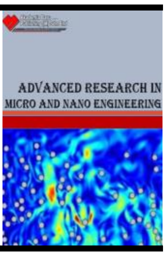
Penerbit
Akademia Baru

Journal of Advanced Research in
Micro and Nano Engineering

Journal homepage:
https://www.akademiabaru.com/submit/index.php/armne/index
ISSN: 2756-8210



A Bibliometric Review on Deoxyribonucleic Acid (DNA) Splicing System

Nooradelena Mohd Ruslim[1], Yuhani Yusof[1,*], Mohd Sham Mohamad[1], Mohammad Hassan Mudaber[2]

[1] Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, 26300 Kuantan, Pahang, Malaysia
[2] Department of Mathematics, Faculty of Natural Sciences, Kabul Education University, 5th District, Afshshar Road, Kabul, Afghanistan

ARTICLE INFO

Article history:
Received 10 January 2024
Received in revised form 14 February 2024
Accepted 19 March 2024
Available online 30 April 2024

Keywords:
DNA; splicing system; bibliometric analysis; DNA recombination

ABSTRACT

This study objects to provide a communicative review for splicing system, focusing on deoxyribonucleic acid (DNA) published in journals and proceedings indexed by the Scopus database. The method of bibliometric analysis is used to describe and analyze the evolution of publication structure and its various co-relationships such as citation, co-citation, co-authorship, and co-occurrence, among the chosen 174 documents indexed by Scopus. This study gives some insightful direction of research where the most influential documents and authors were determined. Besides, the emerging of technologies in most of the research works conducted previously promote the hybridization of background of studies specifically in mathematics, biology, and computer science. The study is limited to English articles which was published by the Scopus database, ranging from the year of 1983 until 2023. The source is also limited to Scopus journals and proceedings only. As far as the authors are aware, this is the first study to address the splicing system topic from a bibliometric perspective. The authors believe that this bibliometric review will help scholars and researchers to stand firm bases regarding the scientific development of this field of study.

1. Introduction

All living organisms are developed and functioned with the influenced of molecules that contain in the genetic instructions called deoxyribonucleic acid (DNA). A long double-stranded helical structure DNA is made up of smaller molecules called nucleotides, which consists of a sugar (deoxyribose), a phosphate group, and one of four nitrogenous bases called adenine (A), guanine (G), cytosine (C) and thymine (T), as explicated by Tamarin [1]. The specific sequence of the nitrogenous bases within the DNA molecule forms the genetic code, which determines the synthesis of proteins and ultimately influences an organism's traits and characteristics. DNA molecules in living organism will experience damage, hence, DNA splicing system is important in substituting the damaged cells. DNA splicing system is a process where DNA molecules from two different sources are combined to form a new genetic sequence. This process plays a crucial role in genetic diversity, evolution, and the repair of damaged DNA.

---

[*] *Corresponding author.*
E-mail address: yuhani@umpsa.edu.my

https://doi.org/10.37934/armne.18.1.123137

There are several mechanisms through which DNA splicing can occur. Li *et al.,* [2] discussed the most common mechanism called homologous recombination, which involves the exchange of genetic material between two similar or identical sequences. This process typically occurs during meiosis, where segments of DNA from each parent are exchanged to generate genetic variation in offspring. Additionally, in 2020, Ensminger and Lobrich [3] deliberated another mechanism of DNA splicing called non-homologous end joining. This process occurs following DNA damage, such as double-stranded breaks, and entails ligating damaged DNA ends directly. It is an error-prone process, often leading to mutations. Furthermore, site-specific recombination is a mechanism through which specific DNA sequences called recombination sites are recognized and exchanged. This category of recombination is involved in DNA arrangements and the control of gene expression. plays a role in the regulation of gene expression and DNA rearrangements. Another process in DNA splicing system also can be seen on the recombinational repair process. Haber [4] in 1999 exposed that this process arises in response to a more prevalent source of such DNA damage.

The DNA splicing process involves the swapping of genetic information among two DNA molecules or different sections in the same DNA molecule. It is mediated by specific enzymes and requires the presence of homologous sequences that act as a template for the recombination event. The steps involved in DNA splicing system include DNA cleavage, DNA strand exchange, DNA branch migration, DNA resolution and DNA ligation. Overall, DNA splicing system plays a crucial role in genetic diversity and evolution by shuffling and combining genetic information from different sources. It is involved in several biological process, including DNA repair, gene expression regulation, and the generation of genetic diversity during meiosis. The study of informational macromolecules and formal language theory, which was started by Head [5] in 1987, has tightened the integration of mathematics and biology, despite the biological process of DNA splicing system, which necessitates financial and temporal commitment.

The data gained from the Scopus database showed a collection of documents that can be clustered into two main area; thus, this paper is meant to see the direction of research in splicing system, either in biology or mathematics perspectives. This paper's primary goal is to give academics and researchers a thorough overview of the DNA splicing system from research works that has been published in proceedings and journals, that the Scopus database has indexed. To accomplish this goal, three research questions are addressed that are; to recognize the most significant publications in DNA splicing system through citation analysis, to explore the relationship among the highly cited publications through co-citation analysis, to investigate the interactions among scholars in DNA splicing system field through co-authorship analysis, and to investigate and foresee future trends in DNA splicing system via co-occurrence analysis. In the next section, a brief introduction and history of splicing system are given.

2. DNA Splicing System and Its History

In the context of theoretical computer science and molecular computing, a splicing system is a formal model, proposed due to the framework of Formal Language Theory. It is a mathematical model used to simulate and explore the operations performed by DNA molecules in biological systems. Head [5] initiated a splicing system namely Head splicing system, which consists of a set of rules or production rules that define how DNA sequences, represented as strings of symbols, can be rearranged and combined to create new sequences. These rules represent the biological processes of cutting and recombining DNA molecules in a laboratory or within living organisms. The model of splicing systems helps researchers understand the behavior and capabilities of DNA computing and its potential applications in solving complex computational problems.

The history of splicing system had begun from the phenomenal work by Head in 1987. The Head splicing system originated by Head was developed based on Formal Language Theory. The Head splicing system makes a connection between molecular biology and mathematics when it represents the set of dsDNA molecules that may arise from the reaction between the initial set of DNA molecules in the enzymatic activities.

