indicates these terms are

increasingly found in the same papers over time not that every study follows a strict chain of biological events.

The trend-topics analysis (Fig. 6b) supports a three-stage progression: an early genomics-tool era (genomics, microarray, biomarker, toxicogenomics; median years ~ 2011–2015); a comparative/physiological-genomics era (bioinformatics, stress response, proteomics, medaka, gill; ~2015–2019); and the current era (oxidative stress, neurotoxicity, cardiotoxicity, mitochondrial dysfunction, microplastics; ~2023–2025), which shows both the most recent median years and the largest term frequencies in the dataset. A three-field plot integrating Sources, Authors, and Author Keywords in a single view (Online Resource 1 (Fig. S10)) provides a complementary visualization of this concentration, though it should be read as an integrated summary rather than independent statistical confirmation of the Bradford and Lotka results reported above. A whole-period strategic thematic map appears in Online Resource 1 (Fig. S12).

Read together, Sections 3.7 through 3.9 support a single, coherent narrative that this study treats as its principal contribution: the literature captured by an abiotic-stress-oriented search increasingly organizes itself around a conceptual sequence running from physiological stress exposure, through transcriptomic and network-level profiling, to oxidative stress as a shared molecular readout, onward to toxicity assessment, and finally to organelle-specific mechanistic endpoints – mitochondrial dysfunction and ferroptosis – that represent the current research front. Each stage in this sequence is supported by independent bibliometric evidence: the co-citation split (Section 3.7) identifies the methodological substrate; the keyword clusters (Section 3.8) identify the thematic destinations; and the trend-topic chronology (Fig. 6b) identifies the temporal order in which these destinations were reached. No single analysis performed here demonstrates that this sequence is causal at the level of molecular biology, but the convergence of independent bibliometric indicators on the same conceptual ordering is, at minimum, strong evidence that the bibliometric record increasingly links these topics as a connected thematic continuum rather than as separate lines of inquiry.

4. DISCUSSION

4.1 An Expansionary, Still-Developing Literature

The annual-production trend and life-cycle model together indicate that this literature remains in an expansionary phase rather than approaching saturation: output has grown markedly since 2018, and the model-projected peak beyond 2035 – while subject to considerable long-range uncertainty (Section 4.5) – is directionally consistent with a field still attracting new researchers and journals rather than one that is plateauing. This projection is model-dependent, derived from a relatively short 21-year series, and should not be read as a precise forecast. In practical terms, an expansionary field of this kind is typically characterized by continuing entry of new authors and institutions rather than consolidation among a fixed set of incumbents; the Lotka exponent and author-productivity distribution (Section 3.4), which show roughly four-fifths of contributing authors publishing only once in the corpus, are consistent with

this interpretation. However, author productivity alone cannot establish the maturity stage of a research area. The same broad signal appears in the Bradford analysis (Section 3.3): the large Zone 3 periphery of occasional-contributor journals is consistent with a field drawing contribution from a broad range of disciplinary venues.

4.2 Convergence: From Physiological Stress to Oxidative Toxicology

The central finding of this analysis is that a search built entirely around abiotic and physiological stressors and comparative-genomics methodology nonetheless returned a corpus whose core journals and dominant keywords are characteristic of ecotoxicology and toxicogenomics. This is best explained by the "stress*" wildcard: because oxidative stress is a shared molecular endpoint of both abiotic stress and chemical exposure, papers on heavy-metal, pharmaceutical, or microplastic toxicity that report oxidative-stress outcomes are captured by the same query as papers on hypoxia or salinity acclimation. This overlap can also be interpreted as a substantive finding, although it should be considered alongside the limitations of the search strategy: it reveals a conceptual convergence between physiological stress research and oxidative-stress toxicology, reflected in the increasing co-occurrence of stress, toxicity, mitochondrial dysfunction, and ferroptosis-related themes (Dixon et al., 2012). Ecotoxicology and toxicogenomics should accordingly be understood as an important emergent component of this literature, expressed most visibly through the recent rise of microplastic-related keywords (Cole et al., 2011; Horton et al., 2017), rather than as the field's original defining identity. This reframing may be relevant to how the field is reviewed and future research priorities are considered: Treating stress physiology and oxidative toxicology as two separate literatures, as earlier disciplinary boundaries might suggest, risks fragmenting a body of work that the bibliometric evidence indicates has substantial thematic and methodological overlap at the level of shared model organisms, overlapping analytical approaches, and shared molecular endpoints. A more accurate framing, consistent with the evidence presented here, is that fish stress genomics has increasingly provided a shared methodological framework for both classical environmental-physiology and contaminant-toxicology research.

