

## A Bibliometric Mapping of Plant-Based Milk

M. Irfan Febriansyah<sup>1</sup>, Nafisah Eka Puteri<sup>2\*</sup>, Mustafa Kamal<sup>2</sup>

<sup>1</sup>Program Studi Gizi, Fakultas Kesehatan Masyarakat, Universitas Teuku Umar,  
Alue Peunyareng, Meurebo, Aceh Barat 23681, Indonesia

<sup>2</sup>Program Studi Teknologi Hasil Pertanian, Fakultas Pertanian, Universitas Teuku Umar,  
Alue Peunyareng, Meurebo, Aceh Barat 23681, Indonesia

\*Email: nafisahekaputeri@utu.ac.id

### ABSTRACT

Non-dairy or plant-based milk is one of the new food products derived from plants. It is essential to explore the plant milk market from the different raw materials that can be used, understand the different products we can obtain from plant milk, as well as improve the characteristics of the plant milk. To do the exploratory, bibliographic research can be performed. Therefore, we applied bibliographic analysis to identify and explore the most advanced technology in plant milk research. Data was collected by using Harzing's Publish or Perish (PoP), saved as CSV and RIS, filtered and refined, and then analyzed using descriptive statistics and VOSviewer. We screened 485 papers (2012-2022) related to plant-based milk from 998 publications since 1981. Related keywords or terms of screened papers were then visualized by VOSviewer into network, overlay, and density visualization map. Based on the visualization map, there were 7 clusters of plant-based milk research field. The novelty of the research covered the importance and development of plant-based milk derivate, correlation to life cycle assessment, and occurrence of milk allergy. The term "diet" became the highest density term in publication between 2012 to 2022. On the other hand, the terms "traditional plant" and "health benefit" became the lowest density term.

**Keywords:** Bibliometric analysis; Non-dairy milk; Plant-based milk; Publish or Perish; VOSviewer

### ABSTRAK

Susu nabati merupakan salah satu pangan masa depan yang diproduksi dari bahan nabati. Hingga saat ini, penelitian dan pengembangan terkait susu nabati masih dilakukan. Maka dari itu, penting untuk melakukan eksplorasi terhadap bahan mentah yang dapat dimanfaatkan, mengenali produk turunan yang dapat dihasilkan dari susu nabati, serta mengembangkan susu nabati dengan karakteristik yang dapat diterima oleh konsumen. Dalam upaya mencapai tujuan tersebut, riset bibliometrik dapat diterapkan. Melalui penelitian ini, peneliti berupaya menerapkan analisis bibliometrik untuk mengidentifikasi dan mengkaji teknologi serta perkembangan riset terkini terkait susu nabati. Metode analisis bibliometrik yang diterapkan mencakup proses panen data artikel dengan menggunakan Harzing's Publish or Perish (PoP), penyimpanan data sebagai CSV and RIS, penapisan dan penyaringan data, serta proses analisis deskriptif dan pemetaan dengan VOSviewer. Melalui proses pemilahan, didapat 485 publikasi dari 998 publikasi terkait susu nabati yang dipublikasi pada rentang tahun 2012 hingga 2022. Kata kunci atau istilah yang berkenaan dengan lingkup riset dari publikasi terpilih kemudian divisualisasikan dengan VOSviewer dalam bentuk peta jaringan (*network*), peta paparan (*overlay*), dan peta densitas (*density*). Berdasar hasil visualisasi tersebut, didapat 7 kelompok istilah yang berkaitan erat. Selanjutnya, kebaruan riset melingkupi tingkat kepentingan dan pengembangan produk turunan susu nabati, termasuk hubungannya dengan efek terhadap lingkungan dalam siklus hidup, serta kejadian alergi susu. Istilah "diet" ditemukan dalam densitas yang paling tinggi pada publikasi terkait susu nabati, sementara istilah "traditional plant" dan "health benefit" ditemukan dalam densitas yang paling rendah.