Paun splicing system introduced by Paun [6] in 1996, considers infinite cases and a distinctive approach due to its set of rules. This model works theoretically rather than biologically since its rule can be presented by infinite cases. Meanwhile, Pixton introduces the Pixton splicing system that same year. This system is shown as a splicing substitution process. Therefore, it is determined that the Pixton splicing system is an informal description of the biological recombinant. Since the crossing site is the most crucial component of the DNA splicing process and all restriction enzymes have three main sections, and these two splicing system formalisms do not focus on it, hence it does not function biologically [7]. Hence, these systems are not related to the cutting and pasting operations as mentioned in the Head splicing system.

The motivation of this study is the new formalism of splicing system proposed by Yusof and Goode in 2012, where Yusof [8] introduced the new writing notation called Yusof-Goode (Y-G) splicing system with the objective of investigating the behavior of restriction enzyme acting on DNA molecule in a translucent approach.

3. Methodology

The bibliometric analysis is a multidisciplinary method that includes both qualitative and quantitative data from published literature. In Donthu *et al.,* [9], the bibliometric analysis is found to be a suitable method for exploring and analysing large scientific data. The bibliometric analysis process has the advantage of being impartial as it assesses all works within a particular topic, limiting the possibility of selection bias. Bibliometric analysis can offer structured, unbiased and trustworthy analysis of trends over time, research streams, and shifts in the field's scope to present a full picture of the subject. Ellegaard and Wallin [10] suggested that this method is widely used in studying various disciplines of knowledge, for example in the aspects of science, and how the universities and institutions are ranked worldwide. Bibliometric analysis can be conducted in different analysis, with different objectives to be achieved. In this paper, four types of analysis are chosen, namely citation analysis, co-citation analysis, co-authorship analysis and co-occurrence analysis.

Citation analysis is a technique used to evaluate the effect and influence of research works by examining the frequency and patterns of citations within academic and scientific publications. It involves analysing the citations in a given document or collection of documents to gain insights into various aspects such as author’s impact, a journal’s prestige, or the influence of a particular publication. It is quantified that assessing and identifying the major contributors will become more challenging, costly, and important as science becomes a larger and more complicated enterprise and plays a more significant role in society. Garfield [11] in 1979 agreed that, an appropriate citation analysis can provide a valuable level of impartiality to the assessment process at a minimal expense.

Co-citation analysis is a method used in bibliometrics and information science to explore the relationship and connection between different scientific papers based on their co-citation patterns. It involves identifying and analysing the patterns of co-citations, which are instances where two or more scholarly works are collectedly cited by other documents. Park [12] suggested that co-citation occurs when two different publications are referenced in the publication. While Small [13] agreed that co-citation is based on the occurrence of two documents from previous works are collectedly cited in a subsequent work. By examining these co-citation patterns, researchers can gain insight into the intellectual structure, evolution, and relationship within a specific area of study.

Co-authorship analysis is a method that focuses on analyzing the collaborative patterns and dynamics among authors in academic publications. It seeks to understand how authors collaborate with each other, how knowledge is shared and disseminated, and how research networks are formed within a particular discipline of field of study. By examining co-authorship patterns, researchers can identify prominent authors, research clusters, and influential collaborations. This analysis can provide insights into the structure of research communities, interdisciplinary collaborations, and evolving trends within a specific domain.

Co-occurrence analysis is a method used in bibliometrics and social network analysis to examine the relationship between terms or keywords in corpus of documents. It involves analysing the co-occurrence patterns of terms to identify meaningful associations or thematic clusters. According to He [14], co-occurrence analysis is a method that examines how co-occurrence patterns are used in textual words to determine the connections between concepts within the topics covered in the texts. In order to examine the relationship between concepts that co-occur among the keywords, the analysis takes the keywords out of the paper’s title, abstract, and keywords.

3.1 Data Collection Strategy

The Scopus database is utilized for bibliometric research works. Table 1 shows the search query applied in Scopus database. Three keywords are used, which produced 218 documents. The search is then restricted to document type of article and conference paper, English language and source type to only journal and conference proceeding, which left the data to 174 documents.

Table 1
Database search string

Database	Search string
Scopus	(“DNA” AND “splicing system*” OR “splicing language*”)

3.2 Data Tools and Analysis

This study is carried out by using VOSviewer package. It is employed to map the field's conceptual framework. The performance analysis is conducted to reveal the number of documents within a time frame, the countries and the sources involved. The citation analysis is provided to disclose the most prominent authors and research works. The academic relationships between scientific groups were described using co-citation and co-authorship. In addition to providing direction for future research, the co-occurrence analysis is finally presented to identify the themes and clusters within the area.

4. Analysis and Interpretation

4.1 Performance Analysis

There are 174 documents associated with 84 journals or conference proceedings sources. The documents are taken within the range from 1983 until 2023. Throughout the years, the most published document is in 2015, with a total of 11 documents. Fluctuation in the publication of documents between 2005 to 2014 occurred, where 10 documents were published in 2005, 2011 and 2014. A steady increase in the publication is observed between the decades, where a total of 12 documents were produced between 1983 to 1990, 40 documents in the next decade, and increase to 47 and 59 documents between 2001 to 2010, and 2011 to 2020, respectively. A detail distribution on the number of documents by year is shown in Figure 1:

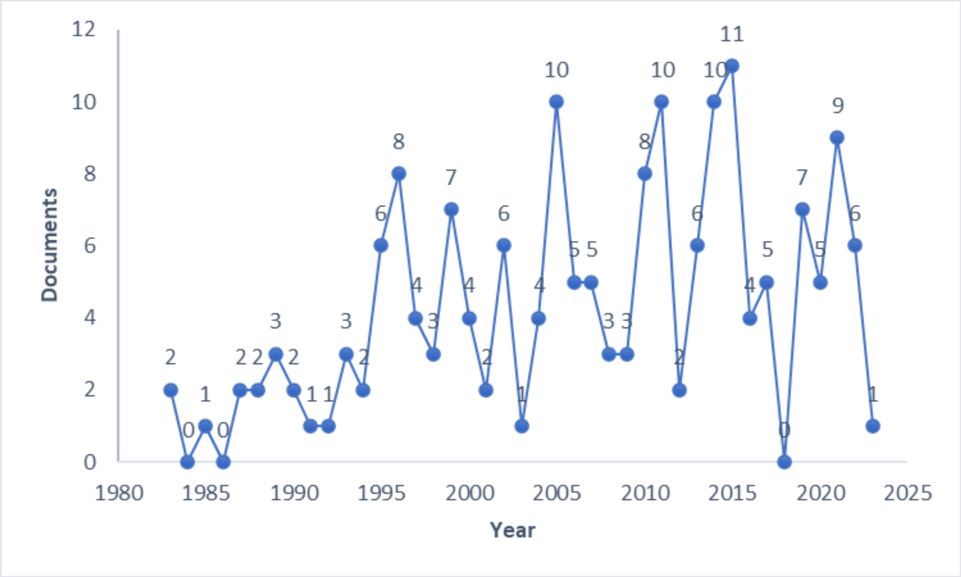


Fig. 1. Documents by year ranging from 1983 - 2023

The top ten countries that contributed to the production of papers is given in the chart below (Figure 2). Malaysia contributed the most with 48 documents and followed by United States (45 documents). India and Italy produced 17 documents, while Japan with 16 documents. The countries are chosen with the minimum of five documents with no minimum number of citations.

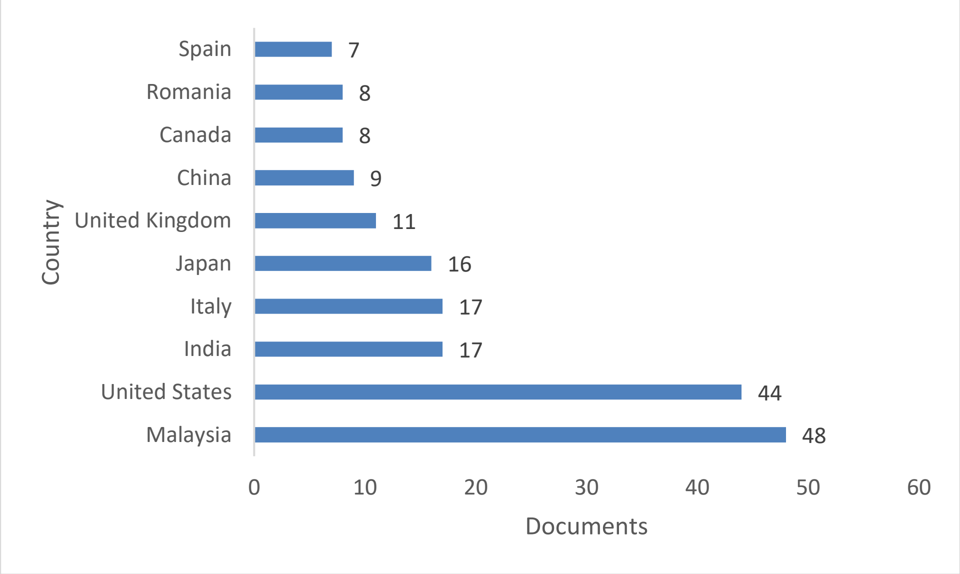


Fig. 2. Documents by country

Other than the countries that produced the most documents, the most relevant journals and conference proceedings are given as follows (Table 2), with a condition of obtaining a minimum of five documents of a source:

Table 2
Most relevant journals and conference proceedings

No.	Source	Source type	Number of citations	Number of documents
1.	Journal of Biological Chemistry	Journal	598	7
2.	Discrete Applied Mathematics	Journal	265	6
3.	Nucleic Acids Research	Journal	213	6
4.	Theoretical Computer Science	Journal	210	15
5.	International Journal of Computer Mathematics	Journal	78	5
6.	AIP Conference Proceedings	Conference Proceeding	33	24
7.	Jurnal Teknologi	Journal	4	6
8.	Journal of Physics: Conference Series	Conference Proceeding	3	6

4.2 Citation Analysis

In the aspect of research policy and research system, citations are becoming more popular as performance indicators. Aksnes *et al.,* [15] suggested that citations are commonly presumed to reflect the influence or quality of the research, while Garfield [11] stressed that when assessing and determining the major contributors in a research work, citation analysis can be a helpful tool for adding impartiality to the review process at a rather low cost. The citation analysis critically overviews the most influential and highly cited documents in the DNA splicing system, as given in section 4.2.1.

4.2.1 Most influential documents

Table below shows the most influential documents based on the citation gathered from a paper published in any Scopus's journals or conference proceedings. By setting up the number of citations per document to 91, of the 174 documents, 10 of them meet the threshold. The top ten most influential documents with its number of citations are given in Table 3.

Table 3
Most influential documents

No.	Authors/ Year	Title	Source title	Citation
1.	Head (1987) [5]	Formal language theory and DNA: An analysis of the generative capacity of specific recombinant behaviors	Bulletin of Mathematical Biology	515
2.	Chong *et al.,* (1996) [16]	Protein Splicing Involving the *Saccharomyces cerevisiae* VMA Intein: The Steps in the Splicing Pathway, Side Reactions Leading to Protein Cleavage, and Establishment of an *In Vitro* Splicing System	Journal of Biological Chemistry	216
3.	Evans *et al.,* (2000) [17]	Protein *trans*-Splicing and Cyclization by a Naturally Split Intein from the dnaE Gene of *Synechocystis* species PCC6803*	Journal of Biological Chemistry	170
4.	Ruby and Abelson (1988) [18]	An Early Hierarchic Role of U1 Small Nuclear Ribonucleoprotein in Spliceosome Assembly	Science	163
5.	Hull *et al.,* (2007) [19]	Identification of Common Genetic Variation that Modulates Alternative Splicing	PLoS Genetics	129
6.	Pixton (1996) [7]	Regularity of Splicing Languages	Discrete Applied Mathematics	119
7.	Li and Sugiura (1990) [20]	Three Distinct Ribonucleoproteins from Tobacco Chloroplasts: Each Contains a Unique	EMBO Journal	113

		Amino Terminal Acidic Domain and Two Ribonucleoprotein Consensus Motifs		
8.	Gordon *et al.,* (2000) [21]	Metal Ion Catalysis during the Exon-Ligation Step of Nuclear Pre-mRNA Splicing: Extending the Parallels between the Spliceosome and Group II Introns	RNA	108
9.	Shi and Muir (2005) [22]	Development of a Tandem Protein Trans-Splicing System based on Native and Engineered Split Inteins	Journal of the American Chemical Society	102
10.	Culik and Harju (1991) [23]	Splicing Semigroups of Dominoes and DNA	Discrete Applied Mathematics	92