4.3 Two Major Intellectual Domains

The co-citation and historiographic analyses show that this literature draws on two major intellectual domains: a compact, widely cited set of bioinformatics and RNA-seq methodology papers, and a larger, increasingly self-referential applied stress-physiology/toxicology literature. The historiographic sequencing places RNA-seq pipeline consolidation (approximately 2010–2015) before the subsequent expansion in applied output; the emergence of RNA-seq preceded and was associated with this expansion, though the data available here support association rather than a demonstrated causal mechanism. This two-domain structure is a common feature of applied genomics fields more broadly: a comparatively small, high-citation methodological layer is imported from computational biology, while a larger and more heterogeneous applied layer develops its own internal citation practices over time. What distinguishes the present field is the specific direction that applied layer has taken — converging on oxidative-stress and mitochondrial-toxicity endpoints (Section 4.2) — rather than remaining focused on

the classical abiotic-stress physiology that the search strategy targeted. It is also worth noting what the co-citation evidence does not show: a clearly separated physiology-only and toxicology-only intellectual architecture within the retrieved corpus. Instead, both draw on the same bioinformatics substrate and the same model organism, which is consistent with the convergence narrative advanced in Section 4.2 rather than with an alternative interpretation in which physiological-stress and toxicological-stress genomics simply happen to be retrieved by the same search string while remaining intellectually distinct.

4.4 Concentration of Productivity and Geography

The slightly super-classical Lotka exponent ($\beta = 2.18$) and the Bradford core of 13 journals both indicate a moderately concentrated field, numerically dominated by a cohort of prolific, largely China-affiliated authors and institutions. Several factors may contribute to this pattern, including differences in research capacity, aquaculture research intensity, institutional concentration, and investment; however, these explanatory factors were not directly evaluated in the present analysis. The institutional network's fragmentation into largely nationally bounded clusters, with only thin bridging ties to Africa, South America, and much of South Asia, indicates persistent geographic underrepresentation in international participation. This concentration is consistent across independent indicators – country-affiliation frequency, the strength of the China–USA collaboration tie, author productivity, and institutional centrality all point in the same direction – which suggests that the observed geographic concentration is unlikely to be attributable to a single metric alone, although the underlying drivers of this pattern were not directly evaluated. From a policy perspective, the thinness of ties to Africa, South America, and South Asia is notable precisely because these regions include aquatic systems (tropical freshwater and coastal fisheries, for example) that differ ecologically from the temperate and subtropical systems most studied in the current literature, meaning that geographic concentration in authorship may also have implications for the ecological representativeness of the systems studied.

4.5 Limitations

Several limitations qualify these results. First, reliance on a single database (Scopus) may underrepresent regionally indexed journals, and the English-language restriction may undercount non-English aquaculture and stress-physiology literature despite the corpus's heavy Chinese authorship. Second, author-name disambiguation by surname and initials can both fragment single authors into multiple network nodes and conflate distinct authors sharing common names; collaboration and productivity statistics should be read as indicative rather than definitive at the individual level. Third, the logistic life-cycle and diffusion extrapolations are model-dependent projections derived from a relatively short 21-year series and carry substantial uncertainty. Fourth, the 2022/2023 thematic-evolution cut-point was chosen at the visible inflection in production; alternative cut-points could redistribute some thematic flows. Finally, keyword co-occurrence and thematic mapping are sensitive to inconsistent keyword usage and synonymy (e.g., "zebrafish" vs. "*Danio rerio*"), and the minimum-occurrence threshold used for network construction (Section 2.2) necessarily excludes rarer terms that might otherwise reveal smaller or more nascent sub-themes. A further limitation specific to the central finding of this study (Section 4.2) is that the stress–toxicology convergence is inferred from corpus-level keyword and journal

composition rather than from full-text screening of individual articles; while the pattern is consistent across the Bradford, keyword-cluster, and trend-topic analyses independently, a manual content analysis of a representative document sample would be needed to confirm the proportion of retrieved articles that are substantively contaminant-exposure studies versus classical abiotic-stress physiology.

