Journal of Natural Sciences and Mathematics Research
J. Nat. Scien. & Math. Res. Vol. 11 No. 2 (2025) 103-112
ISSN: 2460-4453 (Online); 2614-6487 (Printed)
Available online at http://journal.walisongo.ac.id/index.php/jnsmr

Kombucha origin clustering based on 16S metabarcoding datasets analysis

Imam Bagus Nugroho[1,2,*], Darmawan Ari Nugroho[1], Abdul Rahman Siregar[3]
[1] Bioindustry Lab., Department of Agro-industrial Technology, Faculty of Agricultural Technology, Universitas Gadjah Mada, Indonesia.
[2] Genetic Engineering Lab., Research Center for Biotechnology, Universitas Gadjah Mada, Indonesia.
[3] Microbiology Lab., Department of Tropical Biology, Faculty of Biology, Universitas Gadjah Mada, Indonesia.

* Corresponding author's e-mail: imam.bagus.n@ugm.ac.id

ABSTRACT
Consumers of fermented products increasingly demand detailed information on product ingredients, quality, health benefits, and origin. Herein, we have chosen kombucha as a model for a fermented product. This study aims to establish the origin information of kombucha using clustering analysis of 16S metabarcoding datasets. We have downloaded and analysed datasets of kombucha 16S metabarcoding originating from 5 distinct places: Brazil, the United States, the United Kingdom, Turkey, and Thailand. We randomly selected datasets from the collection (n = 32) and analyzed them on the SHAMAN server to develop an initial microbiome profile. We implemented hierarchical agglomerative Clustering and found that Ward's method and the Chao distance produced the best cluster tree, which consistently separates kombucha into five distinct clades, reflecting their origin. We have extended our examination to include more datasets (n=13) to build the final cluster tree (total n = 45). We have also assessed the uncertainty of the final Clustering by pvclust in R. The pvclust cluster tree is comparable in topology to the final cluster tree built using Ward's method and Chao distance. The pvclust cluster tree features stable clades that are highly supported by AU (Approximately Unbiased) values (p-value ≥ 95%). Each kombucha was also placed correctly and consistently according to its respective origin. We have successfully conducted analyses and demonstrated that a simple clustering method, combining Ward's method and the Chao distance, is the most effective for classifying kombucha by origin using a 16S metabarcoding dataset.

Keywords:
Bioinformatics, Bootstrapped Hierarchical Clustering, Fermented Product, Microbiome, Product Origin

Introduction

The development of big data analysis enables the capture of insights for industry use (Li et al., 2022; Dash et al., 2019), primarily focused on productivity, market sentiment, logistics, and supply chain management. The utilisation of insights derived from big data analysis includes, for example, improving manufacturing processes and yields, selecting marketing channels, and optimising product design and engineering. From the consumer perspective, information on product ingredients, bioactive compound contents, health benefits of product consumption, and product origin is essential before they decide to buy the products (Nazzaro et al., 2025). Modern consumers are more likely to be aware of health benefits and trust brands that consistently attribute their products appropriately (Baker et al., 2022).

The bioindustry, including the fermentation industry, has gained significant traction due to the recent advancements in biotechnology. Kombucha is a product of the fermentation industry that has recently gained popularity due to advertised health benefits (Martinez Leal et al., 2018). Kombucha fermentation produces various active biochemical compounds essential for the

health effects on the human body upon consumption (Júnior et al., 2022). The benefits of consuming kombucha are due to its bioactive compounds, such as polyphenols (Aung S Eun, 2022), which exhibit antioxidative, antiproliferative, and antidiabetic effects (Selvaraj S Gurumurthy, 2023; Dechakhamphu et al., 2023).