From these ten most cited documents, the direction of writing can be divided into two; biological and mathematical. Head [5] recorded the highest number of citations, which introduced the most phenomenal findings called Head splicing system. The discussion on the Formal Language Theory which linked the theoretical computer science and discrete applied mathematics, had boost more researches until today. Differ with Head, Chong *et al.,* [16], Evans *et al.,* [17], Ruby and Abelson [18], Hull *et al.,* [19], Li and Sugiura [20], Gordon *et al.,* [21], and Shi and Muir [22], discussed the splicing system in biological perspectives. This is significant with the source of journal that published the papers which fulfil the scope of the journals. Then, Pixton [7] deliberated on the generated splicing language, that is the outcome of splicing system, while Culik and Harju [23] focused on the splicing semigroups of dominoes.

4.3 Co-citation Analysis

Tracking sets of research works that are regularly cited in the same source articles is known as co-citation analysis. Osareh [24] in a research work in 1996 conferred that the process of co-citation links cited documents together. The new method of co-citation analysis can be used to study the cognitive structure of science. When multiple writers co-cite the same pairs of papers, research clusters formed. Usually, these clusters consist of co-cited works with a same theme. According to Surwasee [25], co-citation analysis, when combined with multidimensional scaling techniques and single-link clustering, may physically map the structure of science overall as well as specialised research sectors. In this paper, co-citation analysis is conducted based on cited references.

4.3.1 Co-citation (cited references)

From 3159 cited references, 76 cited references were attained by having the citation threshold at a minimum number of three citations. Figure 3 depicts a network analysis of DNA splicing system based on the cited references.

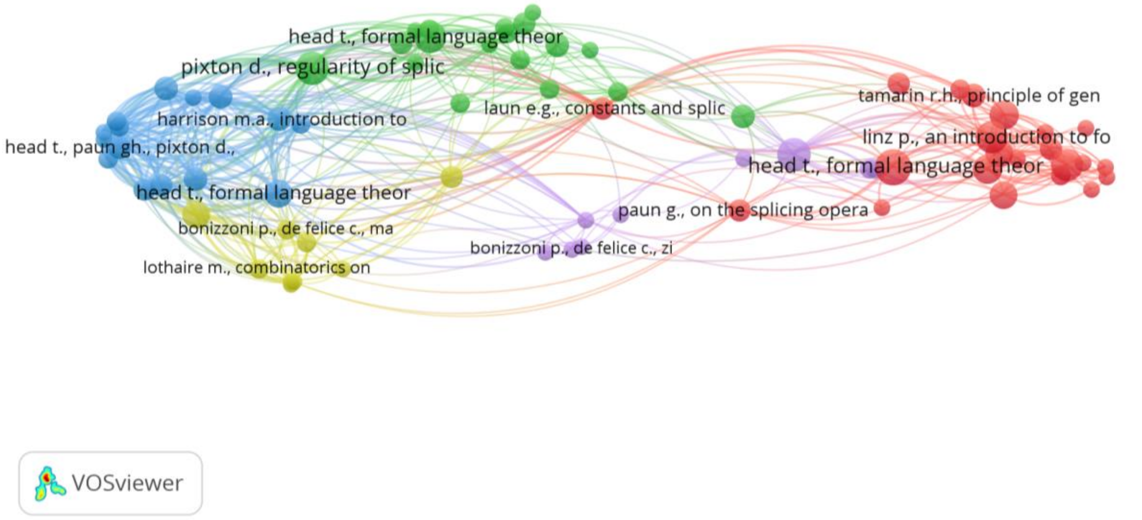


Fig. 3. Co-citation of authors

The list of top ten cited works is then given in Table 4, arranged by the top ten highest co-cited total link strength. Top ten of co-citation (cited references)

Table 4
Top ten of co-citation (cited references)

No.	Authors/ Year	Title	Citation	Total link strength
1.	Pixton (1996) [7]	Regularity of splicing languages	12	87
2.	Hopcroft and Ullman (1979) [26]	Introduction to automata theory	8	86
3.	Head (1987) [5]	Formal language theory and DNA: An analysis of the generative capacity of specific recombinant behaviors	8	68
4.	Bonizzoni and Mauri (2005) [27]	Regular splicing languages and subclasses	5	63
5.	Head *et al.,* (1997) [28]	Language theory and molecular genetics: Generative mechanisms suggested by DNA recombination	5	63
6.	Berstel and Perrin (1985) [29]	Theory of codes	5	62
7.	Paun *et al.,* (1998) [30]	DNA computing	6	57
8.	Paun (1996) [6]	On the splicing operation	6	55
9.	Perrin (1990) [31]	Finite automata	4	55
10.	Culik and Harju (1991) [23]	Splicing semigroups of dominoes and DNA	4	52

4.4 Co-authorship Analysis

Co-authorship analysis illustrates the co-authorship link among top authors. This type of analysis is conducted in order to recognize the main authors who are produced academic work in splicing system that will benefit researchers in this field in initiating new academic works in the future to enhance the field of splicing system, and to acknowledge their contribution in the literature on splicing system. In this paper, the co-authorship analysis is done by seeing the analysis by cited authors.

4.4.1 Co-authorship (cited authors)

Of 4225 authors, the lowest number of citations per author is set to 83, giving 10 authors meet the threshold. From the three clusters, it is observed that the authors are involve in the research of splicing systems concerning on the development of research on splicing system, splicing languages and type of splicing systems.

From the co-authorship map as shown in Figure 4, and its data as given in Table 5, there are research that co-authored from different disciplinary backgrounds, showing the enhancement or adoption of multi-field in DNA splicing system, for instance, Christy *et al.,* [32] that has affiliation from mathematics and computer science departments, Selvarajoo *et al.,* [33] from mathematics, information technology, and science studies departments. In Amos *et al.,* [34], the articles combined authors from mathematics and computer science background, from six countries. The study that discussed on the theory of DNA computing had been cited by 64 other publications. This shows that, the combination of backgrounds and international co-authored had potential in increasing the number of citations. Another factor that is possible to increase the citation is by collaborating with industry. For example, article by Evans *et al.,* [17], that combined authors from England Biolabs Inc., and also authors from different countries; Switzerland, Canada and United Stated. This research work had been cited by 170 times.