4.6 Future Directions

Priority directions include multi-omic integration (transcriptomic, proteomic, metabolomic, epigenomic) to strengthen mechanistic inference beyond transcript-level evidence alone; standardized reporting practices – including consistent author-keyword vocabularies and explicit statements of stressor type (abiotic versus contaminant) – to enable more reliable cross-study meta-analysis and to reduce the synonymy problems noted in Section 4.5; expansion beyond zebrafish to regionally relevant fish species, particularly in under-represented tropical and coastal systems; deeper mechanistic investigation of the microplastic–mitochondrial dysfunction–ferroptosis axis identified as the current research front, ideally paired with the manual content-analysis validation described above; and deliberate strengthening of international collaboration involving under-represented regions, potentially through collaboration and funding mechanisms that connect high-output institutions with research groups in Africa, South America, and South Asia.

5. Conclusion

This bibliometric analysis of 1,126 Scopus-indexed documents (2005–2025) maps aquatic stress transcriptomics and genomics as an expanding field in which the retrieved literature shows substantial thematic convergence between physiological-stress research and oxidative toxicology. Built on an imported RNA-seq/bioinformatics toolkit and organized increasingly around zebrafish-centred oxidative-stress biology, the bibliometric evidence indicates an increasingly visible conceptual convergence among physiological stress, transcriptomic profiling, oxidative stress, toxicity, and mitochondrial dysfunction/ferroptosis with ecotoxicology and toxicogenomics emerging as an important component of this trajectory rather than its starting point. China's prominent position in the corpus and a still-fragmented international collaboration network characterize the field's current geography. As mechanistic environmental-health evidence becomes increasingly central to both physiology and toxicology, this synthesis offers a map of where the field has come from and the dominant thematic directions evident in the recent literature.

Declarations

Funding

No funding was received for conducting this study.

Competing interests

The authors declare that they have no competing interests relevant to the content of this article.

Ethics approval

Not applicable. This study analyzed bibliographic records and did not involve human participants, human data, or experimental animals.

Data availability

The bibliographic records analyzed in this study were retrieved from Scopus on 12 September 2026. The complete search strategy, inclusion criteria, analytical parameters, and derived results are provided in the manuscript and Online Resource 1. The original Scopus export cannot be publicly redistributed because access to Scopus data is subject to the database's licensing and access conditions. Derived datasets, bibliometric analysis outputs, and analysis scripts generated during this study are available from the corresponding author on reasonable request.

Author contributions

Guduru Prasanth Kumar: Conceptualization, methodology, data curation, formal analysis, software, visualization, investigation, writing – original draft, and writing – review and editing. Gutha Rajasekhar: Supervision, validation, critical review and editing, and guidance of the final manuscript. Both authors approved the final version of the manuscript and agree to be accountable for all aspects of the work.

AI-assisted tools

An AI-based language tool was used during manuscript preparation for language refinement, drafting assistance and structural editing. The authors reviewed and verified the manuscript, including its scientific content, analyses, interpretations, and final wording, and remain responsible for the content of the article.

References

1. Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975. <https://doi.org/10.1016/j.joi.2017.08.007>
2. Bradford, S. C. (1934). Sources of information on specific subjects. *Engineering*, 137, 85–86.
3. Cole, M., Lindeque, P., Halsband, C., & Galloway, T. S. (2011). Microplastics as contaminants in the marine environment: A review. *Marine Pollution Bulletin*, 62(12), 2588–2597. <https://doi.org/10.1016/j.marpolbul.2011.09.025>
4. Dixon, S. J., Lemberg, K. M., Lamprecht, M. R., Skouta, R., Zaitsev, E. M., Gleason, C. E., Patel, D. N., Bauer, A. J., Cantley, A. M., Yang, W. S., Morrison, B., & Stockwell, B. R. (2012). Ferroptosis: An iron-dependent form of nonapoptotic cell death. *Cell*, 149(5), 1060–1072. <https://doi.org/10.1016/j.cell.2012.03.042>
5. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21.