**Kata kunci:** Analisis bibliometrik; Susu non-hewani; Susu nabati; Publish or Perish; VOSviewer

## INTRODUCTION

Recent changes in food products derived from plant as alternatives to animal food products, have been driven by an interest in healthier products, more sustainable food production systems and the need of new future food (Bekhit *et al.*, 2022). Non-dairy milk as milk substitution, or simply plant-based milk, is one of new future food product derived from plant. Plant-based milk is produced from the destruction of specific raw materials, subsequent extraction in water, and homogenization. The raw materials can be originated from grains, legumes, nuts, seeds, and vegetable (Silva *et al.*, 2022). Lactose intolerance, cow's milk protein allergy and vegan lifestyle are the main reasons for increased plant-based milk. Furthermore, there are few additional benefits of consuming this product, such as no cholesterol; low in saturated fat; high in fiber; and rich in unsaturated fats.

The market for plant-based products has the opportunity to grow, so there is an increase in research and development of plant-based milk substitutes and their derivatives and this has been implemented to provide more alternatives for consumers (Mc Clements *et al.*, 2019). Besides, plant-based milk is vastly different from cow's milk, so these differences need to be studied, in order to examine where changes can be made to obtain a plant-based milk that closely resembles cow's milk. In addition, it is necessary to study how mixtures of different raw materials can combine and improve the composition of plant-based milk. Furthermore, developing countries could use the abundant indigenous materials and thoroughly examine the production to be able to produce more affordable product. Therefore, it is essential to explore this plant-based milk market, whether in milk or its derivatives, from the different raw materials that can be utilized. It is also important to analyze, explore and understand the different products we can obtain from plant milk, as well as how to obtain them and improve them further for the end consumer. In order to perform the exploration, bibliometric research can be conducted.

Bibliometric research has evolved significantly over the past few years. It is an alternative of systematic literature review that involves the application of scientific mapping, or descriptive statistical analysis, or performance analysis on bibliographic data (Haddow, 2018). Bibliographical studies on the publication characteristics are important because the topics published, geographical distribution, and impact factors will allow scientists to identify key or main topics for research and uncover state of

the art in a particular research area. In this article, we applied bibliometric analysis in order to identify and uncover state of the art of plant-based milk research. Interestingly, to the best of our knowledge, there are no bibliometric studies related to plant-based milk previously. Therefore, this article is aimed to provide information related to plant-based milk publication that already exist, investigate the gap, and analyze the novelty of current research.

## METHODOLOGY

### Collecting and Refining Database

General methodological flow of our research is shown by Figure 1. Data were collected using the search term "plant-based milk". This term is commonly used to describe food product derived from plant which are resembling dairy milk. Data collection were performed in November 2022 by using Harzing's Publish or Perish (PoP) v.8.6.4198.8332 © 1990-2022 from Google Scholar database. The initial data, containing publication and citation information, bibliographic information, abstracts as well as the address of article web page, were used to construct and visualize directory network information. In other words, PoP used Google Scholar queries to get citation information, which is then analyzed and converted into statistics information. Data were saved as .csv files for processing using Microsoft® Excel 2019 and .ris files for processing using Mendeley Desktop v.1.19.8 © 2008-2020.

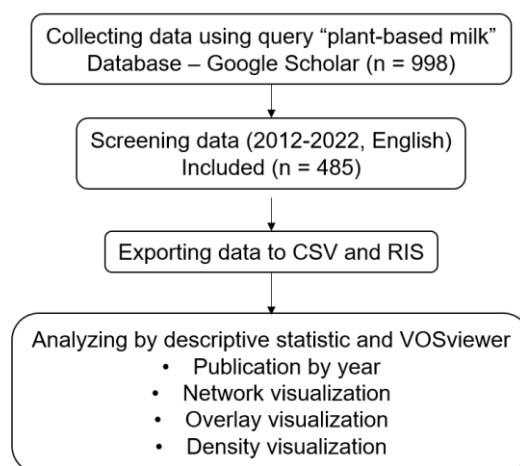


Figure 1. Research's flow chart

In the next step, we refined the initial data and narrowed down the publication year to recent 10 years. Duplicated and non-English articles were also eliminated. Articles were fully screened and reviewed by authors to validate if

they were indeed within the expected scope (Sweileh, 2020). Refined data or relevant articles (485 publication) were then extracted into a structured electronic data table and analyzed by publication year, citation, journal and key results. Descriptive statistics analysis was applied to quantify the difference.