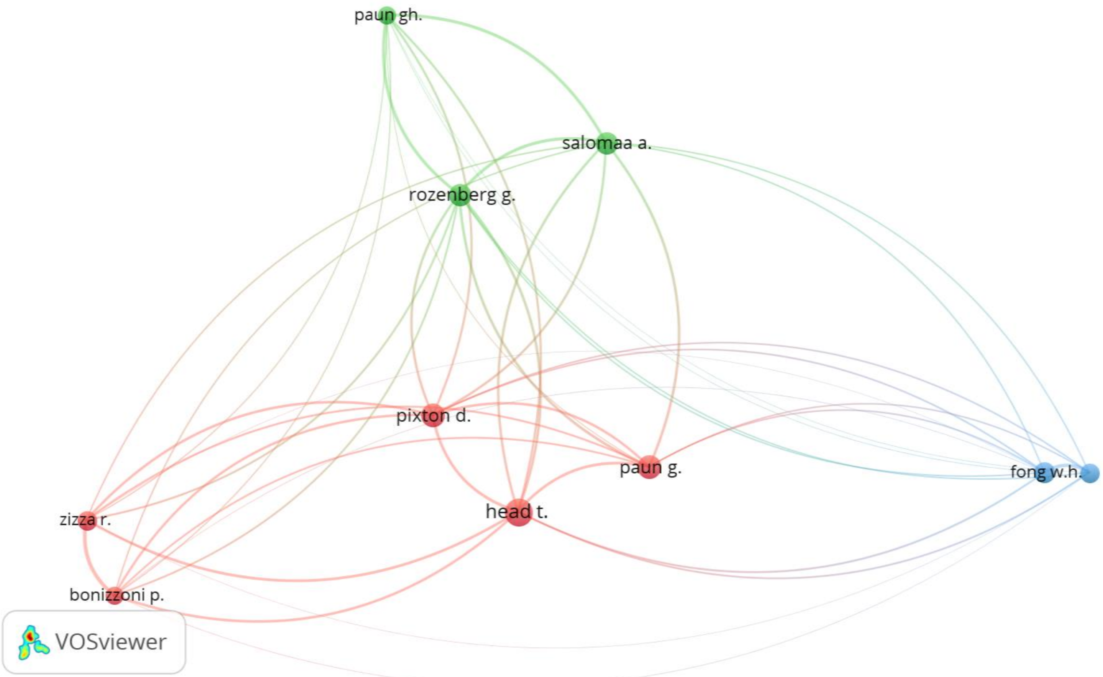


Fig. 4. Co-authorship by authors

Table 5
Co-authorship of top ten authors

No.	Authors	Citations	Total link strength
1.	Head, T.	197	1959
2.	Rozenberg, G.	127	1753
3.	Pixton, D.	139	1749
4.	Salomaa, A.	126	1610
5.	Paun, G.	148	1456
6.	Zizza, R.	94	1410
7.	Bonizzoni, P.	84	1330
8.	Paun, Gh.	90	1145
9.	Fong, W. H.	109	906
10.	Sarmin, N. H.	95	906

4.5 Co-occurrence Analysis

In 1999, He [14] suggested that the co-occurrence analysis is used to track scientific progress by identifying connections between topics in a research field based on the frequency of co-occurrence of pairs of words or phrases. Additionally, Wang [35] discovered that, the solution to issues with the keywords and indexes is to include professionals' proficiency in co-occurrence analysis. In this paper, the co-occurrence analysis is conducted based on author keywords. By setting the least number of incidences of a keyword to three, from 242 keywords, 30 of them meet the threshold, which divided the data into six clusters. The following were expanded upon and further discussed based on each cluster.

Cluster 1 (Red): This cluster is confined with seven keywords (dna, dna computer, restriction enzyme, splicing language, splicing languages, splicing system, and yusof-goode splicing system) with the theme "genetics and biotechnology". The literatures are mainly focused on the research of splicing systems and splicing languages. Ahmad *et al.,* [36] discussed on the presence of second order limit language and its splicing languages when the system involved any strings that has two cutting sites. In addition, Fong *et al.,* [37] focused on the modelling of DNA splicing systems when the system adopted non-palindromic restriction enzymes. More interestingly, Lim *et al.,* [38], Mudaber *et al.,* [39] and Ahmad *et al.,* [40] conducted their research in splicing systems by exploitation of Yusof-Goode splicing systems, that are literally discussed on the elements of splicing systems in producing different types of splicing languages, namely single-stage splicing language, two-stages splicing languages and second-order limit language, respectively.

Cluster 2 (Green): This cluster contains five keywords (automata, computational power, dna computing, probabilistic splicing system, and regular languages). With the theme "emerging technologies in computation", the literatures show an adoption of mathematical approach in solving problems in splicing system. Khairuddin *et al.,* [41], Bonizzoni *et al.,* [27] and Ray *et al.,* [42] indicated the use of automata theory in their study. The use of theory will support the DNA computing, specifically in predicting the output of splicing system, called splicing language.

Cluster 3 (Blue): This cluster contains five keywords (permanent, persistent, second order limit language, two stages splicing languages, and y-g splicing system). This cluster shows a theme of "formal languages and compatibility". In this cluster, the keywords mainly focused on the type of splicing system, for instance, permanent, persistent, second-order limit language, two-stages splicing languages and Y-G splicing system. Permanent splicing system which was first introduced by Head [5] is later adopted in the discussion of two-stages splicing system by Mudaber *et al.,* [43] that was discussed by using Y-G splicing system, proposed by Yusof *et al.,* [8].

Cluster 4 (Yellow): This cluster contains four keywords (automata theory, formal languages, molecular computing, and splicing). This cluster is correlated or linked to cluster 2, where it focused on the adoption of mathematical approach in the literatures. Other than that, it also linked to cluster 5 that discussed on the mathematical approach in solving problems in splicing system. This is supported by comprehensive review conducted by Mohd Ruslim *et al.,* [44] that gave an insightful overview on the adoption of graphical approach in DNA splicing system.