<https://doi.org/10.1093/bioinformatics/bts635>

6. Hill, A. J., Teraoka, H., Heideman, W., & Peterson, R. E. (2005). Zebrafish as a model vertebrate for investigating chemical toxicity. *Toxicological Sciences*, 86(1), 6–19.
<https://doi.org/10.1093/toxsci/kfi110>
7. Horton, A. A., Walton, A., Spurgeon, D. J., Lahive, E., & Svendsen, C. (2017). Microplastics in freshwater and terrestrial environments: Evaluating the current understanding. *Science of the Total Environment*, 586, 127–141. <https://doi.org/10.1016/j.scitotenv.2017.01.190>
8. Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., Collins, J. E., Humphray, S., McLaren, K., Matthews, L., et al. (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, 496(7446), 498–503.
<https://doi.org/10.1038/nature12111>
9. Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, 12(4), 357–360. <https://doi.org/10.1038/nmeth.3317>
10. Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B., & Schilling, T. F. (1995). Stages of embryonic development of the zebrafish. *Developmental Dynamics*, 203(3), 253–310.
<https://doi.org/10.1002/aja.1002030302>
11. Kültz, D. (2005). Molecular and evolutionary basis of the cellular stress response. *Annual Review of Physiology*, 67, 225–257. <https://doi.org/10.1146/annurev.physiol.67.040403.144654>
12. Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
13. Lieschke, G. J., & Currie, P. D. (2007). Animal models of human disease: Zebrafish swim into view. *Nature Reviews Genetics*, 8(5), 353–367. <https://doi.org/10.1038/nrg2091>
14. Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2 – $\Delta\Delta$ CT method. *Methods*, 25(4), 402–408.
<https://doi.org/10.1006/meth.2001.1262>
15. Long, Y., Li, L., Li, Q., He, X., & Cui, Z. (2012). Transcriptomic characterization of temperature stress responses in larval zebrafish. *PLoS ONE*, 7(5), e37209.
<https://doi.org/10.1371/journal.pone.0037209>
16. Long, Y., Song, G., Yan, J., He, X., Li, Q., & Cui, Z. (2013). Transcriptomic characterization of cold acclimation in larval zebrafish. *BMC Genomics*, 14, 612. <https://doi.org/10.1186/1471-2164-14-612>
17. Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16(12), 317–323.
18. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>
19. Pauli, A., Valen, E., Lin, M. F., Garber, M., Vastenhouw, N. L., Long, H. K., McWhinnie, E., Sadelain, M., Schier, A. F., & Rinn, J. L. (2012). Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Research*, 22(3), 577–591.
<https://doi.org/10.1101/gr.133009.111>

20. Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139–140. <https://doi.org/10.1093/bioinformatics/btp616>
21. van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538. <https://doi.org/10.1007/s11192-009-0146-3>
22. Westerfield, M. (2000). *The Zebrafish Book: A Guide for the Laboratory Use of Zebrafish (Danio rerio)* (4th ed.). University of Oregon Press.

Figures

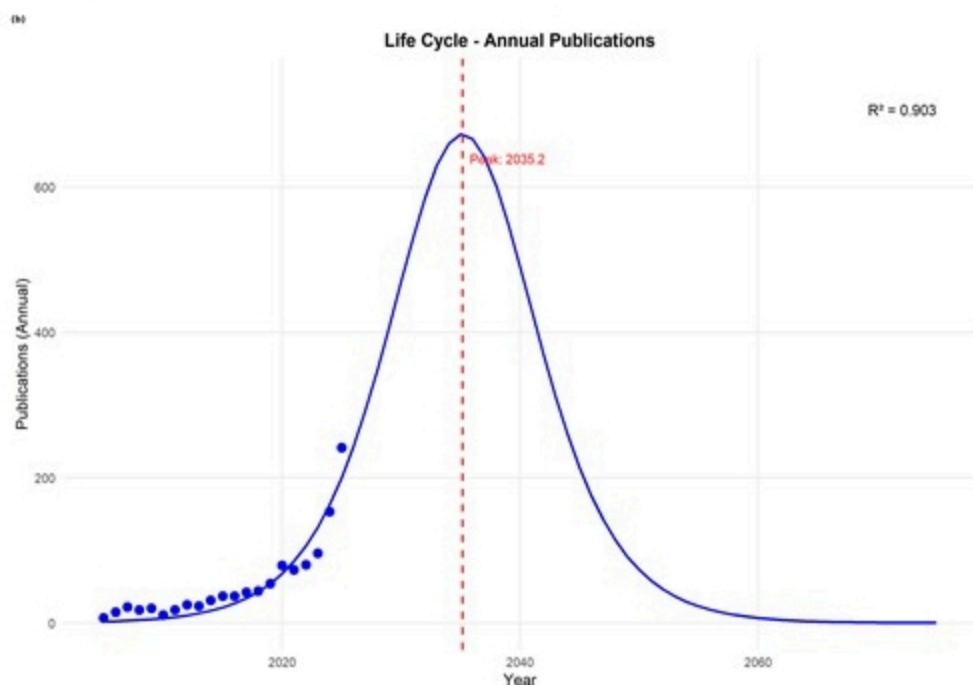
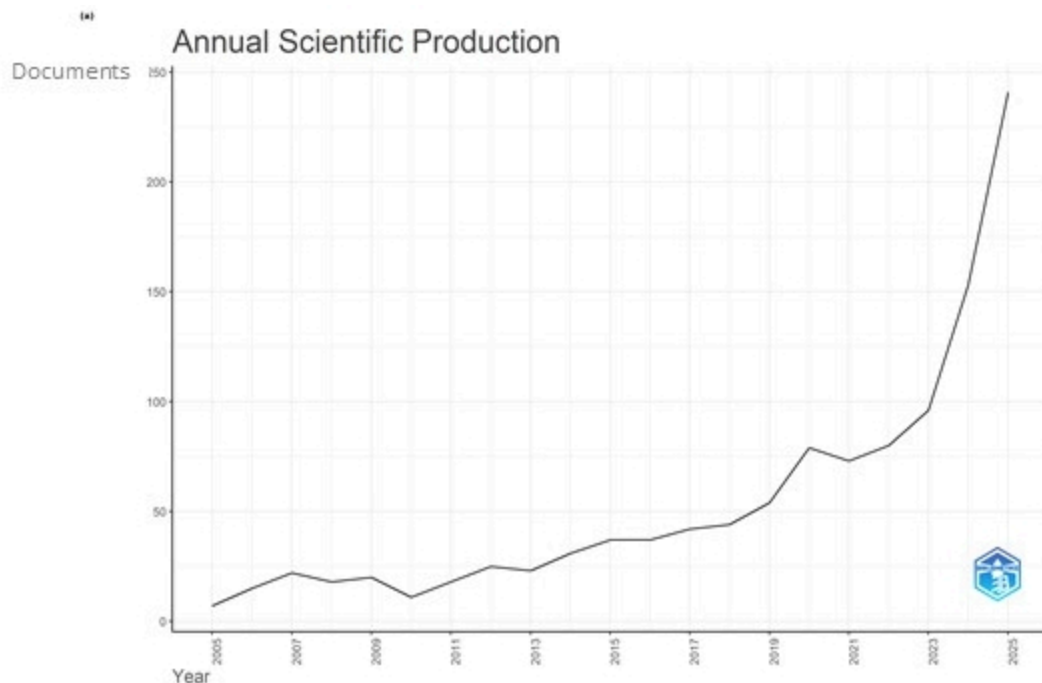