**Bibliometric Analysis**

After refining initial data, the articles from the search term of “plant-based milk” were subjected to bibliographic analysis. Data were imported into VOSviewer v.1.6.18 (Center for Research in Science and Technology (CWTS), Leiden University, Leiden, Netherlands), a software that helped us to process the data and create visualization map of key terms drawn from a scientific literature. In order to improve visualization, we set minimum number of occurrences of term to 5. Then, VOSviewer calculated the relevance score of each term. Due to default choice of VOSviewer, we selected not more than 60% of the most relevant terms. About 60% of the most relevant terms or keywords were then visualized into network, overlay, and density visualization. These visualizations contained nodes and line. Nodes denoted key terms of publications, while line denoted the correlation between key terms.

By default, VOSviewer assigned the nodes or key terms of the network into clusters. A cluster is a compilation of closely related nodes or key terms. Each node in a network is assigned to specifically one cluster. The number of clusters is determined by a resolution parameter and represented by different colour in network visualization. Meanwhile, appearance of key terms by the time of publication was visualized by overlay visualization. VOSviewer may also show the density visualition of each key term. By default, the colors ranged from blue to green to red. The greater the number of acquaintances of a key term and the higher the weights of the acquaintances, the closer the color of the key term to red.

**RESULT AND DISCUSSION**

**Initial Data Collection**

Initial collected data of the term “plant-based milk” were gathered from Google Scholar database by using Publish or Perish (PoP), resulted in information about number of citations, rank, authors, title, year, type of publication, publisher, h-index as well as g-index. Information of publication years, number of papers, citations, h-index and g-index is shown in Table 1. There were 998 papers related to the term “plant-based milk” has been

found by PoP since 1981. However, these metrics data have not been refined and screened.

Table 1. Metrics data of “plant-based milk” between 1981 and 2022

Metrics Data	Terms
	“Plant-Based Milk”
Publication years	1981-2022 (41 years)
Number of papers	998
Number of citations	59050
Cites/year	1440.24
Cites/paper	59.17
h-index	123
g-index	225

**Data Extraction for Recent 10 Years**

After removing irrelevant and double paper, we classified database published in English by year. Then, we gathered the past 10 years papers by eliminating papers published before 2012 due to the low number of papers per year (less than 5 papers per year). We found 485 papers after this screening step. The number of papers published per year is shown in Figure 2.

Based on our initial data, the term “plant-based milk” has been studied in published paper since 1981. However, it was found in less than 5 papers per year in the range of 1981 to 2011. This indicated that the field of plant-based milk research is relatively new due to the long “lag” phase between 1981 and 2011. As can be seen in Figure 2, the annual number of publications related to plant-based milk has grown substantially after 2014. Interestingly, the number of publications has gradually increased since then. In 2022, the number of publications on the topic exceeded 100 papers per year, in November 2022 it was 123 papers published.

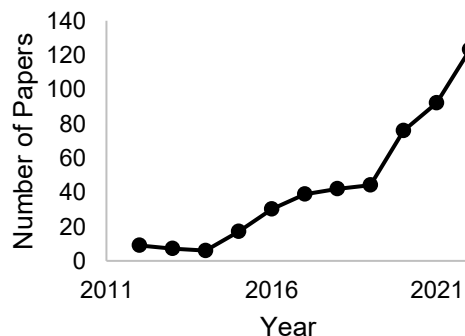


Figure 2. Annual number of papers published related to plant-based milk (2012-2022)

Publication related to plant-based milk has been cited 12105 times in Google Scholar record. The top 5 publication based on the number of citations is explained by Table 2. The most cited publication is a review article with the title of “Plant-based milk alternatives an emerging segment of functional beverages: a review” (Sethi et al., 2016). It is followed by several research articles with the topic of comparison between plant and dairy milk, investigation of characteristics, and also development of particular plant milk.

Table 2. Top 5 publication on the topic of plant-based milk based on number of citations in Google Scholar database

Article	Author	Number of Citations
Plant-based milk alternatives an emerging segment of functional beverages: a review	(Sethi et al., 2016)	549
How well do plant-based alternatives fare nutritionally compared to cow's milk?	(Vanga & Raghavan, 2018)	278
Nutrient density and nutritional value of milk and plant-based milk alternatives	(Chalupa-Krebzdak et al., 2018)	194
Plant-based milk substitutes: Bioactive compounds, conventional and novel processes, bioavailability studies, and health effects	(Aydar et al., 2020)	155
Development of non-dairy milk alternative using soymilk and almond milk	(Kundu et al., 2018)	79

**Trends in Research of “Plant-Based Milk”**

In order to figure out the trends in research correlated to plant-based milk, VOSviewer was applied as a tool for clustering and mapping keywords of database collected. Key terms or keywords can bring information about the main content of an article and can also be used to determine search trends in a particular area. The simultaneous co-

occurrence of keywords can effectively reflect hot topic in specific areas and provide additional support for scientific research.