Cluster 5 (Purple): This cluster contains four keywords DNA molecules (dna recombination, formal language theory, and splicing systems). With the theme of “molecular biology and genetic engineering”, the literatures mainly focused on the foundation knowledge of splicing system and splicing language. From the figure, it can be seen that, Cluster 5 is linked to Cluster 1, 2, 3 and 4, which specify the correlation between the foundation knowledge in splicing system and the adoption of mathematical approach in splicing system. This is relevant with the framework or Formal Language Theory by Head [5] that shows the two field that corresponds to the framework that are, theoretical computer science and discrete applied mathematics.

Cluster 6 (Light Blue): This cluster contains four keywords (dna sequence, image operation, splicing operation, and turing machine). With the theme of “computational biology and bioinformatics”, the research works focused on the adoption of mathematical or other methods in the discussion of splicing system. Zhou *et al.,* [45] transformed the digital image encryption scheme into DNA sequence based on DNA code theory, while Ray *et al.,* [46] developed a Turing machine which transformed the input data into the form of DNA sequences.

From the six clusters gained in the co-occurrence analysis as shown in Figure 5, the top 15 keywords are mainly from articles that were developed based on the hybridization of two fields, as listed in Table 6. The concept presented in the articles were mainly discussed on the splicing system and the generation of splicing languages. Besides, mathematical approach was injected either in finding or validating the results. Modification to the splicing system proposed by Yusof in 2012 and new ideas that applied to the study of wet model of splicing system had proven that this study have a positive direction to be further explored.

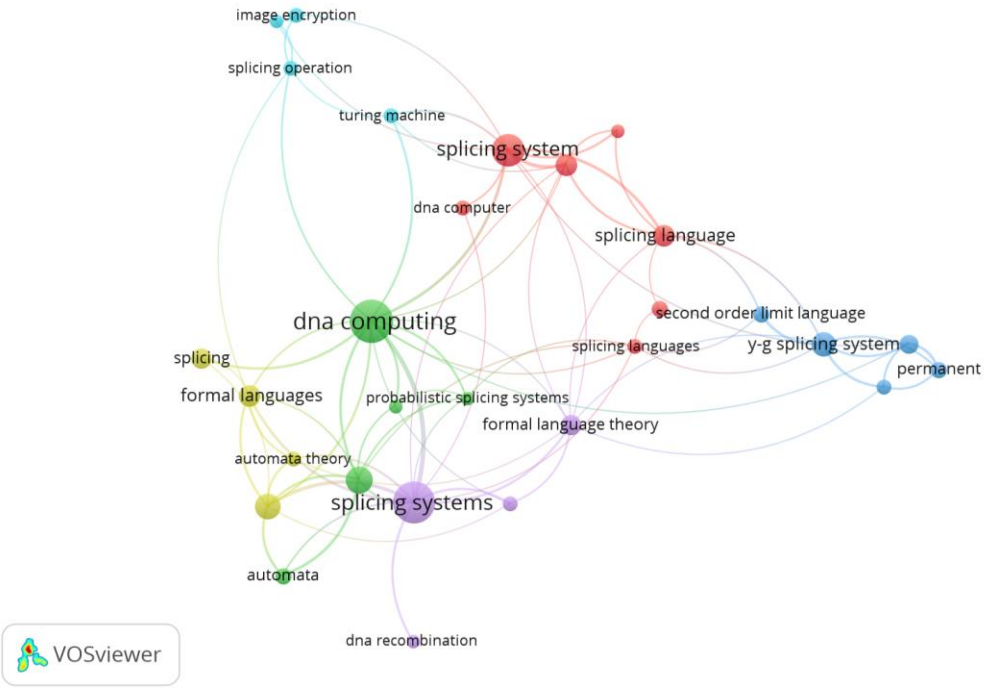


Fig. 5. Co-occurrence of six clusters

Table 6
The co-occurrence keywords analysis

Rank	Keyword	Occurrences	Total link strength
1	DNA computing	31	46
2	Splicing systems	29	44
3	Splicing system	18	28
4	Y-G Splicing System	10	19
5	Formal languages	8	17
6	Splicing language	8	16
7	Formal Language Theory	7	12
8	Permanent	5	11
9	Automata	5	11
10	Automata theory	4	10
11	Second order limit language	5	8
12	Splicing operation	4	8
13	Splicing Languages	4	7
14	Turing machine	4	7
15	Splicing	7	6

5. Conclusion

This paper delivers an indication on the most influential papers, authors and countries in splicing system research founded in the Scopus database. The research was managed by using VOSviewer software.

The performance analysis was first given. The production of the research works was observed between 1983 until 2023. This involves 174 documents that had been published in Scopus database either in journals or proceedings. These documents were originated mostly from Malaysia, followed by United States, India, Italy and Japan. From the list of Scopus indexed sources, it can be categorized into two, which are related to mathematical and biological field.

The performance analysis can also be observed from the citation analysis. Ten most cited documents were discovered, showing the most cited is from a document that was in mathematical scope. However, seven out of ten documents were from biological scope, that shown among the ten most cited documents. This shows that, the splicing system is established in biological scope rather than mathematical scope.

Then, from the co-citation analysis, the trend of the research in the context of DNA splicing system between 1983 until 2023 was captured. The co-citation analysis was conducted based on cited references. From the data obtained, it is observed that, the ten highest total link strength involved ten references that discussed splicing system in mathematical scope.

Then, from the co-authorship analysis, the link between the countries and authors was observed. Most of the authors that co-authored were also found in the top ten authors in splicing system. The collaboration among authors shows a pattern of knowledge hybridization where there are multiple disciplines of authors collaborated in a research work. The collaborations show linkages between authors from Asian, European, and American countries, which can increase the citation of articles in the future.

Finally, from the co-occurrence analysis, the trend of research in the future can be predicted. The highest number of occurrences of a word shows that the word frequently appeared either in the titles, abstracts, or full texts. This will give direction to scholars in predicting the future trend of research. For example, in this paper, the future of research involving DNA computing can be forecasted, where from the six clusters gained, the merging between mathematical and biological perspectives in DNA splicing systems can be observed. The research in splicing system is expand with the implementation of mathematical approach when most of the clusters suggested that the emerging technologies being the theme of the keywords in the clusters, which is also deliberated in other research work in other fields as proposed by Johan *et al.,* [47].