Figure 1

Growth dynamics of the literature. (a) Annual scientific production, 2005–2025 (n = 1,126 documents). (b) Logistic life-cycle model fitted to annual counts, with a model-projected peak year of 2035.2 ($R^2 = 0.903$); this is a statistical extrapolation and should be read as a directional trend signal rather than a precise forecast (see Section 4.1)

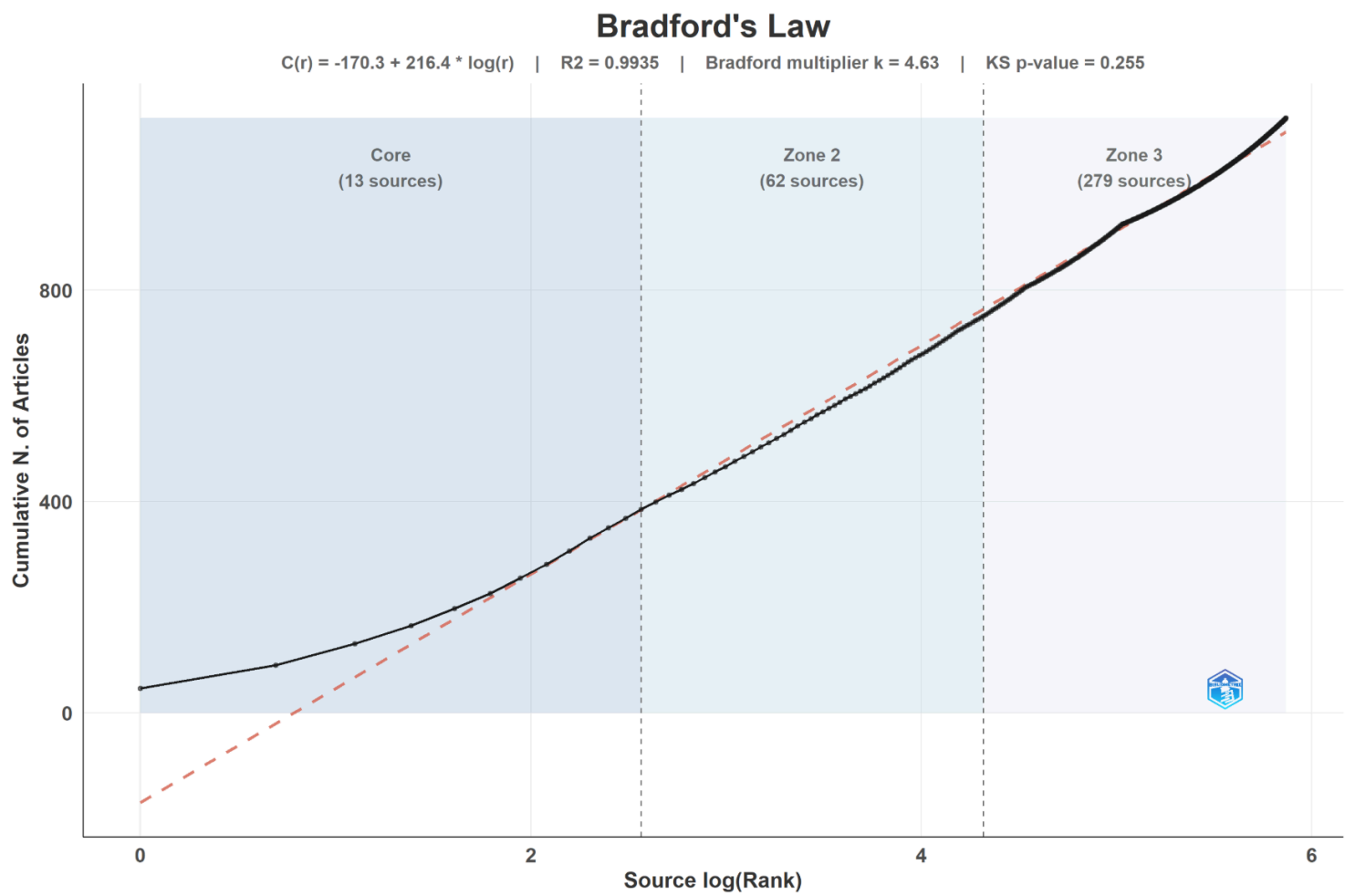


Figure 2

Bradford's Law of Scattering applied to the source distribution of the corpus (Core = 13; Zone 2 = 62; Zone 3 = 279 sources; $k = 4.63$; $R^2 = 0.9935$)

Lotka's Law

Estimated: Beta = 2.18, C = 0.3069 | R² = 0.9129 | KS p-value (Beta=2): 0.00141 | KS p-value (fitted): 0.0794