From the 485 papers obtained after data extraction, VOSviewer retrieved 3108 keywords or terms. Then, we set minimum number of occurrences of term to 5. As the result, of the 3108 terms found, 156 met threshold. Then, VOSviewer calculated the relevance score of each term. Due to default choice of VOSviewer, we selected not more than 60% of the most relevant terms. The terms *skimmed milk* and *traditional plant* had the highest relevance score, while the terms *addition* and *fruit* had the lowest relevance score. On the other hand, the terms *diet* and *intake* had the highest occurrence score, while the terms *traditional plant* and *health benefit* had the lowest occurrence score. As many as 71 terms or keywords were then visualized into network, overlay, and density visualization.

Based on co-occurrence of keywords, there were three visualizations provided by VOSviewer, they are network visualization, overlay visualization; and density visualization. The keyword co-occurrence network and overlay visualization map are shown in Figure 3 and 4. In these figures, the proportion area of the nodes represents keyword strength and the lines between the keywords represent co-occurrence. The shorter the distance of lines, the stronger relationship between keywords. And the thicker lines mean the more often those keywords appear together in publications. VOSviewer classified the keywords into seven clusters represented by different colour (Figure 3). The classification was based on the closeness of every keywords.

Cluster 1 consisted of 25 terms represented by red colour, they were *case, characterization, coconut, dairy milk, drink, fluid milk, formulation, health benefit, ice cream, ingredient, lactic acid bacterium, milk allergy, milk beverage, non-dairy milk, oat, optimization, physicochemical properties, seed, sensory attribute, soybean, soymilk, vegan, and sugar*. Cluster 1 were mainly consisted of terms from publication on development and characterization of plant-based milk (Bakiya et al., 2019).

Cluster 2 (green) consisted of 14 terms related to publication with respect to comparison between plant-based milk and animal-based milk, including the impact on infant (Fiocchi et al., 2022). The terms were *animal milk, animal protein, child, coconut milk, comparison, difference, infant, infant formula, milk powder, milk protein, novel plant, rice, rice milk, skim milk*. Cluster 3 (blue) is made up of terms in publication about the association of

plant-based milk consumption and diet (He et al., 2016). Terms in cluster 3 were *adult, association, diet, diet index, dietary pattern, health plant, intake, risk, whole food plant*. Cluster 4 represented by yellow consisted of 8 terms, they were *almond milk, cow milk, kefir, milk analogue, oat milk, oil, soy milk, yoghurt*. This cluster was found in publication with respect to non-dairy milk from nut and cereal and the derivative product (Kim et al., 2021).

Cluster 5 (purple) consisted of 7 items, they were *addition, bioactive compound, importance, life cycle assessment, prevention, research, treatment*. This cluster was found in publication related to the importance and impact of plant-based milk development on human body and environment (Winans et al., 2020). Cluster 6, with sky blue colour, highlight topics that were correlated to prospective new material for plant-based milk (Mukhtarovna et al., 2022). Terms in cluster 6 consisted of *butter, fruit, nut, popularity, skim milk powder, vegetable*. Cluster 7, represented by orange colour, were consisted of *skimmed milk and water*.

The overlay visualization can be used to show development and trends over time, including state-of-the-art of related research or novelty. The colour of overlay map indicates

different meaning to network map in Figure 3. Variations in node colour indicates variations in the year of publication, purple to yellow colour represents past to recent terms. Based on our study, recent terms on plant-based milk research were related to *oat milk, importance, life cycle assessment, rice milk, coconut, ice cream, sensory attribute, milk allergy, and whole food plant*. It means that novelty of plant-based milk research was correlated to importance of plant-based milk development and occurrence of milk allergy. Occurrence of lactose intolerance and cow's milk protein allergy affected the importance of plant-based milk (Ayah et al., 2022). In addition, accessibility, variation of raw material, and efficacy also decreased the importance of plant-based milk (Vandenplas et al., 2021). Moreover, the development of milk products derivatives (such as ice cream), especially in sensory attributes (Chung et al., 2022), based on whole food (Wicks & Wentzel-Viljoen, 2021), including oat; rice; and coconut, had novelty value. Recent study on plant-based milk was also connected to life cycle assessment or effect on environment by increasing efficiency and decreasing liability (Riofrio & Baykara, 2022; Winans et al., 2020).

