 https://doi.org/10.24191/jeesr.v29i1.022

Mapping the Scholarly Landscape of DNA Sequence Alignment: Trends, Themes, and Future Directions

'Iffah Athirah Zaidi, Syed Abdul Mutalib Al-Junid, Abdul Karimi Halim and Abdul Hadi Abdul Razak*

***Abstract*—DNA sequence alignment remains a cornerstone of genomics, underpinning the identification of functional, structural, and evolutionary relationships across biological sequences. However, the rapid expansion of sequencing technologies and methodological diversification has fragmented scholarly output, making it challenging to capture the field's intellectual evolution comprehensively. This study addresses this gap through a bibliometric analysis of DNA sequence alignment research published between 1968 and 2025. A total of 1,848 unique documents were retrieved from Scopus and Web of Science and analyzed using ScientoPy, VOSviewer, and Biblioshiny. Performance indicators assessed publication trends, document types, prolific journals, and country-level contributions, while science mapping and thematic evolution analyses identified conceptual structures and emerging trajectories. Results indicate exponential growth beginning in the mid-2000s, coinciding with the rise of next-generation and long-read sequencing technologies. Four dominant thematic clusters were identified: multiple sequence alignment and phylogenetics, accelerated read mapping, microbial rRNA sequencing, and short-read error correction. Core journals such as Nucleic Acids Research and Bioinformatics remain central publication venues, while geographic trends highlight sustained leadership by the United States alongside growing contributions from China, India, and Australia. Thematic mapping further reveals emerging research fronts in nanopore sequencing, bioinformatics integration, and AI-driven analytical frameworks. This study consolidates the intellectual foundations and thematic evolution of DNA sequence alignment research, offering a holistic view of its historical development and future trajectory. Future investigations should incorporate alternative databases and altmetric indicators to capture interdisciplinary convergence and societal impact within the evolving genomics landscape.**

***Index Terms*—bibliometric analysis, Biblioshiny, bioinformatics, DNA sequence alignment, long-read sequencing, ScientoPy, VOSviewer.**

This manuscript is submitted on August 21, 2026, accepted on 8 October, 2026 and published on 31 October, 2026.

'Iffah Athirah Zaidi, Syed Abdul Mutalib Al-Junid, Abdul Karimi Halim, Abdul Hadi Abdul Razak are from the Electronics Architecture and Applications (EArA), School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.

*Corresponding author
Email address: hadi@ieee.org

1985-5389/© 2026 The Authors. Published by UiTM Press. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

I. INTRODUCTION

In the era of precision biology and big data, DNA sequence alignment remains a cornerstone technique in genomics, bioinformatics, and evolutionary biology. It enables the detection of regions of similarity that reveal functional, structural, or evolutionary relationships among biological sequences [1]. From genome annotation and phylogenetic analysis to the identification of disease-causing mutations, sequence alignment is pivotal to unravelling the information encoded within DNA [2]. With the proliferation of next-generation sequencing (NGS) technologies, the demand for accurate, efficient, and scalable alignment methods has become increasingly urgent [3]–[5].

The field has progressed through successive technological waves. Early dynamic programming algorithms such as Needleman-Wunsch [6] and Smith-Waterman [7] provided rigorous yet computationally intensive solutions for pairwise alignment. The advent of NGS in the 2000s introduced an unprecedented scale of short-read data, which spurred the development of heuristic and probabilistic methods exemplified by BWA [1], BLAST [8], and Bowtie [9]. More recently, third-generation sequencing platforms such as PacBio and Oxford Nanopore have produced long reads capable of resolving structural complexity, but they demand novel computational approaches for error correction, structural variant detection, and pan-genomic analysis [10]–[12].

This rapid evolution has fragmented DNA sequence alignment research into specialized subdomains, from metagenomic profiling and microbial community analysis [13] to transcript isoform detection [14] and real-time epigenomic monitoring [15]. While this diversification has driven innovation, it has also generated a heterogeneous knowledge base that is increasingly difficult to synthesize holistically. Evaluating scholarly output, collaboration networks, and thematic development is therefore essential to capture the field's intellectual structure, identify emerging research fronts, and guide future innovation [16], [17].

A. Need for Bibliometric Analysis

Although numerous narrative reviews have synthesized particular aspects of sequence alignment, few studies have offered a holistic, quantitative view of its broader scientific evolution. Traditional literature reviews are constrained by subjective scope and manual selection, which risk overlooking cross-disciplinary linkages. In contrast, bibliometric analysis

provides an objective and reproducible methodology to evaluate scholarly output, track collaboration networks, and uncover thematic structures across large datasets [16], [18], [19]. By mapping citation relationships, co-authorship patterns, and keyword trajectories, bibliometric techniques can systematically reveal both established research clusters and nascent topics.

Such an approach is especially valuable in DNA sequence alignment, where rapid technological advances have produced fragmented and overlapping subfields. Bibliometric mapping enables the detection of hidden intellectual connections, research silos, and conceptual shifts that may remain invisible in conventional reviews [20]–[22]. Furthermore, by integrating data from both Scopus and Web of Science, this study ensures broader coverage and mitigates database bias, which strengthens the reliability of the results [23], [24]. The combined use of ScientoPy for performance indicators and trend analysis, alongside VOSviewer for network visualization, offers a comprehensive analytical framework capable of capturing both the quantitative growth and the conceptual evolution of the field [25].

B. Objectives

The objectives of this study are threefold:

1) To characterize the scholarly landscape of DNA sequence alignment research.

Map the scholarly landscape of DNA sequence alignment research by analyzing publication trends, document types, leading journals, citation performance, and global geographic distribution of research output.

2) To identify the intellectual and thematic structure of DNA sequence alignment research.

Identify the intellectual and thematic structure of DNA sequence alignment research through bibliographic coupling and keyword co-occurrence analyses, identifying major research domains and knowledge communities.

3) To determine emerging research trends and future directions in DNA sequence alignment research.

Determine the emerging research fronts and future trajectories in DNA sequence alignment research by highlighting methodological innovations, interdisciplinary linkages, and strategic directions for researchers, tool developers, and funding agencies.

II. METHODOLOGY

This study adopts a quantitative, descriptive, and exploratory bibliometric approach to map the intellectual landscape of DNA sequence alignment research [16], [19]. The methodology integrates performance indicators with science-mapping techniques to capture both the field's productivity and its conceptual structure. The section is structured into two parts: Bibliometric Analysis and Procedural Analysis.

A. Bibliometric Analysis

Bibliometric indicators were applied to evaluate the productivity, impact, and intellectual structure of the field. Following established bibliometric practices [16], [19], the analysis was conducted from two complementary perspectives:

1) Productivity and Impact Indicators:

Annual scientific production, document type distributions, source title contributions, country-level outputs, and citation counts were analyzed. These metrics provide a macro-level view of research activity, its global diffusion, and its influence across disciplines.

2) Knowledge and Conceptual Structures:

Two science-mapping techniques were employed. Bibliographic coupling was used to detect thematic clusters by linking documents that share common references, while keyword co-occurrence was used to trace conceptual evolution and emerging research fronts by analyzing the frequency and co-appearance of keywords [17].

3) Thematic Evolution and Strategic Mapping:

To identify the maturity, centrality, and developmental stage of research themes, thematic mapping was performed [18]. This approach classified research topics into four quadrants, which are Motor Themes, Basic Themes, Niche Themes, and Emerging or Declining Themes, based on Callon's centrality and density measures, providing additional insights into the field's emerging research fronts and future trajectories.

Together, these perspectives offer both a quantitative and structural understanding of DNA sequence alignment research.

B. Procedural Analysis

This section describes the step-by-step procedures undertaken to ensure methodological transparency and analytical reproducibility. It covers database selection, search strategy development, criteria for document inclusion, and metadata extraction. It also outlines the cleaning and standardization of records prior to analysis and the use of specialized bibliometric software, including ScientoPy for performance analysis [25] and VOSviewer for network visualization and cluster identification [23], [26], and Biblioshiny [18] for thematic evolution and strategic mapping of emerging research fronts.

Fig. 1 presents the systematic workflow of data collection, screening, and preparation, from the initial database queries to the final consolidated dataset used for analysis.

C. Data Source and Search Strategy

The data were collected from Scopus (Elsevier) and the Web of Science Core Collection (Clarivate Analytics), two of the most widely recognized and reputable scientific databases. These sources were selected due to their broad coverage of multidisciplinary, peer-reviewed literature and their established reliability in citation analysis [23], [24]. Scopus provides extensive indexing across applied and theoretical disciplines, while Web of Science emphasizes high-impact journals and long-standing citation data. Using both databases reduces coverage bias and ensures a more comprehensive representation of the research landscape [27], [28].