Acknowledgement

The authors would like to express their appreciation to Ministry of Higher Education for providing financial support under Fundamental Research Grant Scheme (FRGS) No. FRGS/1/2021/STG06/UMP/02/1 (University reference RDU210107) and Universiti Malaysia Pahang Al-Sultan Abdullah under UMP Fundamental Research Grant RDU220344.

References

[1] Tamarin, Robert H. (2001). *Chapter 1 - Genetics and the Scientific Method,* Page 1-10 - 1-11.

[2] Li, Jinbao, Huize Sun, Yulin Huang, Yali Wang, Yuyan Liu, and Xuefeng Chen. "Pathways and assays for DNA double-strand break repair by homologous recombination." *Acta biochimica et biophysica Sinica* 51, no. 9 (2019): 879-889. https://doi.org/10.1093/abbs/gmz076

[3] Ensminger, Michael, and Markus Löbrich. "One end to rule them all: Non-homologous end-joining and homologous recombination at DNA double-strand breaks." *The British journal of radiology* 93, no. 1115 (2020): 20191054. https://doi.org/10.1259/bjr.20191054

[4] Epum, Esther A., and James E. Haber. "DNA replication: the recombination connection." *Trends in cell biology* 32, no. 1 (2022): 45-57. https://doi.org/10.1016/j.tcb.2021.07.005

[5] Head, Tom. "Formal language theory and DNA: an analysis of the generative capacity of specific recombinant behaviors." *Bulletin of mathematical biology* 49 (1987): 737-759. https://doi.org/10.1016/S0092-8240(87)90018-8

[6] Păun, Gheorghe. "On the splicing operation." *Discrete Applied Mathematics* 70, no. 1 (1996): 57-79. https://doi.org/10.1016/0166-218X(96)00101-1

[7] Pixton, Dennis. "Regularity of splicing languages." *Discrete Applied Mathematics* 69, no. 1-2 (1996): 101-124. https://doi.org/10.1016/0166-218X(95)00079-7

[8] Yusof, Yuhani. "Dna splicing system inspired by bio molecular operations." PhD diss., Universiti Teknologi Malaysia, 2012.

[9] Donthu, Naveen, Satish Kumar, Debmalya Mukherjee, Nitesh Pandey, and Weng Marc Lim. "How to conduct a bibliometric analysis: An overview and guidelines." *Journal of business research* 133 (2021): 285-296. https://doi.org/10.1016/j.jbusres.2021.04.070

[10] Ellegaard, Ole, and Johan A. Wallin. "The bibliometric analysis of scholarly production: How great is the impact?." *Scientometrics* 105 (2015): 1809-1831. https://doi.org/10.1007/s11192-015-1645-z

[11] Garfield, Eugene. "Is citation analysis a legitimate evaluation tool?." *Scientometrics* 1 (1979): 359-375. https://doi.org/10.1007/BF02019306

[12] Park, Hyejin, and Peter Shea. "A Review of Ten-Year Research through Co-citation Analysis: Online Learning, Distance Learning, and Blended Learning." *Online Learning* 24, no. 2 (2020): 225-244. https://doi.org/10.24059/olj.v24i2.2001

[13] Small, Henry. "Citation indexing revisited: Garfield's early vision and its implications for the future." *Frontiers in Research Metrics and Analytics* 3 (2018): 8. https://doi.org/10.3389/frma.2018.00008

[14] He, Qin. "Knowledge discovery through co-word analysis." (1999).

[15] Aksnes, Dag W., Liv Langfeldt, and Paul Wouters. "Citations, citation indicators, and research quality: An overview of basic concepts and theories." *Sage Open* 9, no. 1 (2019): 2158244019829575. https://doi.org/10.1177/2158244019829575

[16] Chong, Shaorong, Yang Shao, Henry Paulus, Jack Benner, Francine B. Perler, and Ming-Qun Xu. "Protein splicing involving the Saccharomyces cerevisiae VMA intein: the steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system." *Journal of Biological Chemistry* 271, no. 36 (1996): 22159-22168. https://doi.org/10.1074/jbc.271.36.22159

[17] Evans, Thomas C., Deana Martin, Reto Kolly, Daniel Panne, Luo Sun, Inca Ghosh, Lixin Chen, Jack Benner, Xiang-Qin Liu, and Ming-Qun Xu. "Protein trans-Splicing and Cyclization by a Naturally Split Intein from the dnaE Gene ofSynechocystis Species PCC6803." *Journal of Biological Chemistry* 275, no. 13 (2000): 9091-9094. https://doi.org/10.1074/jbc.275.13.9091

[18] Ruby, Stephanie W., and John Abelson. "An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly." *Science* 242, no. 4881 (1988): 1028-1035. https://doi.org/10.1126/science.2973660

[19] Hull, Jeremy, Susana Campino, Kate Rowlands, Man-Suen Chan, Richard R. Copley, Martin S. Taylor, Kirk Rockett et al., "Identification of common genetic variation that modulates alternative splicing." *PLoS genetics* 3, no. 6 (2007): e99. https://doi.org/10.1371/journal.pgen.0030099

[20] Li, Y. Q., and Masahiro Sugiura. "Three distinct ribonucleoproteins from tobacco chloroplasts: each contains a unique amino terminal acidic domain and two ribonucleoprotein consensus motifs." *The EMBO journal* 9, no. 10 (1990): 3059-3066. https://doi.org/10.1002/j.1460-2075.1990.tb07502.x

[21] Gordon, Peter M., ERIK J. Sontheimer, and Joseph A. Piccirilli. "Metal ion catalysis during the exon-ligation step of nuclear pre-mRNA splicing: extending the parallels between the spliceosome and group II introns." *Rna* 6, no. 2 (2000): 199-205. https://doi.org/10.1017/S1355838200992069

[22] Shi, Jianxin, and Tom W. Muir. "Development of a tandem protein trans-splicing system based on native and engineered split inteins." *Journal of the American Chemical Society* 127, no. 17 (2005): 6198-6206. https://doi.org/10.1021/ja042287w

[23] Culik II, Karel, and Tero Harju. "Splicing semigroups of dominoes and DNA." *Discrete Applied Mathematics* 31, no. 3 (1991): 261-277. https://doi.org/10.1016/0166-218X(91)90054-Z

[24] Osareh, Farideh. "Bibliometrics, citation analysis and co-citation analysis: A review of literature I." (1996): 149-158. https://doi.org/10.1515/libr.1996.46.3.149

[25] Surwase, Ganesh, Anil Sagar, B. S. Kademani, and K. Bhanumurthy. "Co-citation analysis: An overview." (2011): 179-185.