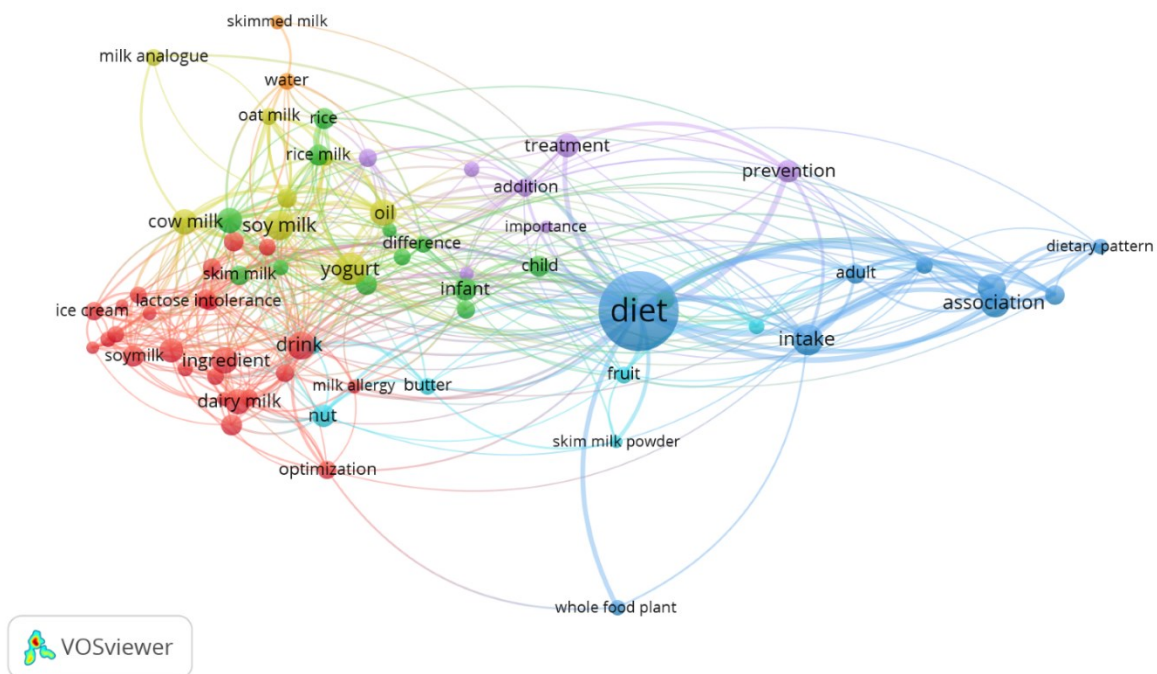


Figure 3. Network visualization map of keywords in the paper related to plant-based milk

Density visualization is performed by Figure 5. The density visualization presented a quick overview of the major areas from

bibliometric network. There is no line linked between keywords. However, every keyword is represented by different colour intensity. The

colour of a node in the visualization was determined by the density of the terms at that point. The terms were then converted to a colour value. The higher the density, the higher the colour value. The colour of a node was determined by matching the color value of the

node to the colour values and interpolating between the two colors. The term *diet* appeared as the highest density term due to highest occurrence score in publication between 2012 to 2022.