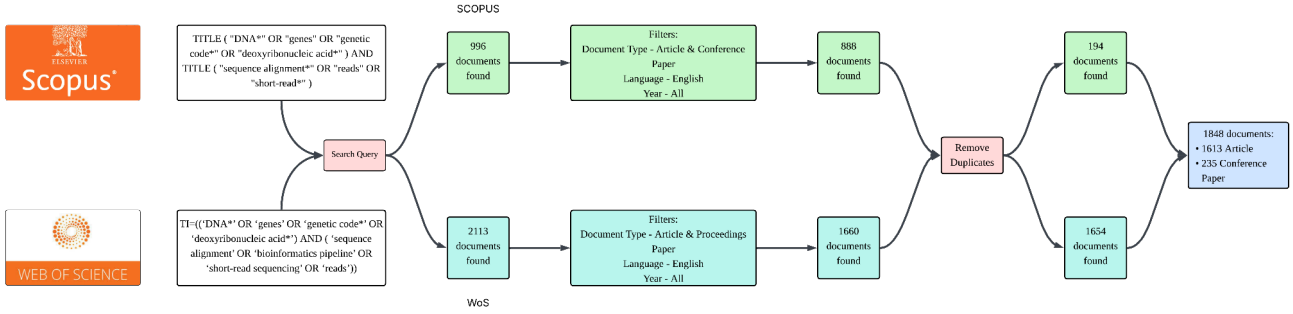


Fig. 1. Search strategy and document selection process.

Boolean operators (AND, OR, NOT) were applied to craft the search query, combining DNA-related terms with those describing alignment methodologies and sequencing reads. This strategy ensured that the retrieved publications accurately reflected the study's core themes [16], [18].

As shown in Fig. 1, the initial search, performed in September 2025, yielded 996 documents from Scopus and 2,113 from WoS. To ensure consistency and relevance, the dataset was refined using the following inclusion criteria: (i) English-language publications, (ii) document types limited to journal articles and conference (or proceeding) papers, and (iii) no restriction on publication year to enable longitudinal analysis. After applying these filters, the datasets were reduced to 888 records in Scopus and 1,660 in WoS, which were deemed eligible for bibliometric analysis. Metadata, including titles, abstracts, authors, affiliations, keywords, publication year, and citation counts, were exported in CSV format for analysis. Following deduplication, a consolidated dataset of 1,848 unique documents was obtained, comprising 1,613 journal articles and 235 conference papers. This curated dataset formed the basis for the subsequent bibliometric analysis, providing both historical and longitudinal insights into the evolution of DNA sequence alignment research.

D. Software Tools and Functions

This study employed three widely recognized bibliometric tools, ScientoPy, VOSviewer and Biblioshiny to conduct a comprehensive analysis of the dataset.

ScientoPy [25] was used primarily for performance analysis, offering visualizations and detailed metrics related to publication outputs, citation patterns, and topic evolution. Its capabilities include trend timelines, author productivity graphs, and keyword growth trajectories, enabling a clearer understanding of the domain's temporal development and concentration of research efforts. This made ScientoPy especially useful for identifying core journals, prolific contributors, and emerging themes in DNA sequence alignment [16].

VOSviewer [26] was employed for science mapping, with a focus on uncovering the intellectual and conceptual structures of DNA sequence alignment research. Two types of maps were generated:

1) Bibliographic coupling maps.

Bibliographic coupling maps, which identify thematic clusters by linking documents that share common references [17].

2) Keyword co-occurrence maps.

Keyword co-occurrence maps, which reveal conceptual linkages and thematic evolution by analyzing the frequency and co-appearance of keywords [16], [21].

VOSviewer applies association strength normalization and modularity-based clustering, enabling the detection of distinct research communities and conceptual structures. Items positioned closer together on the map represent stronger bibliographic or keyword ties, while color coding indicates cluster membership. These maps provided the foundation for the thematic interpretations reported in the Results section.

Biblioshiny, a web-based interface of the Bibliometrix R package [18], was utilized to perform thematic evolution and strategic mapping. This tool enabled the identification of research maturity, centrality, and developmental stages through Callon's centrality and density measures, supporting the analysis of emerging research fronts and future trajectories.

Together, ScientoPy, VOSviewer, and Biblioshiny formed a cohesive analytical workflow: ScientoPy handled data preprocessing and performance analysis, VOSviewer visualized relational structures, and Biblioshiny complemented these with strategic mapping of thematic evolution. This integration enabled both macro-level trend analysis and micro-level conceptual exploration, consistent with the widely applied three-level bibliometric framework of performance analysis, science mapping, and network analysis [16], [18].

III. Results

Prior to analysis, the retrieved records from Scopus and Web of Science were merged and cleaned to remove duplicates. Fig. 2 shows the preprocessing stage. Out of the 888 Scopus documents and 1,660 WoS documents initially retained after filtering, a large proportion of Scopus entries (78%) were identified as duplicates and removed, leaving unique contributions. In contrast, the overlap in WoS records was minimal (0%), confirming its strong representation of unique documents.

This cleaning step ensured that the final dataset of 1,848 non-redundant documents was both representative and reliable, preventing bias from duplicate counting across databases. By consolidating sources in this manner, the analysis captures the most accurate scholarly landscape of DNA sequence alignment research.

A. Publication Trends

Fig. 3 shows the annual evolution of DNA sequence alignment publications indexed in Scopus and Web of Science from 1968 to 2025.

The earliest contributions appeared in the late 1960s, coinciding with foundational work on pairwise alignment algorithms such as Needleman-Wunsch [6] and Smith-Waterman [7]. A modest but steady increase is visible through the 1980s and 1990s, corresponding to the rise of heuristic approaches and early bioinformatics applications.

A marked growth phase begins in the mid-2000s, reflecting the widespread adoption of next-generation sequencing (NGS) technologies and the corresponding demand for scalable alignment tools. Publications indexed in WoS show a sharp upward trend from 2008 onwards, peaking above 100 documents per year in 2024. Scopus demonstrates a lower but consistent contribution, indicating partial database coverage rather than a lack of research activity.

Overall, the trajectory highlights both the maturation and expansion of DNA sequence alignment research, with the past decade showing accelerated growth that aligns with the advent of long-read sequencing, pan-genomics, and real-time analysis pipelines. This confirms the field's transition from foundational algorithmic development to a diversified and application-driven domain.

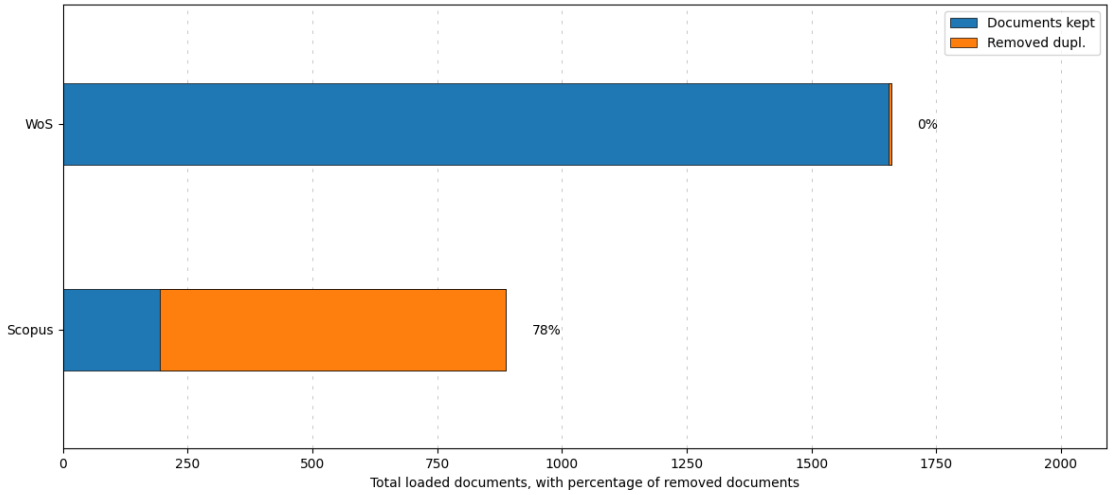


Fig. 2. Total loaded documents and percentage of removed duplicates from Scopus and WoS databases (Source: ScientoPy).