Kombucha is manufactured through the fermentation of tea and sugar using a symbiotic culture starter containing bacteria and yeast (SCOBY) (Laavanya et al., 2021; Xia et al., 2019). The bacterial community is typically dominated by acetic acid bacteria, such as *Komagataeibacter*, *Acetobacter*, and *Gluconacetobacter*, which influence the beverage's acidity (Kaashyap et al., 2021; Wang et al., 2022). Yeasts, including *Brettanomyces*, *Zygosaccharomyces*, *Saccharomyces*, and *Candida*, hydrolyse sugar and produce ethanol and other metabolites that feed the acetic acid bacteria and contribute to flavour (Landis et al., 2022; Wang et al, 2025). Metagenomic surveys reveal that, despite variations in raw materials and climate, kombucha systems tend to share a core microbiota dominated by *Komagataeibacter* and *Brettanomyces* or *Zygosaccharomyces*, highlighting both diversity and ecological stability in this fermented beverage (Wang et al., 2022; Wang et al., 2025).

As mentioned above, the diversity of microorganisms involved in kombucha fermentation can be studied using independent culture approaches, i.e., metabarcoding (Forsman et al., 2022) and metagenomic (Pérez-Cobas et al., 2020) analyses. The metabarcoding analysis involves a sequencing method that relies on DNA markers, such as the 16S rRNA gene for bacterial and archaeal communities or the ITS sequence for fungi (Patin et al., 2025). Meanwhile, metagenomics provides greater depth, as sequencing encompasses not just specific markers but also the entire DNA pool within the sample.

In this study, we focused on the metabarcoding datasets, particularly the 16S datasets of kombucha, which are readily available and sufficient to develop the bacterial diversity profile of kombucha. This study is inspired by the recognisably distinct microbial diversity within kombucha, as demonstrated by a previous study (Wang et al., 2025). We envision incorporating the bacterial diversity profile of kombucha into a machine learning approach in the future. Therefore, the classification or clustering process can be more straightforward, automated, and more accurate.

Methods

Datasets Collection

The datasets for developing the initial microbiome profile were retrieved from the public European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena/browser/home) using the keyword "kombucha". The datasets were collected as fastq.gz files and filtered to represent five places of origin, i.e., Brazil, the United States, the United Kingdom, Turkey, and Thailand. We were unable to find any datasets from Indonesia; therefore, we were unable to build a profile to include Indonesian kombucha. The retrieved datasets were split into a reference set (for developing the initial cluster tree) and a test set (for creating the final cluster tree), as shown in Table 1.

Metabarcoding Analysis and Visualisations

Metabarcoding analysis was done using the SHAMAN server (https://shaman.pasteur.fr) (Volant et al., 2020). The first analysis was conducted on the collected datasets (n=32) to build the initial microbiome profile of kombucha originating from five regions (the US, the UK, Brazil, Turkey, and Thailand). The profile was then downstream analysed using several techniques in the R environment within SHAMAN and aided by visualisations, namely Principal Component Analysis (PCA) (Jolliffe S Cadima, 2016), Principal Coordinate Analysis (PCoA) (Gower, 2014), Non-Metric Dimensional Scaling (NMDS) (Alotaibi, 2015) and Ward's Method of Hierarchical Agglomerative Clustering (Murtagh S Legendre, 2014). The analyses were combined using two distance matrices: Bray-Curtis (Ricotta S Podani, 2017) and Chao (Chiu S Chao, 2014) to construct the initial cluster tree. A second analysis included more data (total n = 45) and utilized Ward's Method and Chao's distance to build the final cluster tree.

Addressing Clustering Uncertainty Using Multi-Scale Bootstrap Resampling

Multi-scale bootstrap resampling analysis was used, as reported in previous studies (Shimodaira, 2002; Shimodaira, 2004; Suzuki & Shimodaira, 2006). This method measures the robustness and uncertainty in clustering analysis. We implemented a hierarchical clustering method.hclust" = "ward.D2" to apply Ward's Method as in the previous clustering analysis. Since pvclust does not support the Chao distance, we seek an alternative distance method or "method.dist" (i.e., "euclidean", "maximum", "manhattan", "canberra", "minkowski", or "binary") to reproduce a similar cluster tree topology as done using Ward's Method and Chao distance. We implemented bootstrap resampling ("nboot = 10000") and parallel computing ("parallel = TRUE") to speed up the calculation.