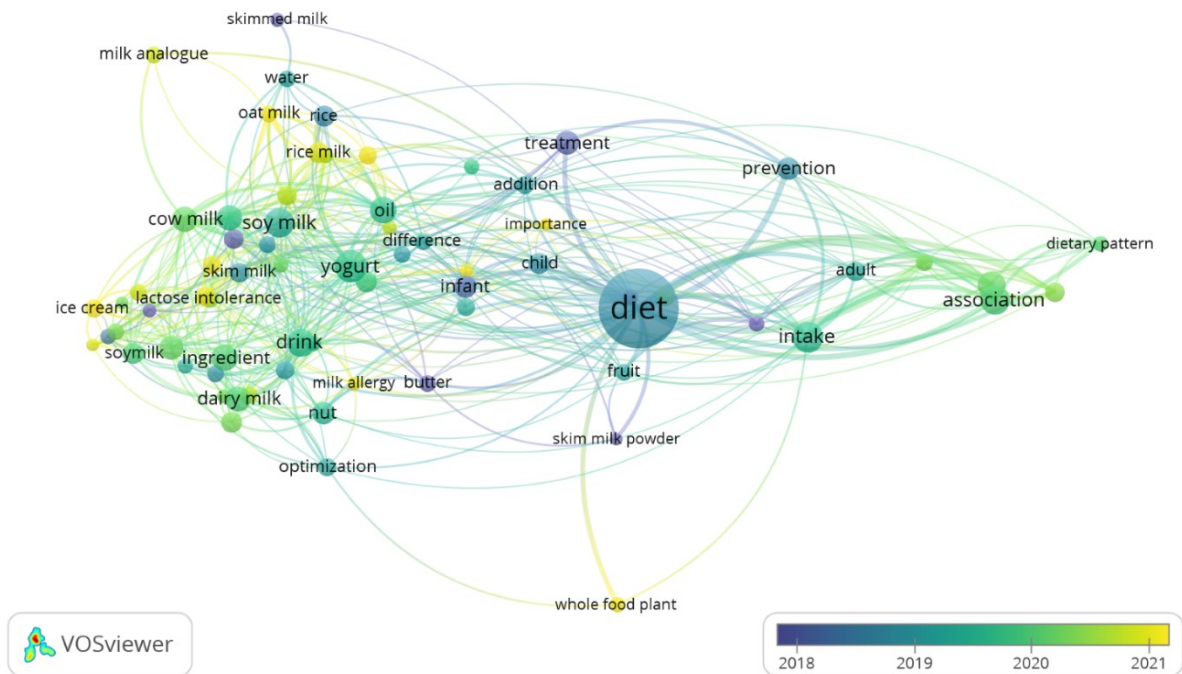


Figure 4. Overlay visualization map of keywords in the paper related to plant-based milk