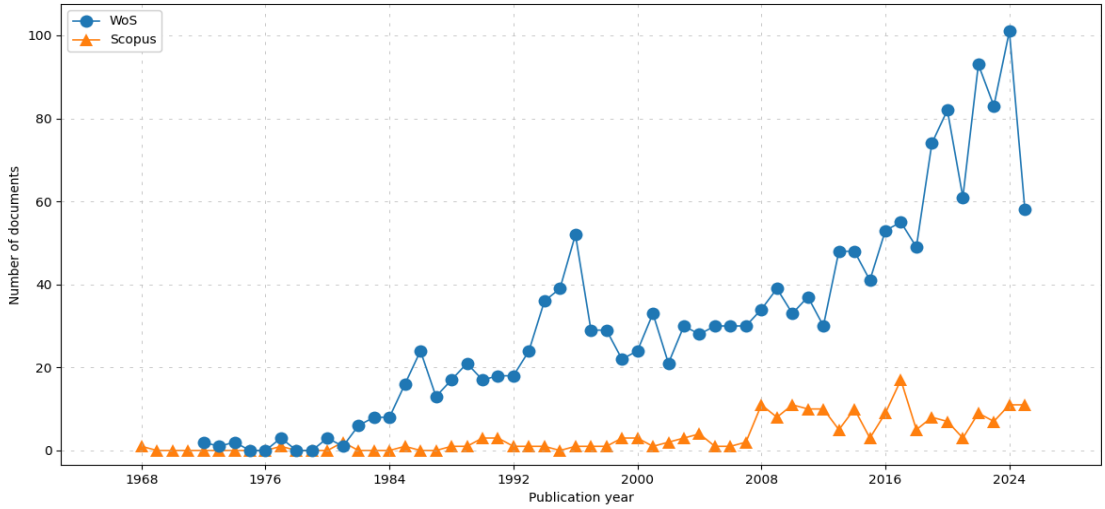


Fig. 3. Annual scientific production in Scopus and WoS, 1968-2025 (Source: ScientoPy).

To further understand the publication landscape, the dataset was analyzed by document type. As shown in Fig. 4, journal articles consistently represent the majority of contributions to DNA sequence alignment research. Their output has steadily increased since the 1980s, reaching a peak of more than 100 publications in 2024.

Conference papers, by contrast, began to appear in the early 1990s and gained momentum in the mid-2000s, coinciding with the emergence of next-generation sequencing and computational bioinformatics pipelines. Their share, however, has remained relatively modest compared to journal articles, reflecting the field's strong emphasis on peer-reviewed journal dissemination rather than conference proceedings.

Overall, this distribution highlights the predominantly journal-driven nature of DNA sequence alignment research, with conference papers contributing more during phases of

rapid methodological innovation but not maintaining the same long-term trajectory as journal articles.

The distribution of publications across journals provides insight into the disciplinary orientation and preferred venues of DNA sequence alignment research. As illustrated in Fig. 5, Nucleic Acids Research is the most prolific outlet, contributing over 85 documents. This is consistent with its longstanding role as a leading journal in molecular biology and bioinformatics tool development.

Other highly represented journals include Yeast and the Journal of Virology, reflecting the application of sequence alignment techniques to microbial genomics and viral research. Specialized computational outlets such as Bioinformatics and BMC Bioinformatics also feature prominently, underscoring the methodological dimension of the field. Broadly scoped journals such as PLOS ONE and Scientific Reports appear as additional contributors, highlighting the cross-disciplinary relevance of alignment studies.

Overall, the concentration of publications in bioinformatics and molecular biology journals demonstrates the central role of sequence alignment as both a computational and experimental enabler across diverse biological domains.

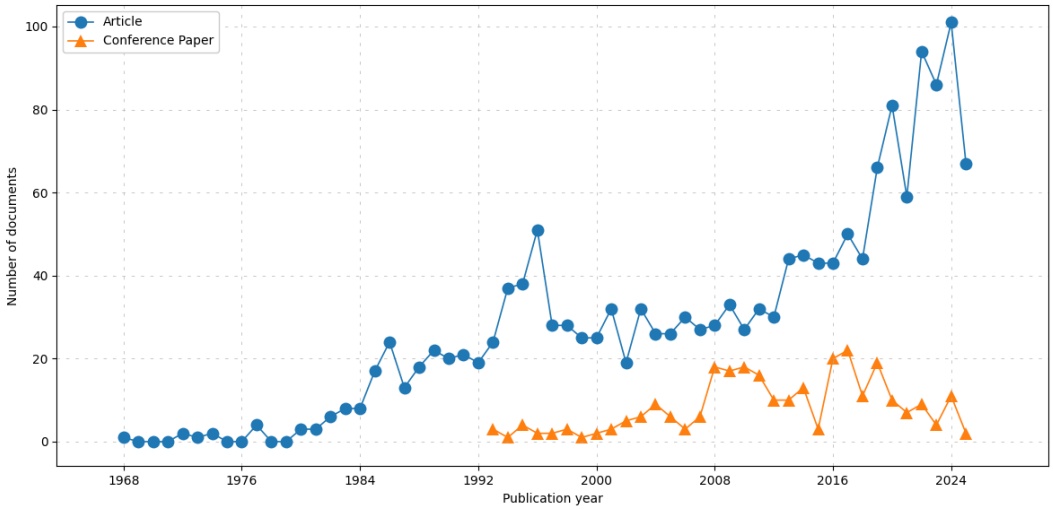


Fig. 4. Annual publication trends of articles and conference papers related to DNA sequence alignment (Source: ScientoPy).

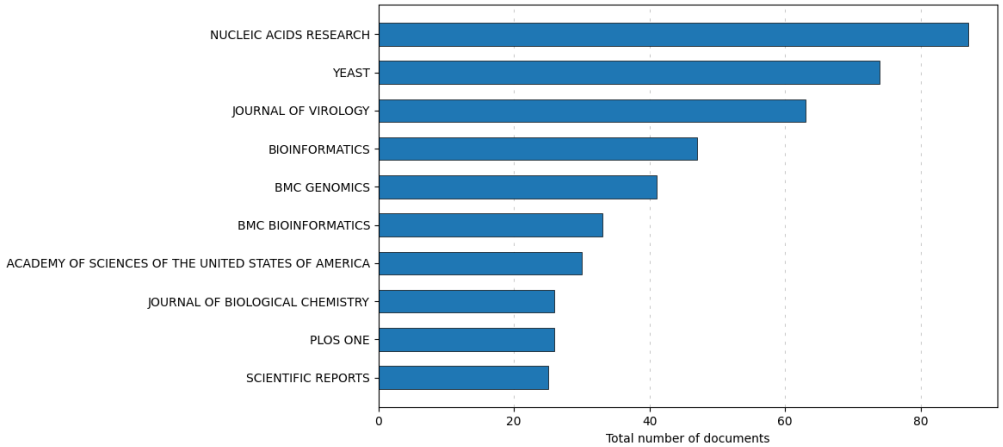


Fig. 5. Trend of publications distribution across various journals and conference proceedings of DNA sequence alignment research (Source: ScientoPy).

The geographic distribution of DNA sequence alignment research is shown in Fig. 6. The United States is the most prolific contributor, accounting for the highest number of publications and maintaining consistent dominance across the study period. China ranks second, with a notable 20% of its publications appearing in the most recent two years (2024-2025), highlighting its growing momentum in the field.

European countries, including the United Kingdom, Germany, France, and Italy, demonstrate steady contributions, reflecting strong research communities in both algorithmic development and applied genomics. Japan and Canada also maintain a visible presence, while India and Australia show rising activity, with 15% of their publications emerging in 2024-2025, suggesting increased engagement with advanced sequencing and bioinformatics applications.

This distribution illustrates the globalized yet regionally concentrated nature of DNA sequence alignment research. Established leaders such as the United States provide foundational contributions, while rapidly emerging nations such as China, India, and Australia signal a diversification of research hubs and a potential rebalancing of global influence in the near future.

B. Key Research Areas

Keyword analysis offers insight into the conceptual orientation of DNA sequence alignment research. Fig. 7 displays the most frequently occurring author keywords in the dataset, distinguishing between historical use (before 2024) and more recent activity (2024-2025).

Unsurprisingly, "DNA" dominates the field, appearing in more than 200 documents. Terms such as "expression," "alignment," "identification," "protein," and "genome" also feature prominently, reflecting the methodological and biological breadth of sequence alignment applications. More technical descriptors such as "bioinformatics," "sequence," and "evolution" underscore the interdisciplinary integration of computational and biological sciences.

Of particular interest, "bioinformatics" shows a 12% concentration of publications in the past two years, suggesting increasing prominence as sequence alignment becomes more deeply embedded in computational pipelines. Similarly, keywords related to "genome" and "sequence alignment" continue to expand, reflecting the influence of long-read sequencing and pan-genomics.

Overall, the keyword dynamics indicate that while core concepts remain stable, the field is increasingly oriented toward integrative bioinformatics approaches and genome-wide applications, with signs of diversification into emerging computational subfields.