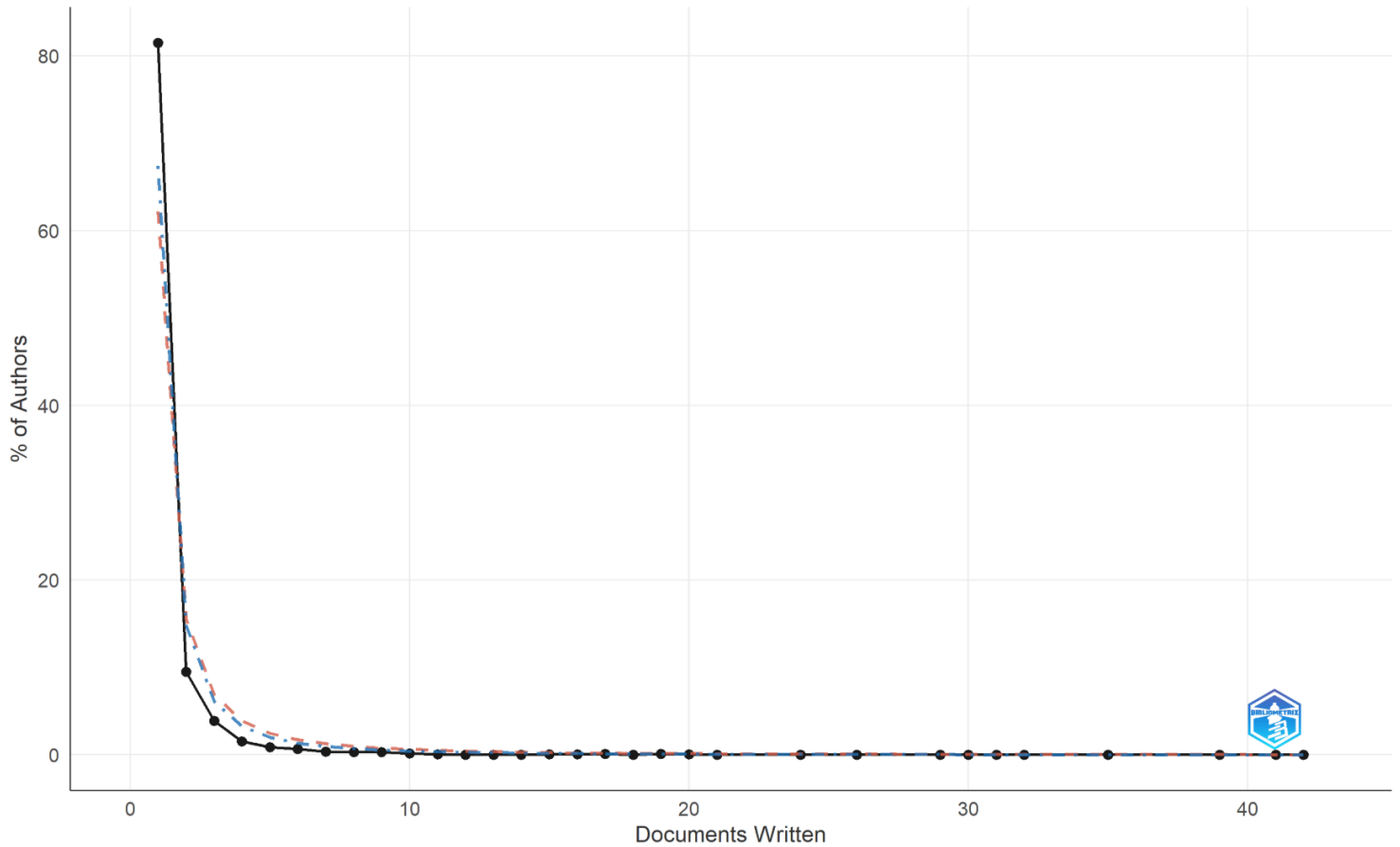


Figure 3

Lotka's Law applied to author productivity (fitted $\beta = 2.18$, $C = 0.3069$, $R^2 = 0.9129$)

(a) Country Scientific Production

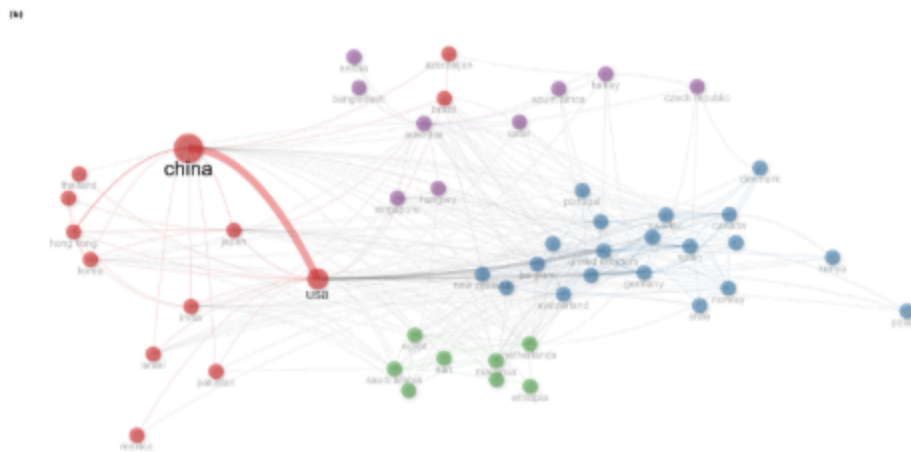
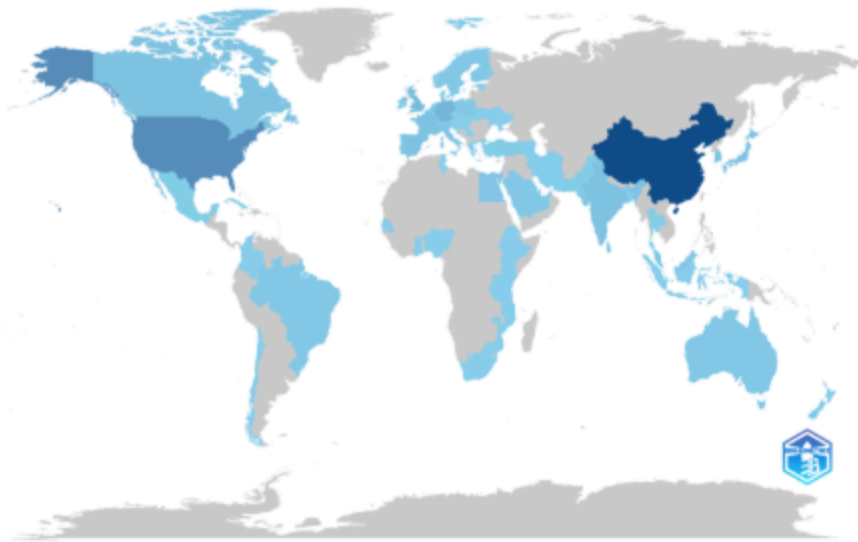