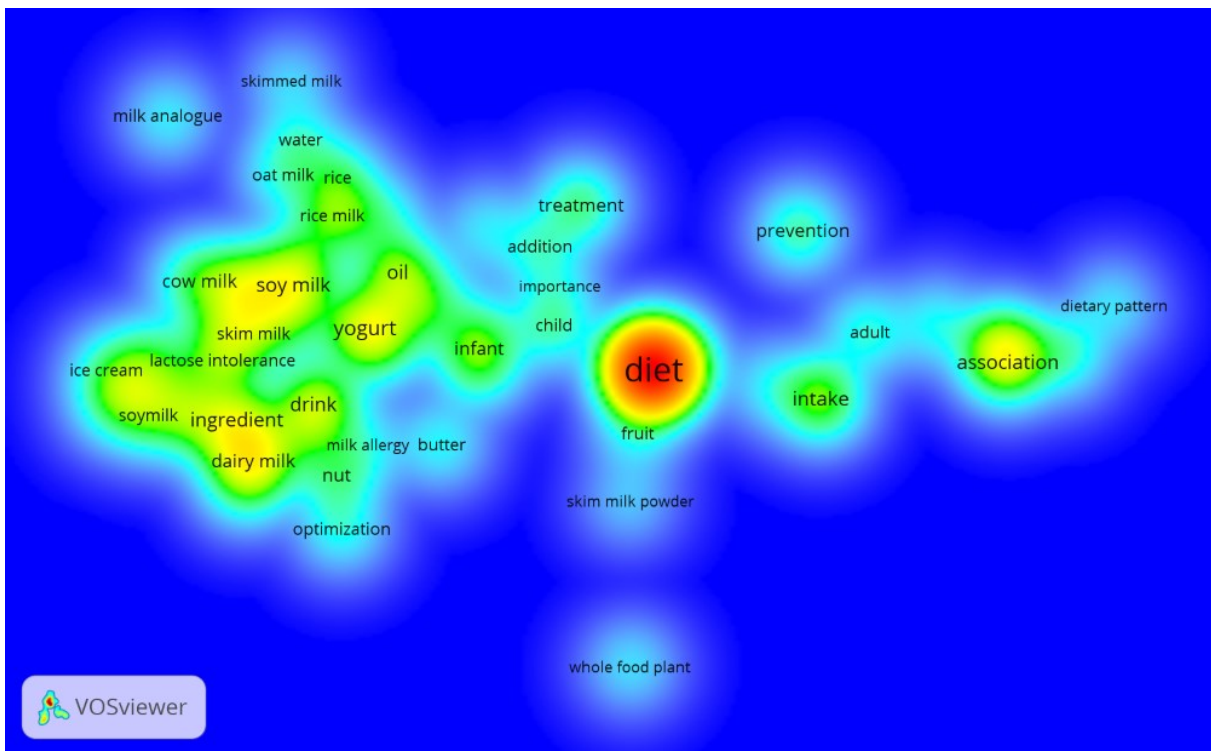


Figure 5. Density visualization map of keywords in the paper related to plant-based milk



However, the terms *traditional plant* and *health benefit* were hard to be seen on density visualization map due to the lowest occurrence score. It means that the gap of plant-based milk research was related to traditional plant and health benefit. There was inadequate information about health benefit of plant-based milk. Research in traditional plant or local plant as raw material of plant-based milk was also limited. Therefore, health benefit of plant-based milk and also utilization of traditional plant as raw material were potentially to be explored further.

### Limitation of Study

Our current study applied the information from Google Scholar, which is a helpful tool that support researchers locate relevant articles or papers in seconds from billions of archives on the web. However, it must be acknowledged that there is also a substantial amount of information from other sources and database, which is not included here. Constant alteration of content, algorithms, and structure, made Google Scholar to be a bad option for systematic review (Giustini & Boulos, 2013). In addition, our database is limited to articles provided in English, it is recognized that valid database is available in languages other than English.

Self-citations included in current review, which could create bias in the total citation, is another limitation of our research. In addition, since the subject matter of "plant-based milk" is vast and in-depth, it is not easy to ensure that all database is included.

### CONCLUSION

The research field of plant-based milk had been found since 1981. There is an increase of attention to the topic of plant-based milk in the last decade. We found the low number of publications per year from 1981 to 2011. Interestingly, the number of publications has gradually increased since then. Based on visualization map, we found cluster of the research field, novelty, recent study, and research gap.

Novelty of the research covered the importance and development of plant-based milk derivate. Recent study of plant-based milk also covered the correlation to life cycle assessment on environment and occurrence of milk allergy. Research on plant-based milk correlated to diet had been extensively studied. However, research correlated to traditional plant and health benefit of plant-based milk had not been studied comprehensively.

### REFERENCES

- Ayah, M., Ahmed, S. M., Sohair, F. N. (2022). Formulation and Evaluation of Some Novel Food Products Containing Plant (Vegan) Milks. In *Alexandria Journal ajfs.journals.ekb.eg*.  
[https://ajfs.journals.ekb.eg/article\\_217782\\_1326424dfecc3fcc59ee0404fd4a4c5f.pdf](https://ajfs.journals.ekb.eg/article_217782_1326424dfecc3fcc59ee0404fd4a4c5f.pdf)
- Aydar, E. F., Tutuncu, S., & Ozcelik, B. (2020). Plant-based milk substitutes: Bioactive compounds, conventional and novel processes, bioavailability studies, and health effects. In *Journal of Functional Foods*. Elsevier.  
<https://www.sciencedirect.com/science/article/pii/S1756464620301997>
- Bakiya, P., Arrivukkarasan, S., (2019). Optimization And Its Characterization Of Legumes Based Milk Chocolate To Enhance Its Folic Acid Content. *Think India Journal*.  
<https://thinkindiaquarterly.org/index.php/think-india/article/view/10975>
- Bekhit, A. E.-D. A., Ahmed, I. A. M., & Al-Juhaimi, F. Y. (2022). Non-bovine milk: Sources and future prospects. *Foods*, 11(13).  
<https://doi.org/10.3390/foods11131967>
- Chalupa-Krebzdak, S., Long, C. J., & Bohrer, B. M. (2018). Nutrient density and nutritional value of milk and plant-based milk alternatives. *International Dairy Journal*.  
<https://www.sciencedirect.com/science/article/pii/S0958694618301900>
- Chung, Y. L., Kuo, W. Y., Liou, B. K., Chen, P. C., (2022). Identifying sensory drivers of liking for plant-based milk coffees: Implications for product development and application. *Journal of Food ....*  
<https://doi.org/10.1111/1750-3841.16373>
- Fiocchi, A., Barrio-Torres, J., Dupont, C., (2022). Hydrolyzed rice formula for dietary management of infants with cow's milk allergy. *Organization Journal*. Elsevier.  
<https://www.sciencedirect.com/science/article/pii/S193945512200093X>
- Giustini, D., & Boulos, M. N. K. (2013). Google Scholar is not enough to be used alone for systematic reviews. *Online Journal of Public Health Informatics*, 5(2), 214.  
<https://doi.org/10.5210/ojphi.v5i2.4623>
- Haddow, G. (2018). Bibliometric research. In K. Williamson & G. Johanson (Eds.),