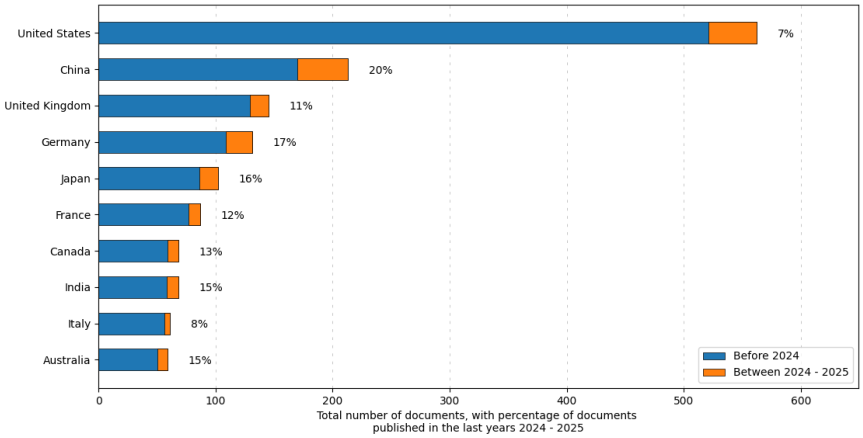


Fig. 6. Trend of publications distribution of DNA sequence alignment research across countries (Source: ScientoPy).

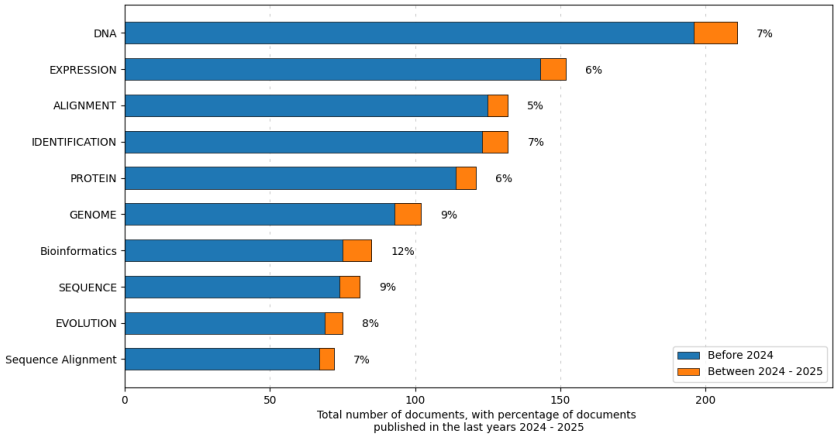


Fig. 7. Annual publication trends of DNA sequence alignment research on keyword analysis (Source: ScientoPy).

C. Citation Analysis

To identify the intellectual foundations of the field, the most cited documents in the dataset were analyzed. Table I lists the top-cited works, which include landmark algorithmic contributions and widely adopted tools.

For example, [9] introduced Bowtie, a memory-efficient aligner optimized for short-read data, which has been cited more than 19,000 times. Similarly, [29] presented SINA, a high-throughput multiple sequence alignment tool tailored for ribosomal RNA gene sequences, while [30] developed a method for mapping short reads and calling variants, central to the NGS era.

These highly cited works highlight the dual emphasis in DNA sequence alignment research:

1. Methodological innovation (e.g., Bowtie, MAFFT, BWA).
2. Applied frameworks for microbial, genomic, and evolutionary analysis.

Collectively, they illustrate the field's enduring reliance on computational efficiency and accuracy, while also reflecting the methodological shifts accompanying each generation of sequencing technology.

D. Bibliographic Coupling Analysis

Bibliographic coupling was performed to identify thematic

clusters within the DNA sequence alignment literature. Using a minimum citation threshold of 105, 60 documents met the inclusion criteria. For each, VOSviewer calculated the total strength of bibliographic coupling links, grouping the documents into four clusters as shown in Fig. 8. These clusters represent documents that share strong reference overlaps, indicating proximity in research focus.

TABLE I. MOST CITED ARTICLES RELATED TO DNA SEQUENCE ALIGNMENT RESEARCH

Documents	Title	Citation	Insight
[9]	Ultrafast and memory-efficient alignment of short DNA sequences to the human genome	19807	Bowtie is a memory-efficient alignment program for short DNA sequences to the human genome, achieving over 25 million reads per hour with a 1.3 GB memory footprint.
[29]	SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes	2549	Pruesse et al. (2012) introduce SINA, a high-throughput tool for accurate ribosomal RNA gene sequence alignment, outperforming existing methods like PyNAST and mothur in various benchmarks.
[30]	Mapping short DNA sequencing reads and calling variants using mapping quality scores	2347	Li, Ruan, and Durbin (2008) introduce MAQ software for mapping short DNA sequencing reads to a reference genome, improving variant calling accuracy, including SNPs and short indels.
[31]	Alternative Reading Frames Of The Ink4a Tumor-Suppressor Gene Encode 2 Unrelated Proteins Capable of Inducing Cell-Cycle Arrest	1438	The INK4a tumor suppressor gene encodes two proteins, p16^INK4a and p19^ARF, which link Rb and p53 tumor-suppressor pathways within a single genetic locus.
[32]	Multiple Alignment of DNA Sequences with MAFFT	1415	The paper discusses MAFFT's algorithms for DNA sequence alignment, emphasizing scalability and accuracy, and introduces alternative tools like TBA and MAUVE for genome alignment.
[33]	Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays	1135	Drmanac et al.'s 2010 study introduces a cost-effective, precise genome sequencing method using self-assembling DNA nanoarrays, achieving 45-87-fold coverage and identifying millions of sequence variants.
[34]	Identifying DNA and protein patterns with statistically significant alignments of multiple sequences	1111	Hertz and Stormo's 1999 paper explores methods for aligning DNA and protein sequences, focusing on statistical significance through information content scoring, P value estimation, and a greedy algorithm.
[35]	Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase	1046	The paper presents a single-nucleotide DNA sequencing method using a mutant MspA nanopore and phi29 DNA polymerase, overcoming key challenges in nanopore sequencing.
[36]	Substantial biases in ultra-short read data sets from high-throughput DNA sequencing	979	The paper highlights biases in ultra-short read data from high-throughput sequencing, affecting data quality and interpretation in genomic analyses like SNP identification and transcriptome analysis.
[37]	NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes	954	The paper introduces NAST, an online tool that efficiently aligns thousands of 16S rRNA genes, enabling microbiologists to profile bacterial and archaeal communities for comparative analysis.

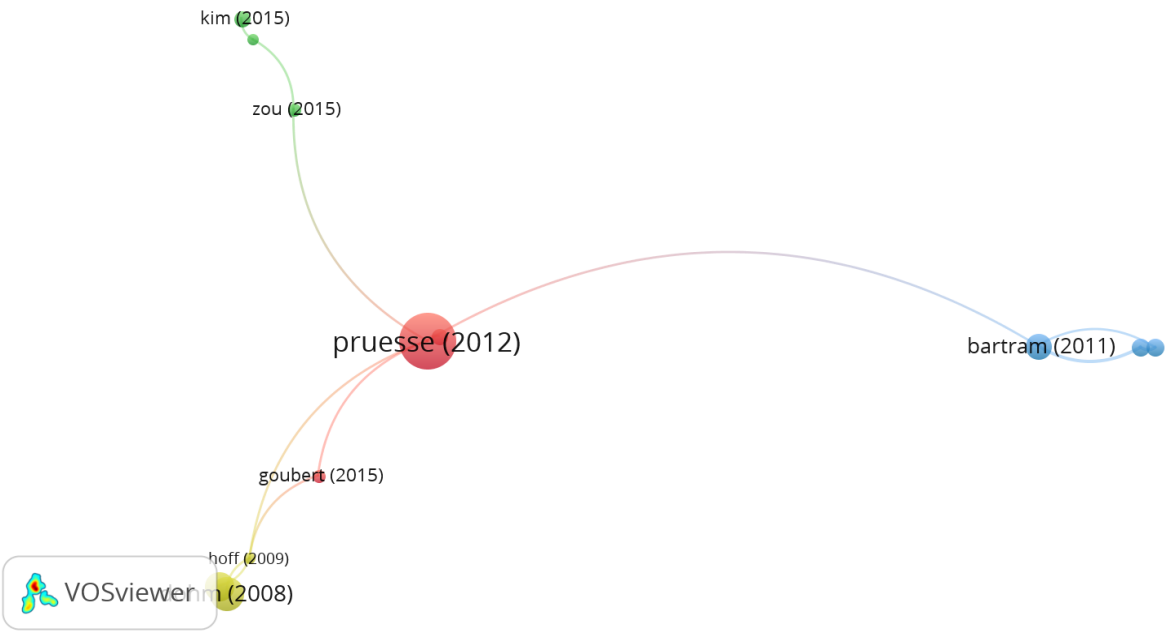


Fig. 8. Network visualization of DNA sequence alignment (Source: VOSviewer).

1) Cluster 1 (Red) - Multiple Sequence Alignment and Phylogenetics

This cluster emphasizes algorithmic and methodological advances in multiple sequence alignment. Representative works include [29], who developed SINA for high-throughput ribosomal RNA alignment, and [38], who assessed automated

filtering methods for phylogenetic inference. These contributions highlight ongoing challenges in balancing alignment accuracy and computational efficiency when dealing with complex genomic data.