Results and Discussions

The Initial Microbiome Profile of Kombucha from Five Different Regions

The fermentation of kombucha utilises a microbial consortium consisting of bacteria and yeast (Laavanya et al., 2021). The microbial consortia of each kombucha can harbour diverse microorganisms with distinct characteristics (alpha and beta diversity indices and relative abundances of each taxon) (Landis et al., 2022). The study by Landis et al. (2022) also revealed that each kombucha from across the United States harbours subtle variations in microbial species. Therefore, we hypothesized that the diversity of microbial consortium revealed by metabarcoding analysis may correlate with the place of origin of kombucha. The metabarcoding profile can help determine the origin or geographic assignment of the kombucha.

Table 1. Datasets used in this study

Dataset Type	Sample Code	Accession Number	Place of Origin	Sample Code	Accession Number	Place of Origin
Initial	RefB1	SRR13280207	Brazil[3]	RefUS1	SRR15972179	US[1]
	RefB2	SRR13280229	Brazil	RefUS2	SRR15972180	US
	RefB3	SRR13280206	Brazil	RefUS3	SRR15972182	US
	RefB4	SRR13280228	Brazil	RefUS4	SRR15972177	US
	RefB5	SRR13280212	Brazil	RefUS5	SRR15972153	US
	RefB6	SRR13280231	Brazil	RefUS6	SRR15972155	US
	RefB7	SRR13280204	Brazil	RefUS7	SRR15972181	US
	RefB8	SRR13280205	Brazil	RefUS8	SRR15972157	US
	RefB9	SRR13280213	Brazil	RefUS9	SRR15972154	US
	RefB10	SRR13280230	Brazil	RefUS10	SRR15972156	US
	RefT1	SRR23314101	Thailand[5]	RefUS11	SRR15972178	US
	RefT2	SRR23314102	Thailand	RefTu1	SRR15017084	Turkey[4]
	RefUK1	SRR12833115	UK[2]	RefTu2	SRR15017086	Turkey
	RefUK2	SRR12833096	UK	RefUK5	SRR12833102	UK
	RefUK3	SRR12833100	UK	RefUK6	SRR12833097	UK
	RefUK4	SRR12833101	UK	RefUK7	SRR12833098	UK
Additional	B1	SRR13280208	Brazil	UK1	SRR12833099	UK
	B2	SRR13280214	Brazil	UK2	SRR12833103	UK
	B3	SRR13280215	Brazil	UK3	SRR12833114	UK
	B4	SRR13280232	Brazil	US1	SRR15972158	US
	B5	SRR13280233	Brazil	US2	SRR15972161	US
	US3	SRR15972172	US	US5	SRR15972184	US
	US4	SRR15972183	US			