- Research methods: Information, systems, and contexts: Second Edition* (2nd ed., pp. 241–266). Chandos Publishing. <https://doi.org/10.1016/B978-0-08-102220-7.00010-8>
- He, J., Shen, X., Fang, A., Song, J., Li, H., Guo, M., (2016). Association between predominantly plant-based diets and iron status in Chinese adults: A cross-sectional analysis. *British Journal of Nutrition*. <https://www.cambridge.org/core/journals/british-journal-of-nutrition/article/association-between-predominantly-plantbased-diets-and-iron-status-in-chinese-adults-a-crosssectional-analysis/A8DF7A46BFCE4F939E681D65EEE6B58D>
- Kim, T. J., Seo, K. H., Chon, J. W., Youn, H. Y., (2021). Development of Plant-Based Milk Analogues as Alternatives to Cow Milk: Current Status and Future Prospects. *Journal of Dairy Science*. <https://www.koreascience.or.kr/article/JAKO202111261588259.view?orgId=anpor&hide=breadcrumb,journalinfo>
- Kundu, P., Dhankhar, J. K. A., (2018). Development of non dairy milk alternative using soymilk and almond milk. *Food Science Journal*. <http://www.foodandnutritionjournal.org/volume6number1/development-of-non-dairy-milk-alternative-using-soymilk-and-almond-milk/>
- McClements, D. J., Newman, E., & McClements, I. F. (2019). Plant-based milks: A review of the science underpinning their design, fabrication, and performance. *Comprehensive Reviews in Food Science and Food Safety*, 18(6), 2047–2067.
- Mukhtarovna, N. R., Hokimjon, A., (2022). Production of alternative vegetable milk. *Research Journal*. <https://wos.academiascience.org/index.php/wos/article/view/2067>
- Riofrio, A., & Baykara, H. (2022). A techno-environmental and life cycle assessment of 'oat-milk' production in Ecuador: a cradle-to-retail life cycle assessment. *International Journal of Food Science*. <https://doi.org/10.1111/ijfs.15828>
- Sethi, S., Tyagi, S. K., & Anurag, R. K. (2016). Plant-based milk alternatives an emerging segment of functional beverages: a review. *Journal of Food Science and Technology*. <https://doi.org/10.1007/s13197-016-2328-3>
- Silva, A. R. A., Silva, M. M. N., & Ribeiro, B. D. (2022). Plant-based milk products. In R. Bhat (Ed.), *Future foods: Global trends, opportunities, and sustainability challenges* (1st ed., pp. 233–249). Academic Press. <https://doi.org/10.1016/B978-0-323-91001-9.00025->
- Sweileh, W. M. (2020). Bibliometric analysis of scientific publications on “sustainable development goals” with emphasis on “good health and well-being” goal (2015–2019). *Globalization and Health*, 16(1), 68. <https://doi.org/10.1186/s12992-020-00602-2>
- Vandenplas, Y., Brough, H. A., Fiocchi, A., (2021). Current guidelines and future strategies for the management of cow's milk allergy. In *Journal of Asthma*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8548055/>
- Vanga, S. K., & Raghavan, V. (2018). How well do plant based alternatives fare nutritionally compared to cow's milk? *Journal of Food Science and Technology*. <https://doi.org/10.1007/s13197-017-2915-y>
- Wicks, M., & Wentzel-Viljoen, E. (2021). The whole-food plant-based diet: what does it entail and what lessons can it offer South African dietitian. *African Journal of Clinical Nutrition*. <https://doi.org/10.1080/16070658.2021.1943165>
- Winans, K. S., Macadam-Somer, I., Kendall, A., (2020). Life cycle assessment of California unsweetened almond milk. *Journal of Life Cycle*. Springer. <https://doi.org/10.1007/s11367-019-01716-5>