2) Cluster 2 (Green) - Read Mapping and Alignment Acceleration

Focused on technological innovations for faster alignment of sequencing data, this cluster includes [39], who pioneered fetal DNA detection through sequence read counts, and [40], who developed HAlign, a fast multiple sequence alignment tool. The inclusion of GRIM-Filter [41] underscores the trend toward hardware-accelerated pipelines for high-throughput DNA mapping.

3) Cluster 3 (Blue) - Microbial Community Analysis and rRNA Sequencing

This cluster centres on methodological frameworks for analyzing microbial diversity using rRNA sequencing. Key works include [42], who generated large 16S rRNA libraries, and [43], who introduced EMIRGE for reconstructing full-length ribosomal genes. Together, these contributions establish the role of sequence alignment in metagenomics and microbial ecology.

4) Cluster 4 (Yellow) - Error Correction and Gene Prediction in Short Reads

This cluster reflects efforts to address limitations of early high-throughput sequencing technologies. [36] revealed systematic biases in short-read data, while [44] and [45] developed Orphelia and FragGeneScan for gene prediction in noisy metagenomic reads. These works illustrate the critical role of error handling in ensuring reliable biological interpretation from next-generation sequencing outputs.

Overall, the bibliographic coupling analysis reveals a field organized around four thematic pillars:

1. Multiple sequence alignment & phylogenetics (Cluster 1).
2. Alignment acceleration & read mapping pipelines (Cluster 2).
3. Microbial community analysis using rRNA sequencing (Cluster 3).
4. Error correction and gene prediction in short reads (Cluster 4).

Together, these clusters capture the methodological and applied diversity of DNA sequence alignment research. Rather than citation prestige alone, bibliographic coupling reveals the intellectual proximity of documents that share common reference foundations, mapping the field's current thematic organization.

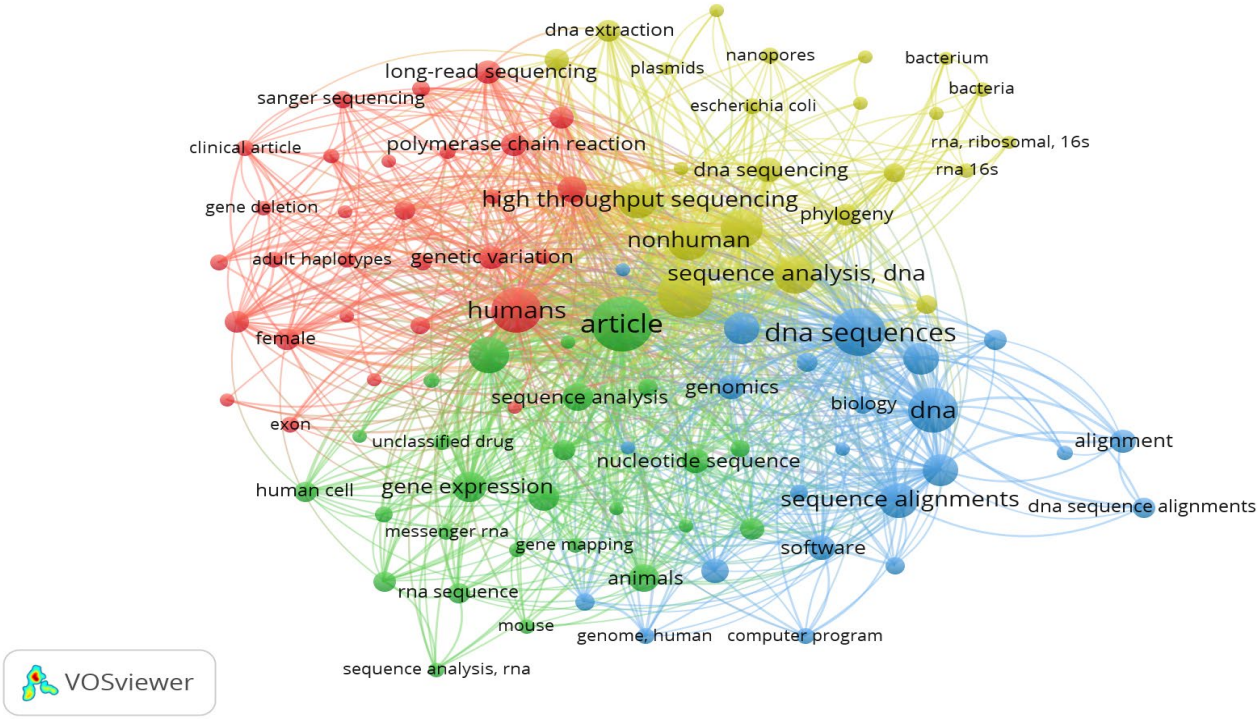


Fig. 9. Co-occurrence analysis on DNA sequence alignment (Source: VOSviewer).

E. Keyword Co-Occurrence Analysis

The keyword co-occurrence network, as shown in Fig. 9, was generated with a minimum threshold of 25 occurrences. Out of 8,012 keywords, 101 met the threshold and were grouped into four clusters, each representing a distinct thematic area in DNA sequence alignment research.

1) Cluster 1 (Red) - Human Genomics and Clinical Applications

Anchored by terms such as humans, long-read sequencing, genetic variation, and gene deletion. This cluster highlights applications of DNA sequence alignment in human genomics, population genetics, and clinical studies (e.g., identifying variants, structural changes, and haplotypes).

2) Cluster 2 (Green) - Functional Genomics and Gene Regulation

Defined by keywords like gene expression, sequence analysis, human cell, and gene mapping. This cluster represents studies focusing on transcriptomics, RNA sequencing, and gene regulation, reflecting the integration of alignment with functional biology and expression profiling.

3) Cluster 3 (Blue) - Computational Tools and Core Alignment Methods

Featuring DNA, DNA sequence alignment, genomics, biology, and software. This cluster is centred on algorithm development, bioinformatics software, and computational strategies for efficient sequence alignment, showing the methodological backbone of the field.

4) Cluster 4 (Yellow) - High-Throughput Sequencing and Microbial Genomics

Characterized by high-throughput sequencing, nanopores, bacteria, and DNA extraction. This cluster points to technological advances in sequencing platforms (e.g., nanopore, next-generation sequencing) and their applications in microbial community studies and metagenomics.

Together, these clusters reveal a balanced landscape: from clinical genomics (Cluster 1), to functional biology (Cluster 2), to computational innovation (Cluster 3), and emerging sequencing technologies (Cluster 4). This structure underscores how DNA sequence alignment serves as a methodological bridge linking human health, computational biology, and microbial genomics.

Table II presents the most frequently occurring keywords in DNA sequence alignment research, along with their total link strength (TLS) values. Occurrences represent how many times a keyword appeared across the dataset, while TLS reflects the extent of its co-occurrence with other terms, indicating its connectivity within the research network.

F. Emerging Research Fronts and Future Trajectories

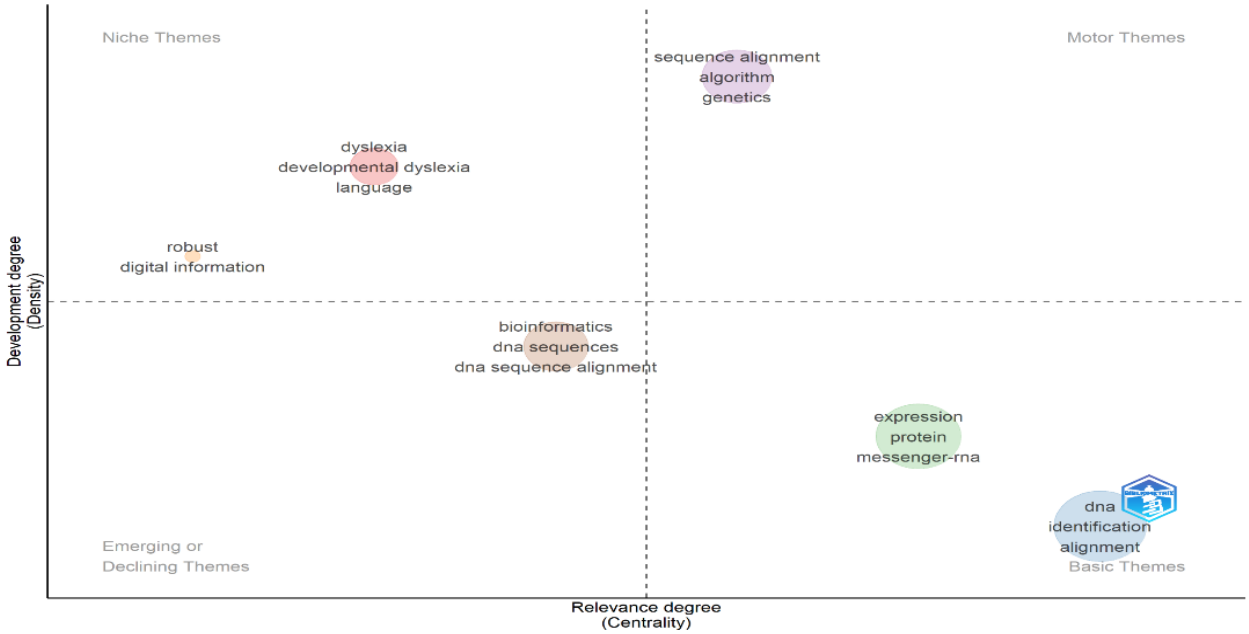
The thematic evolution analysis using Biblioshiny was conducted to identify the conceptual maturity and developmental potential of research themes within the DNA sequence alignment domain. Fig. 10 presents the thematic map, where clusters are classified along two dimensions: centrality, which indicates their importance and connection to other themes, and density, which represents their internal cohesion and maturity. The map is divided into four quadrants representing Motor Themes, Basic Themes, Niche Themes, and Emerging or Declining Themes, each signifying distinct levels of thematic significance and evolution.