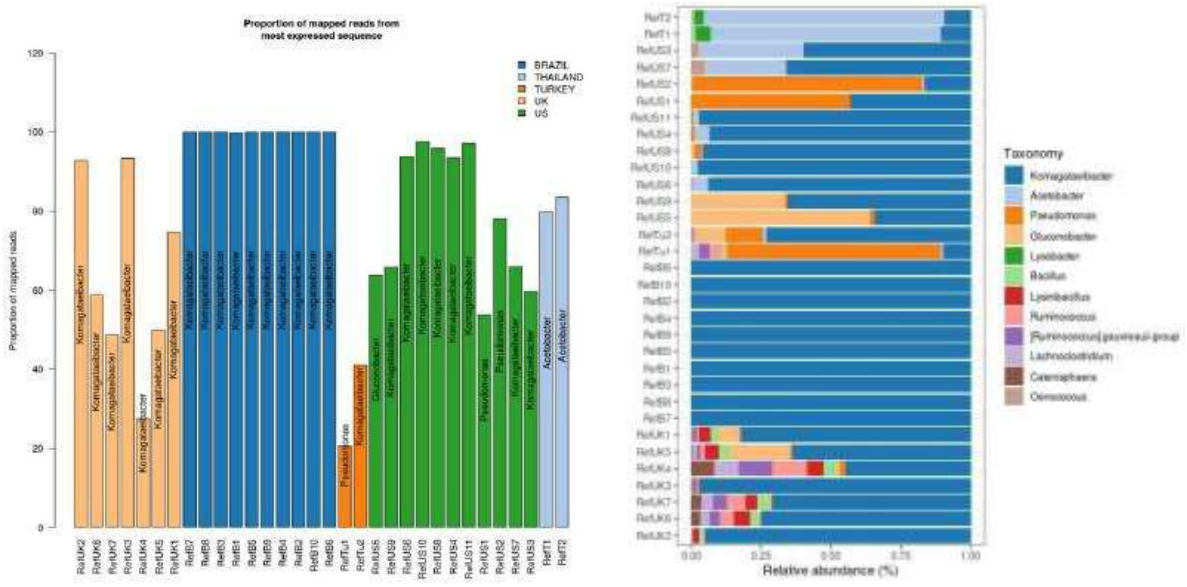
Notes: [1]US: The United States (PRJNA764354); [2]UK: The United Kingdom (PRJNA669148); [3]Brazil (PRJNA686871); [4]Turkey (PRJNA742890); and [5]Thailand (PRJNA930599).

The metabarcoding analysis revealed the diversity of microbes in kombucha from five regions, i.e., the US, UK, Brazil, Turkey, and Thailand. The metabarcoding analysis utilizes amplicon sequencing generated by Next-Generation Sequencing (NGS). The NGS sequencer produces short reads up to 500 bp. The sequencing process relies on the reads of the 16S rRNA

gene marker. The raw sequences generated by the NGS machine were processed using the SHAMAN-built metabarcoding workflow. First, raw reads were subjected to quality controls, including cleaning (removal) of low-quality sequences and sequencing adaptors, merging forward and reverse reads, and removing singletons and chimaeras. Cleaned and optimised sequences were binned and matched to the Silva database to build the taxonomy and count files.

Metabarcoding analysis of the initial datasets from five regions revealed a specific pattern of genera by kombucha origin (Figure 1(a)). The genus *Komagataeibacter* dominated kombucha samples from the UK (RefUK), the US (RefUS), Brazil (RefB), and one sample from Turkey. Genus *Acetobacter* dominated the kombucha from Thailand (RefT), while genus *Pseudomonas* dominated one of the samples from Turkey (RefTu). Additionally, *Gluconobacter* was a dominant genus in one of the US samples.

The microbial diversity of kombucha exhibited abundant, taxon-specific profiles (Figure 1(b)). Although almost all samples from different regions harboured *Komagataeibacter* as the most abundant taxon, several other taxa were also detected, including *Pseudomonas*, *Gluconobacter*, *Lysobacter*, *Bacillus*, *Lysinibacter*, *Ruminococcus*, *Lachnoclostridium*, *Catenisphaera*, and *Oenococcus*.



(a) (b)

Figure 1. Metabarcoding analysis results. (a) Primary taxonomy representing the most abundant taxon in the reference datasets (b) Microbial diversity of kombucha products from five regions showing the most abundant taxa.

Kombucha Classification Based on Its Origin Using Clustering

The source of microbial consortia in the starter cultures yielded distinct metabarcoding profiles for each kombucha originating from different sources. Presumably, kombucha from the UK, US, and Turkey utilised natural sources for starter culture, which have various species and abundant naturally occurring microbes. We employed several analytical techniques to characterise the metabarcoding profile of kombucha across these five regions. Principal Component Analysis (PCA; Figure 2) revealed consistent groupings among kombucha samples from the UK, Turkey, and Thailand. However, the kombuchas from the US and Brazil were grouped and could not be distinguished clearly.