[26] Hopcroft, John E., Rajeev Motwani, and Jeffrey D. Ullman (1979). *Chapter 1 – Automata: The Methods and the Madness*, Page 1-5 - 1-17.

[27] Bonizzoni, Paola, and Giancarlo Mauri. "Regular splicing languages and subclasses." *Theoretical computer science* 340, no. 2 (2005): 349-363. https://doi.org/10.1016/j.tcs.2005.03.035

[28] Head, Tom, Gheorghe Paun, and Dennis Pixton (1997). "Language Theory and Molecular Genetics: Generative Mechanisms Suggested by DNA Recombination." *In Handbook of Formal Languages, edited by Rozenberg, G. and Salomaa, A. Berlin, Heidelberg: Springer* (1997): 295 – 360. https://doi.org/10.1007/978-3-662-07675-0_7

[29] Berstel, Jean and Dominique Perrin. "Theory of Codes." *Bulletin of the American Mathematical Society* 17, no. 1 (1987): 161 – 167.

[30] Paun, Gheorghe, Grzegorz Rozenberg, and Arto Salomaa (1998). *Chapter 2 – DNA: Its Structure and Processing*, Page 2-9 – 2-41.

[31] Perrin, Dominique (1990). *Chapter 1 – Finite Automata*, Page 1-1 – 1-57.

[32] Christy, D. K., P. John Paul, and D. G. Thomas. "H-array system on picture languages with array rewriting rules." In *AIP Conference Proceedings*, vol. 2277, no. 1. AIP Publishing, 2020. https://doi.org/10.1063/5.0025583

[33] Selvarajoo, Mathuri, Sherzod Turaev, Wan Heng Fong, and Nor Haniza Sarmin. "The Generative Capacity of Probabilistic Splicing Systems." *Applied Mathematics & Information Sciences* 9, no. 3 (2015): 1191. https://doi.org/10.1063/1.4882571

[34] Amos, Martyn, Gheorghe Păun, Grzegorz Rozenberg, and Arto Salomaa. "Topics in the theory of DNA computing." *Theoretical computer science* 287, no. 1 (2002): 3-38. https://doi.org/10.1016/S0304-3975(02)00134-2

[35] Wang, Zhong-Yi, Gang Li, Chun-Ya Li, and Ang Li. "Research on the semantic-based co-word analysis." *Scientometrics* 90, no. 3 (2012): 855-875. https://doi.org/10.1007/s11192-011-0563-y

[36] Ahmad, Muhammad Azrin, Nor Haniza Sarmin, Wan Heng Fong, Yuhani Yusof, and Noraziah Adzhar. "Second Order Limit Language with Two Cutting Sites."

[37] Fong, Wan Heng, Nurul Izzaty Ismail, and Nor Haniza Sarmin. "The Modelling of DNA Splicing Systems with Two Non-Palindromic Restriction Enzymes." *Indian Journal of Public Health Research & Development* 10, no. 6 (2019). https://doi.org/10.5958/0976-5506.2019.01515.8

[38] Lim, Wen Li. "Modelling the Behaviour of Single Stage Splicing Language: A Yusof Goode Computational Approach." *Jurnal Teknologi* 73, no. 1 (2015): 135-138. https://doi.org/10.11113/jt.v73.3560

[39] Mudaber, Mohammad Hassan. "PERSISTENCY AND PERMANENCY OF TWO STAGES SPLICING LANGUAGES VIA YUSOF-GOODE APPROACH: TWO INITIAL STRINGS AND TWO RULES." *JURNAL TEKNOLOGI* 77, no. 1 (2015). https://doi.org/10.11113/jt.v77.3701

[40] Ahmad, Muhammad Azrin, Nor Haniza Sarmin, Yuhani Yusof, and Wan Heng Fong. "Some restrictions on the existence of second order limit language." In *AIP Conference Proceedings*, vol. 1682, no. 1. AIP Publishing, 2015. https://doi.org/10.1063/1.4932457

[41] Khairuddin, Siti Hajar Mohd, Muhammad Azrin Ahmad, and Noraziah Adzhar. "Application of Automata Theory On n-th Order Limit Language." *Review of International Geographical Education Online* 11, no. 4 (2021): 817-824.

[42] Ray, Kumar S., and Mandrita Mondal. "Splicing operation and fuzzy molecular automaton." *International Journal of Intelligent Computing and Cybernetics* 5, no. 2 (2012): 206-225. https://doi.org/10.1108/17563781211231543

[43] Mudaber, Mohammad Hassan, Yuhani Yusof, Mohd Sham Mohamad, Aizi Nor Mazila Ramli, and Wen Li Lim. "Modelling of two stages DNA splicing languages on de Bruijn graph." *Jurnal Teknologi* 78, no. 1 (2016): 73-78. https://doi.org/10.11113/jt.v78.4236

[44] Ruslim, Nooradelena Mohd, Marta Elizabeth, Yuhani Yusof, Mohd Sham Mohamad, and Noraziah Adzhar. "Deoxyribonucleic acid (DNA) splicing system from graph theoretic perspective." In *Journal of Physics: Conference Series*, vol. 1988, no. 1, p. 012081. IOP Publishing, 2021. https://doi.org/10.1088/1742-6596/1988/1/012081

[45] Zhou, Changjun, Xiaopeng Wei, Qiang Zhang, and Rui Liu. "DNA sequence splicing with chaotic maps for image encryption." *Journal of Computational and Theoretical Nanoscience* 7, no. 10 (2010): 1904-1910. https://doi.org/10.1166/jctn.2010.1558

[46] Ray, Kumar Sankar, and Mandrita Mondal. "Fuzzy molecular automaton using splicing theory." *International Journal of Bio-Inspired Computation* 3, no. 5 (2011): 320-330. https://doi.org/10.1504/IJBIC.2011.042258

[47] Johan, Azri, Nurhayati Issa, Yoshiyuki Matsuura, and Mohammad Ali Tareq. "Technology Exploration and Technology Exploitation in Japanese Chemical Companies." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 32, no. 3 (2023): 75-91. https://doi.org/10.37934/araset.32.3.7591