Motor Themes are both highly developed and essential to the research domain. The cluster positioned in this quadrant contains sequence alignment, algorithms, and genetics. These keywords form the methodological backbone of the field, highlighting areas where computational precision and biological insight converge. The cluster centred on sequence alignment shows a high Callon density (19.26) and moderate Callon centrality (1.638), indicating that it is both well-developed and conceptually robust. The strong centrality and density confirm that this theme remains a driving force in the field, continually refined through algorithmic innovation to handle increasingly complex genomic data. Overall, this quadrant represents the active engine of DNA sequence alignment research, which is mature, relevant, and continually evolving.

Basic Themes are well-connected to the field but less internally developed, serving as the conceptual foundation for broader research activities. Two clusters appear in this quadrant. The first includes DNA, identification, and alignment, representing the fundamental aspects of sequence comparison and genetic identification. The second cluster includes expression, protein, and messenger RNA, reflecting applications of alignment in functional genomics and transcriptomics. The DNA cluster, with the highest Callon centrality (4.073) and cluster frequency (2627), represents the core conceptual foundation of the field. Its extensive connectivity reflects the central role of DNA analysis in genomics, bioinformatics, and molecular biology.

TABLE II . MOST TOP 15 KEYWORDS ON DNA SEQUENCE ALIGNMENT

Ranks	Keyword	Occurrences	Total Link Strength (TLS)
1	genetics	346	1245
2	dna sequences	337	1126
3	dna	301	905
4	humans	299	1071
5	nonhuman	205	788
6	sequence analysis, dna	204	923
7	high-throughput sequencing	199	847
8	sequence alignments	185	588
9	bioinformatics	153	432
10	gene expression	133	445
11	animals	110	455
12	sequence analysis	108	454
13	gene sequence	98	405
14	dna sequencing	95	415
15	genome	91	365



Cluster	CallonCentrality	CallonDensity	RankCentrality	RankDensity	ClusterFrequency
dyslexia	0.214	18.571	2	5	133
dna	4.073	12.764	6	1	2627
expression	1.853	13.421	5	2	1560
sequence alignment	1.638	19.26	4	6	571
robust	0	16.964	1	4	15
bioinformatics	0.695	15.534	3	3	396

Fig. 10. The thematic map (Source: Biblioshiny).

The expression cluster, with a Callon centrality of 1.853 and density of 13.421 (cluster frequency 1560), captures research on gene regulation and transcriptomic profiling, highlighting its biological significance. Together, these clusters indicate that the basic scientific principles and biological contexts of alignment continue to underpin ongoing studies and methodological advancements. These themes represent well-established research bases that support cross-disciplinary integration between molecular biology and computational science.

Niche Themes represent specialized, highly developed topics with limited connectivity to the core of the field. Two distinct clusters were identified in this quadrant. The first cluster, comprising dyslexia, developmental dyslexia, and language, reflects an interdisciplinary application of sequence alignment principles beyond genomics, particularly in computational linguistics and cognitive studies. Centred on dyslexia (Callon centrality = 0.214; Callon density = 18.571), this cluster illustrates intersections between genetics, neurodevelopment, and language processing. The second cluster, containing robust and digital information, emphasizes algorithmic robustness and digital information management. This cluster, centred on robust (Callon centrality = 0; Callon density = 16.964), signifies focused computational research on improving reliability and accuracy in alignment algorithms. Although peripheral, these clusters contribute meaningful methodological diversity within the broader domain.

Emerging or Declining Themes occupy a transitional position in the field's evolution. The cluster located in this quadrant includes bioinformatics, DNA sequences, and DNA sequence alignment. Centred on bioinformatics, this theme exhibits moderate Callon density (15.534) and lower Callon centrality (0.695). Despite its limited centrality, the cluster's growing frequency (396) reflects increasing research attention. Although its position might suggest a declining trend, this cluster more likely represents a transitional phase from classical sequence alignment studies toward integrated bioinformatics systems. These themes point to future trajectories where alignment methods are embedded within broader analytical frameworks, incorporating real-time data processing, machine learning, and long-read sequencing technologies. Thus, this quadrant signifies transformation rather than decline, marking the progression toward next-generation computational genomics.

Overall, the thematic structure presented in Fig. 10 captures both the stability and dynamism of DNA sequence alignment research. Core domains such as DNA analysis, gene expression, and sequence alignment continue to anchor the research ecosystem, while emerging areas like bioinformatics highlight a shift toward automation, data integration, and interdisciplinary collaboration. Collectively, these patterns portray the future trajectory of DNA sequence alignment as increasingly intelligent, scalable, and deeply integrated with broader bioinformatics and genomic technologies.

IV. Discussion

The bibliometric analysis reveals a research domain characterized by rapid growth, thematic diversification, and expanding international collaboration. Publication trends indicate a modest beginning in the 1970s, followed by a steep acceleration in the mid-2000s, coinciding with the widespread adoption of next-generation sequencing (NGS). This trajectory mirrors the transformative impact of NGS in genomics, where affordability, speed, and scalability revolutionized biological discovery [3], [4]. The more recent proliferation of long-read sequencing platforms, such as PacBio and Oxford Nanopore, has further broadened the scope of DNA sequence alignment, enabling structural variant detection, pan-genomic studies, and clinical diagnostics [11], [12]. These technological shifts are reflected in the diversification of thematic clusters and emerging research fronts identified in this study, demonstrating how alignment research continually adapts to evolving sequencing innovations.

The ScientoPy analysis emphasizes the predominance of journal articles over conference proceedings, reinforcing the importance of archival publishing in advancing methodological and applied genomics research. Highly productive journals such as Nucleic Acids Research and Bioinformatics highlight the dual nature of the field, where computational algorithms and biological applications evolve in parallel. Meanwhile, Biblioshiny's thematic mapping confirms this diversification by distinguishing between foundational and emerging areas of research, from traditional alignment algorithms to integrative bioinformatics frameworks. Country-level analysis also reaffirms the leadership of the United States, while noting the accelerating contributions of China, India, and Australia, aligning with broader trends in global research capacity and scientific collaboration [46], [47]. Together, these patterns underscore the dynamic and increasingly international character of DNA sequence alignment research.

A. Theoretical Contributions

This study enhances theoretical understanding of the intellectual and conceptual landscape of DNA sequence alignment by integrating insights from bibliographic coupling, keyword co-occurrence, and thematic mapping. The bibliographic coupling analysis identified four dominant clusters, which are multiple sequence alignment and phylogenetics, accelerated read mapping, microbial rRNA sequencing, and short-read error correction that collectively define the methodological and applied pillars of the field. These themes align with prior literature emphasizing the dual progression of algorithmic innovation and biological application [1], [2]. By situating these clusters within a unified conceptual framework, this study demonstrates the coherence of a field that has diversified across specialized domains yet remains intellectually interconnected through shared computational foundations.

The keyword co-occurrence and thematic mapping analyses further reveal the field's conceptual evolution. Traditional terms such as DNA, alignment, and protein continue to anchor the field, while emergent keywords such as nanopores, long-read sequencing, and clinical genomics indicate a shift toward translational and integrative bioinformatics applications [13], [15]. Thematic mapping reinforces this trajectory by positioning sequence alignment and algorithm development as motor themes which highly developed and central to the domain, while identifying bioinformatics as an emerging theme, signaling the transition toward automated, large-scale genomic analysis. This dynamic interplay between technological innovation and conceptual diversification exemplifies the evolution of a scientific field as an adaptive network, where methodological advances continually reshape intellectual structures [19].

B. Practical Implications

The findings offer valuable practical insights for researchers, developers, and policymakers. For researchers, the identification of core journals and influential thematic clusters provides strategic direction for publication and collaboration. Emerging areas such as nanopore sequencing, metagenomics, and AI-assisted alignment algorithms represent promising frontiers for future exploration [48], [49]. The growing prominence of bioinformatics as a bridging keyword underscores the interdisciplinary nature of the field, suggesting the increasing importance of collaborative efforts across biology, computer science, and engineering to address challenges in scalability, data integration, and real-time analysis [50], [51].