We employed Principal Coordinate Analysis (PCoA) to further analyse and distinguish the unresolved groups (see Figure 2; cluster Brazil and US). We also performed a similar test using Non-Metric Dimensional Scaling (NMDS). PCoA and NMDS used two distance measures: Bray-Curtis and Chao. The results of PCoA and NMDS analyses showed distinct grouping patterns. The PCoA and NMDS, combined with the Bray-Curtis distance, grouped kombucha from the UK, US, and Brazil into overlapping clusters, while kombucha from Thailand and Turkey formed separate groups (Figures 3(a) and 3(b)). Meanwhile, the PCoA and NMDS, combined with the

Chao distance, resulted in more distinct groupings (Figures 4(a) and 4(b)).

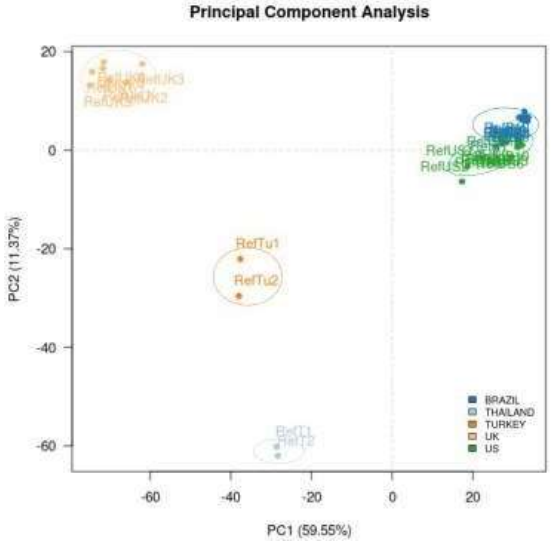
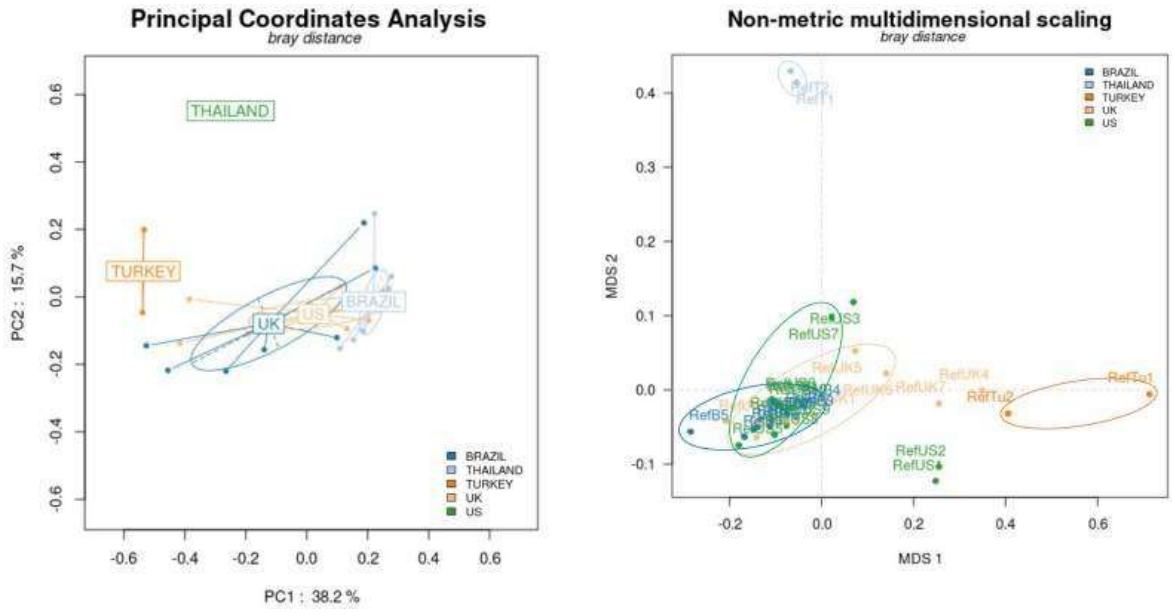


Figure 2. Dimensional reduction analysis using Principal Component Analysis.