Figure 4

Geographic distribution and international collaboration. (a) Choropleth map of country scientific production (country-affiliation frequency; full counting). (b) International country collaboration network (node size = country-affiliation frequency; edge width = collaboration frequency)

Page 22/23

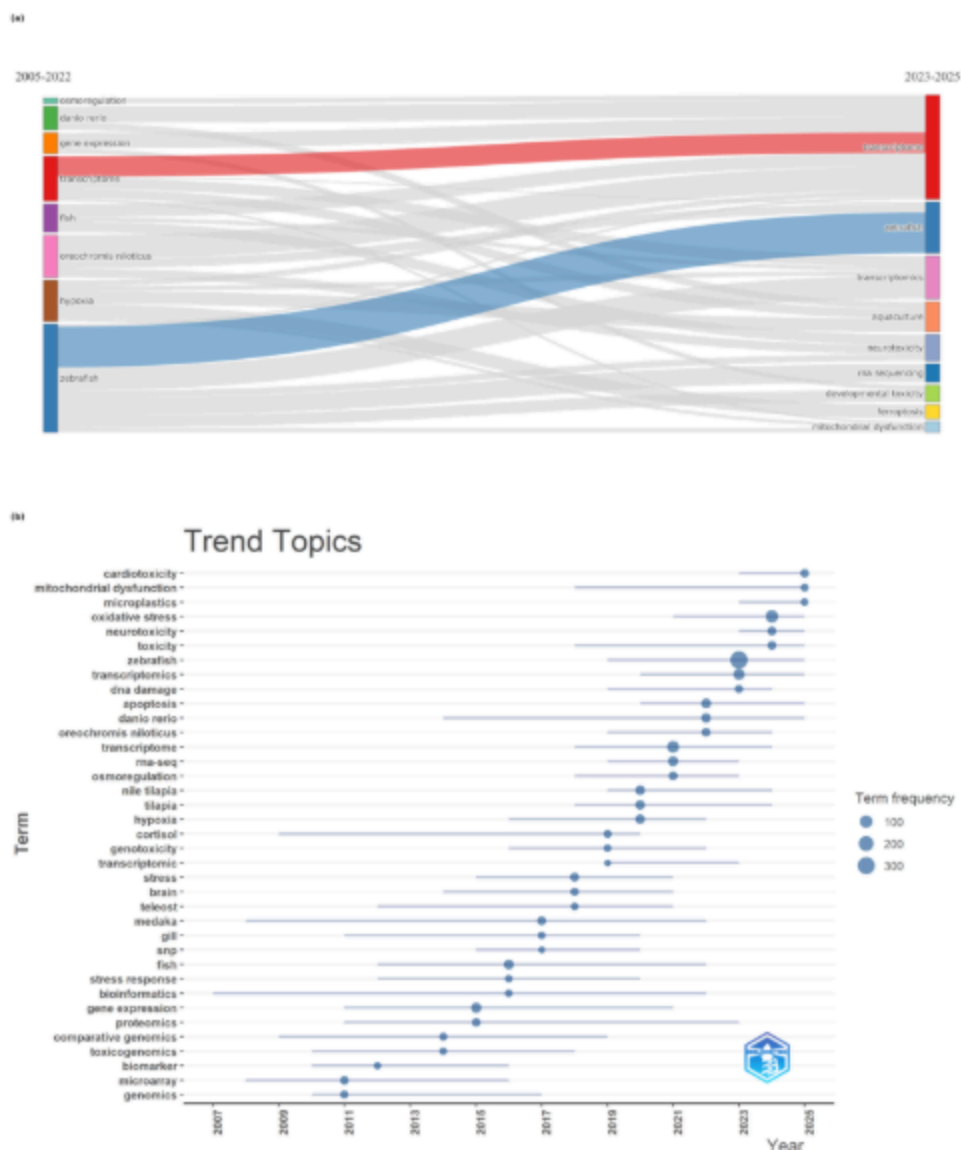


Figure 6

Temporal evolution of research themes. (a) Thematic evolution of author keywords between two periods (2005–2022 and 2023–2025); flows represent conceptual keyword associations, not biological causal pathways. (b) Trend-topics chart showing the chronological emergence, peak, and decline of key research terms (bubble size = term frequency)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ESM1.pdf](#)