From a policy and strategic standpoint, the geographic distribution of research output points to an ongoing rebalancing of global expertise. While the United States continues to lead in publication volume and citation impact, nations such as China, India, and Australia have rapidly expanded their contributions, reflecting broader transformations in the global research ecosystem [23]. These shifts suggest that DNA sequence alignment research is becoming more distributed and collaborative, with emerging economies playing vital roles in computational innovation and bioinformatics tool development. For funding agencies and policymakers, these results emphasize the value of fostering international partnerships, infrastructure investment, and capacity-building programs to sustain the momentum of global scientific advancement in genomic technologies.

C. Limitations and Future Directions

Despite offering a comprehensive perspective, this study has several limitations. The analysis was restricted to Scopus and Web of Science, which, while widely recognized as reliable bibliometric databases, do not fully capture all scholarly outputs, particularly those indexed in PubMed, Google Scholar, or regional databases [24], [27]. Furthermore, the focus on English-language publications may underrepresent research from non-English-speaking regions, and the reliance on author-supplied metadata introduces variability in keyword accuracy, completeness, and standardization.

Tool-specific constraints should also be acknowledged. ScientoPy is powerful for descriptive performance analysis but lacks advanced relational mapping functions, whereas

VOSviewer is sensitive to parameter settings such as counting methods and clustering thresholds [22]. Similarly, while Biblioshiny provides an accessible interface for thematic evolution and strategic mapping, its outputs depend on Callon's centrality and density measures, which may vary depending on keyword frequency thresholds. These tool-related factors could influence the granularity of thematic interpretation.

Future research could integrate complementary platforms such as CiteSpace [20] to enable longitudinal and dynamic mapping of conceptual evolution. Combining bibliometric analysis with systematic reviews or machine learning-based text mining could further enrich the qualitative interpretation of trends. Expanding coverage to include preprints, grey literature, and altmetric indicators (e.g., social media and policy citations) would also provide a more holistic and real-time view of this fast-evolving domain. Such extensions would strengthen understanding of how DNA sequence alignment research continues to adapt within the broader landscape of genomics and computational biology.

V. Conclusion

This bibliometric study provides a comprehensive overview of the intellectual and thematic evolution of DNA sequence alignment research from 1968 to 2025. By integrating descriptive performance indicators with science-mapping and thematic-evolution techniques, the study illustrates how the field has advanced from foundational pairwise alignment algorithms to a highly diversified domain encompassing multiple sequence alignment, accelerated read mapping, microbial metagenomics, and clinical genomics. The combined use of ScientoPy, VOSviewer, and Biblioshiny enabled both macro-level insights into research productivity, collaboration, and global distribution, and micro-level mapping of thematic clusters, conceptual linkages, and emerging research trajectories.

The findings highlight the dual nature of sequence alignment as both a computational framework and a biological enabler. Core publication outlets such as Nucleic Acids Research and Bioinformatics remain central, while countries including the United States, China, and the United Kingdom continue to lead in global output. The increasing contributions from India, Australia, and other emerging research hubs reflect the internationalization and diversification of the field. Thematic mapping further revealed dynamic transitions toward novel sequencing platforms, nanopore-based technologies, and AI-assisted analytical frameworks, signifying the ongoing convergence of computational bioinformatics and applied genomics.

Looking ahead, future bibliometric inquiries could extend this work by incorporating alternative databases such as Dimensions or Google Scholar and by integrating altmetric indicators that capture societal and policy influence. Longitudinal tracking of thematic transitions and interdisciplinary collaborations would deepen understanding of how DNA sequence alignment continues to shape and adapt within the broader landscape of genomics, precision medicine, and systems biology. Collectively, this study not only consolidates the intellectual foundations of sequence alignment but also charts its future trajectory in the era of big data and integrative biology.

Acknowledgement

The authors would like to express their deepest appreciation to the Ministry of Higher Education Malaysia (MOHE) through the Fundamental Research Grant Scheme (FRGS/1/2023/TK07/UITM/02/1) for funding this research. This research is also supported by Universiti Teknologi MARA (UiTM).

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of the paper.

Author Contribution

The authors confirm their contribution to the paper as follows: ***Writing–Original Draft:*** *'IAZ;* ***Conceptualization:*** *AHAR;* ***Formal Analysis:*** *'IAZ, AHAR and SAMAJ;* ***Investigation:*** *'IAZ and AKH;* ***Writing–Review & Editing:*** *AHAR, SAMAJ and AKH. All authors reviewed and approved the final version of the manuscript.*

References

[1] H. Li and R. Durbin, "Fast and accurate short read alignment with Burrows–Wheeler transform," *Bioinformatics*, vol. 25, no. 14, pp. 1754–1760, 2009, doi: 10.1093/bioinformatics/btp324.

[2] B. Paten, D. Earl, N. Nguyen, M. Diekhans, D. Zerbino, and D. Haussler, "Cactus: Algorithms for genome multiple sequence alignment," *Genome Res.*, vol. 21, no. 9, pp. 1512–1528, 2011, doi: 10.1101/gr.123356.111.

[3] S. Goodwin, J. D. McPherson, and W. R. McCombie, "Coming of age: Ten years of next-generation sequencing technologies," *Nat. Rev. Genet.*, vol. 17, no. 6, pp. 333–351, 2016, doi: 10.1038/nrg.2016.49.

[4] E. R. Mardis, "DNA sequencing technologies: 2006–2016," *Nat. Protoc.*, vol. 12, no. 2, pp. 213–218, 2017, doi: 10.1038/nprot.2016.182.

[5] J. Shendure *et al.*, "DNA sequencing at 40: Past, present and future," *Nature*, vol. 550, no. 7676, pp. 345–353, 2017, doi: 10.1038/nature24286.

[6] S. B. Needleman and C. D. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *J. Mol. Biol.*, vol. 48, no. 3, pp. 443–453, 1970, doi: 10.1016/0022-2836(70)90057-4.

[7] T. F. Smith and M. S. Waterman, "Identification of common molecular subsequences," *J. Mol. Biol.*, vol. 147, no. 1, pp. 195–197, 1981, doi: 10.1016/0022-2836(81)90087-5.

[8] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *J. Mol. Biol.*, vol. 215, no. 3, pp. 403–410, 1990, doi: 10.1016/S0022-2836(05)80360-2.

[9] B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg, "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome," *Genome Biol.*, vol. 10, no. 3, Art. no. R25, 2009, doi: 10.1186/gb-2009-10-3-r25.

[10] S. L. Amarasinghe, S. Su, X. Dong, L. Zappia, M. E. Ritchie, and Q. Gouil, "Opportunities and challenges in long-read sequencing data analysis," *Genome Biol.*, vol. 21, no. 1, Art. no. 30, 2020, doi: 10.1186/s13059-020-1935-5.

[11] M. Jain, H. E. Olsen, B. Paten, and M. Akeson, "The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community," *Genome Biol.*, vol. 17, no. 1, Art. no. 239, 2016, doi: 10.1186/s13059-016-1103-0.

[12] F. J. Sedlazeck, H. Lee, C. A. Darby, and M. C. Schatz, "Piercing the dark matter: Bioinformatics of long-range sequencing and mapping," *Nat. Rev. Genet.*, vol. 19, no. 6, pp. 329–346, 2018, doi: 10.1038/s41576-018-0003-4.

[13] J. Qin *et al.*, "A human gut microbial gene catalogue established by metagenomic sequencing," *Nature*, vol. 464, no. 7285, pp. 59–65, 2010, doi: 10.1038/nature08821.

[14] C. Trapnell *et al.*, "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation," *Nat. Biotechnol.*, vol. 28, no. 5, pp. 511–515, 2010, doi: 10.1038/nbt.1621.

[15] J. T. Simpson, R. E. Workman, P. C. Zuzarte, M. David, L. J. Dursi, and W. Timp, "Detecting DNA cytosine methylation using nanopore sequencing," *Nat. Methods*, vol. 14, no. 4, pp. 407–410, 2017, doi: 10.1038/nmeth.4184.

[16] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," *J. Bus. Res.*, vol. 133, pp. 285–296, 2021, doi: 10.1016/j.jbusres.2021.04.070.

[17] N. J. van Eck and L. Waltman, "Visualizing bibliometric networks," in *Measuring Scholarly Impact: Methods and Practice*, Y. Ding, R. Rousseau, and D. Wolfram, Eds. Cham, Switzerland: Springer, 2014, pp. 285–320, doi: 10.1007/978-3-319-10377-8_13.

[18] M. Aria and C. Cuccurullo, "bibliometrix: An R-tool for comprehensive science mapping analysis," *J. Informetr.*, vol. 11, no. 4, pp. 959–975, 2017, doi: 10.1016/j.joi.2017.08.007.