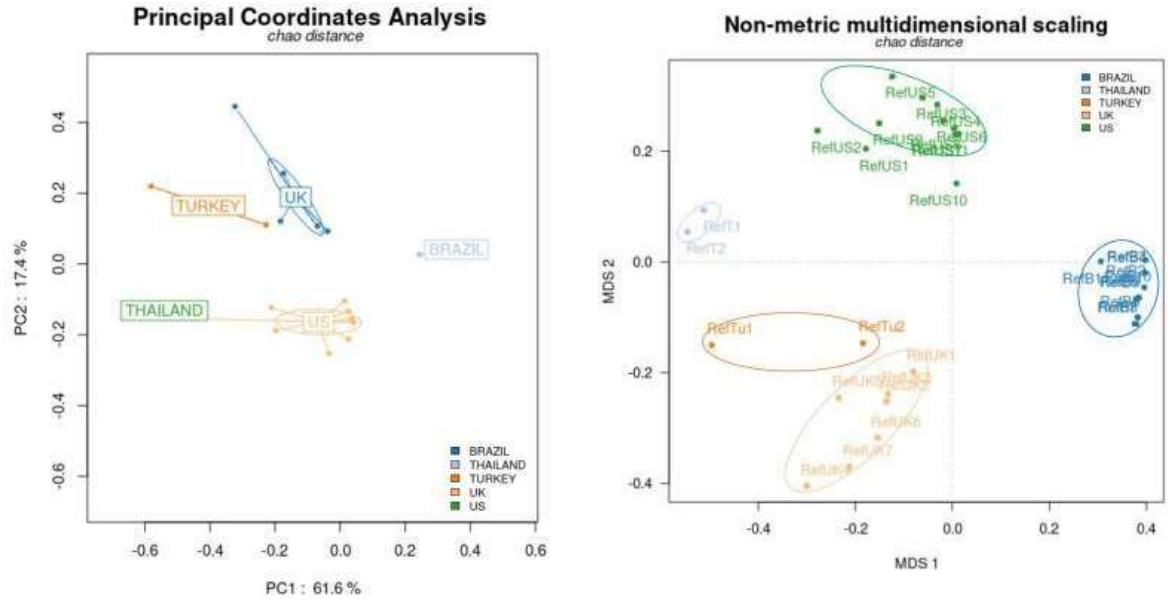


(a) (b)

Figure 3. Analysis using Bray-Curtis's distance. (a) Principal coordinate analysis. (b) Non-metric multidimensional scaling.

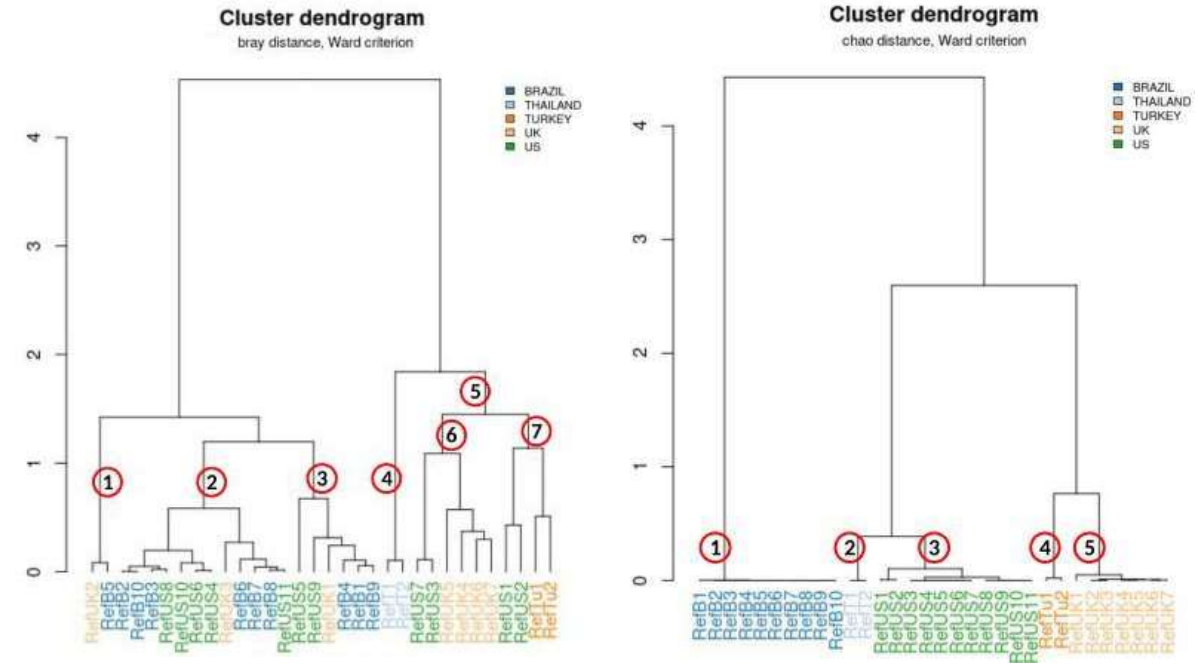
We have extended our examination of the initial microbiome profile using Hierarchical Agglomerative Clustering, namely Ward's Method, to determine whether clustering analysis can group and discriminate kombucha based on their origin. The clustering analysis also applied Bray-Curtis and Chao distances to ascertain their robustness in producing optimal clusters.