[19] I. Zupic and T. Čater, "Bibliometric methods in management and organization," *Organ. Res. Methods*, vol. 18, no. 3, pp. 429–472, 2015, doi: 10.1177/1094428114562629.

[20] C. Chen, Z. Hu, S. Liu, and H. Tseng, "Emerging trends in regenerative medicine: A scientometric analysis in CiteSpace," *Expert Opin. Biol. Ther.*, vol. 12, no. 5, pp. 593–608, 2012, doi: 10.1517/14712598.2012.674507.

[21] M. J. Cobo, A. G. López-Herrera, E. Herrera-Viedma, and F. Herrera, "Science mapping software tools: Review, analysis, and cooperative study among tools," *J. Am. Soc. Inf. Sci. Technol.*, vol. 62, no. 7, pp. 1382–1402, 2011, doi: 10.1002/asi.21525.

[22] J. A. Moral-Muñoz, E. Herrera-Viedma, A. Santisteban-Espejo, and M. J. Cobo, "Software tools for conducting bibliometric analysis in science: An up-to-date review," *Prof. Inf.*, vol. 29, no. 1, 2020, doi: 10.3145/epi.2020.ene.03.

[23] P. Mongeon and A. Paul-Hus, "The journal coverage of Web of Science and Scopus: A comparative analysis," *Scientometrics*, vol. 106, no. 1, pp. 213–228, 2016, doi: 10.1007/s11192-015-1765-5.

[24] R. Pranckutė, "Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world," *Publications*, vol. 9, no. 1, Art. no. 12, 2021, doi: 10.3390/publications9010012.

[25] J. Ruiz-Rosero, G. Ramírez-González, and J. Viveros-Delgado, "Software survey: ScientoPy, a scientometric tool for topics trend analysis in scientific publications," *Scientometrics*, vol. 121, no. 2, pp. 1165–1188, 2019, doi: 10.1007/s11192-019-03213-w.

[26] N. J. van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523–538, 2010, doi: 10.1007/s11192-009-0146-3.

[27] A. Aghaei Chadegani *et al.*, "A comparison between two main academic literature collections: Web of Science and Scopus databases," *Asian Soc. Sci.*, vol. 9, no. 5, pp. 18–26, 2013, doi: 10.5539/ass.v9n5p18.

[28] V. K. Singh, P. Singh, M. Karmakar, J. Leta, and P. Mayr, "The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis," *Scientometrics*, vol. 126, no. 6, pp. 5113–5142, 2021, doi: 10.1007/s11192-021-03948-5.

[29] E. Pruesse, J. Peplies, and F. O. Glöckner, "SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes," *Bioinformatics*, vol. 28, no. 14, pp. 1823–1829, 2012, doi: 10.1093/bioinformatics/bts252.

[30] H. Li, J. Ruan, and R. Durbin, "Mapping short DNA sequencing reads and calling variants using mapping quality scores," *Genome Res.*, vol. 18, no. 11, pp. 1851–1858, 2008, doi: 10.1101/gr.078212.108.

[31] D. E. Quelle, F. Zindy, R. A. Ashmun, and C. J. Sherr, "Alternative reading frames of the INK4a tumor suppressor gene encode two unrelated proteins capable of inducing cell cycle arrest," *Cell*, vol. 83, no. 6, pp. 993–1000, 1995, doi: 10.1016/0092-8674(95)90214-7.

[32] K. Katoh, G. Asimenos, and H. Toh, "Multiple alignment of DNA sequences with MAFFT," in *Bioinformatics for DNA Sequence Analysis*. Totowa, NJ, USA: Humana Press, 2009, pp. 39–64, doi: 10.1007/978-1-59745-251-9_3.

[33] R. Drmanac *et al.*, "Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays," *Science*, vol. 327, no. 5961, pp. 78–81, 2010, doi: 10.1126/science.1181498.

[34] G. Z. Hertz and G. D. Stormo, "Identifying DNA and protein patterns with statistically significant alignments of multiple sequences," *Bioinformatics*, vol. 15, no. 7, pp. 563–577, 1999, doi: 10.1093/bioinformatics/15.7.563.

[35] E. A. Manrao *et al.*, "Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase," *Nat. Biotechnol.*, vol. 30, no. 4, pp. 349–353, 2012, doi: 10.1038/nbt.2171.

[36] J. C. Dohm, C. Lottaz, T. Borodina, and H. Himmelbauer, "Substantial biases in ultra-short read data sets from high-throughput DNA sequencing," *Nucleic Acids Res.*, vol. 36, no. 16, Art. no. e105, 2008, doi: 10.1093/nar/gkn425.

[37] T. Z. DeSantis *et al.*, "NAST: A multiple sequence alignment server for comparative analysis of 16S rRNA genes," *Nucleic Acids Res.*, vol. 34, no. suppl_2, pp. W394–W399, 2006, doi: 10.1093/nar/gkl244.

[38] G. Tan *et al.*, "Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference," *Syst. Biol.*, vol. 64, no. 5, pp. 778–791, 2015, doi: 10.1093/sysbio/syv033.

[39] S. K. Kim *et al.*, "Determination of fetal DNA fraction from the plasma of pregnant women using sequence read counts," *Prenat. Diagn.*, vol. 35, no. 8, pp. 810–815, 2015, doi: 10.1002/pd.4615.

[40] Q. Zou, Q. Hu, M. Guo, and G. Wang, "HAlign: Fast multiple similar DNA/RNA sequence alignment based on the centre star strategy," *Bioinformatics*, vol. 31, no. 15, pp. 2475–2481, 2015, doi: 10.1093/bioinformatics/btv177.

[41] J. S. Kim *et al.*, "GRIM-Filter: Fast seed location filtering in DNA read mapping using processing-in-memory technologies," *BMC Genomics*, vol. 19, no. Suppl. 2, Art. no. 89, 2018, doi: 10.1186/s12864-018-4460-0.

[42] A. K. Bartram, M. D. Lynch, J. C. Stearns, G. Moreno-Hagelsieb, and J. D. Neufeld, "Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads," *Appl. Environ. Microbiol.*, vol. 77, no. 11, pp. 3846–3852, 2011, doi: 10.1128/AEM.02772-10.

[43] C. S. Miller, B. J. Baker, B. C. Thomas, S. W. Singer, and J. F. Banfield, "EMIRGE: Reconstruction of full-length ribosomal genes from microbial community short read sequencing data," *Genome Biol.*, vol. 12, no. 5, Art. no. R44, 2011, doi: 10.1186/gb-2011-12-5-r44.

[44] K. J. Hoff, T. Lingner, P. Meinicke, and M. Tech, "Orphelia: Predicting genes in metagenomic sequencing reads," *Nucleic Acids Res.*, vol. 37, no. suppl_2, pp. W101–W105, 2009, doi: 10.1093/nar/gkp327.

[45] M. Rho, H. Tang, and Y. Ye, "FragGeneScan: Predicting genes in short and error-prone reads," *Nucleic Acids Res.*, vol. 38, no. 20, Art. no. e191, 2010, doi: 10.1093/nar/gkq747.

[46] É. Archambault, D. Campbell, Y. Gingras, and V. Larivière, "Comparing bibliometric statistics obtained from the Web of Science and Scopus," *J. Am. Soc. Inf. Sci. Technol.*, vol. 60, no. 7, pp. 1320–1326, 2009, doi: 10.1002/asi.21062.

[47] J. Wang and P. Shapira, "Is there a relationship between research sponsorship and publication impact? An analysis of funding acknowledgments in nanotechnology papers," *PLoS One*, vol. 10, no. 2, Art. no. e0117727, 2015, doi: 10.1371/journal.pone.0117727.

[48] M. Alser *et al.*, "Technology dictates algorithms: Recent developments in read alignment," *Genome Biol.*, vol. 22, no. 1, Art. no. 249, 2021, doi: 10.1186/s13059-021-02443-7.

[49] M. H. A. Ghafar *et al.*, "Comparative performance of urine and blood-based attributes in chronic kidney disease prediction using artificial neural network," *J. Electr. Electron. Syst. Res.*, vol. 17, pp. 17–28, 2020, doi: 10.24191/jeesr.v17i1.003.

[50] M. H. Abdul Ghafar, S. A. M. Al-Junid, M. S. A. Megat Ali, F. Mohamad, and A. H. Abdul Razak, "Chronic kidney disease diagnostic tools based on machine learning algorithms: A review," *J. Electr. Electron. Syst. Res.*, vol. 26, no. 1, pp. 18–26, 2025, doi: 10.24191/jeesr.v26i1.003.

[51] N. A. Mohd Farid, M. F. Md Idros, A. K. Halim, and A. H. Abdul Razak, "Amplifying 23-years insights of Doherty power amplifier: A review and trend analysis," *J. Electr. Electron. Syst. Res.*, vol. 25, no. 1, pp. 34–45, 2024, doi: 10.24191/jeesr.v25i1.004.