The clustering analysis produced distinct cluster profiles, as shown in the initial cluster tree (Figure 5). The cluster dendrogram created using Ward's Method in combination with Bray-Curtis revealed five distinct clades (branches) 1-5 with 2 (two) subclades (sub-branches) 6 and 7 from clade 5 (Figure 5(a)). Kombucha from Thailand (RefT1 and RefT2) clustered in clade 4, while kombuchas from Turkey (RefTu1 and RefTu2) were in subclade 6. Meanwhile, the other kombuchas from the US, UK, and Brazil (RefUS, RefUK, and RefB) were scattered across clades 1-6.



(a) (b)

Figure 4. Analysis using the Chao distance. (a) Principal coordinate analysis. (b) Non-metric multidimensional scaling.



(a) (b)

Figure 5. Initial cluster trees. (a) using Bray-Curtis's distance. (b) using the Chao distance.

The Ward's Method and the Chao distance yielded a more distinctive dendrogram. This analysis separated all the kombuchas from five different regions into five distinguished clades (branches) (Figure 5(b)). The kombuchas from Brazil, Thailand, the US, and Turkey were clustered in clades 1-5, respectively. This result demonstrated the robustness of Ward's Method and the Chao distance in constructing optimal Clustering. This finding also supports previous findings from PCA, PCoA, and NMDS analyses.

Next, we performed the metabarcoding analysis in SHAMAN to include more data (total data, n = 45) (Table 1). The previous optimal clustering method (Ward's Method combined with the Chao distance) enables further analysis of kombucha origin by combining the initial and additional datasets. We observed six clades formed in the final cluster tree (Figure 6). An

additional clade to the initial cluster tree (Figure 5(b)) was due to the split of formerly US clades into two subclades: US subclades 1 and 2. All five additional Brazilian datasets were in clade 1. All three UK datasets were in clade 6. Two US datasets clustered in Clade 2 (US Clade 1), while the other three clustered in Clade 4 (US Clade 2). Therefore, we have achieved an optimal and consistent cluster tree, placing each kombucha on its correct clades and representing its origin.

We recognised the limitations of clustering analysis using Ward's Method and the Chao distance, which resulted in a cluster tree without clade/branch/node support values. The constructed cluster tree may be unstable and exhibit a different topology if the datasets change or become more extensive. Therefore, we implemented multi-scale bootstrap resampling using the pvclust package in R (Shimodaira, 2002; Shimodaira, 2004; Suzuki and Shimodaira, 2006). We implemented a multi-scale bootstrap resampling to test whether a cluster tree will occur and to calculate the cluster tree's probability/confidence (p-value), with results harbouring similar characteristics across iterated resampling.

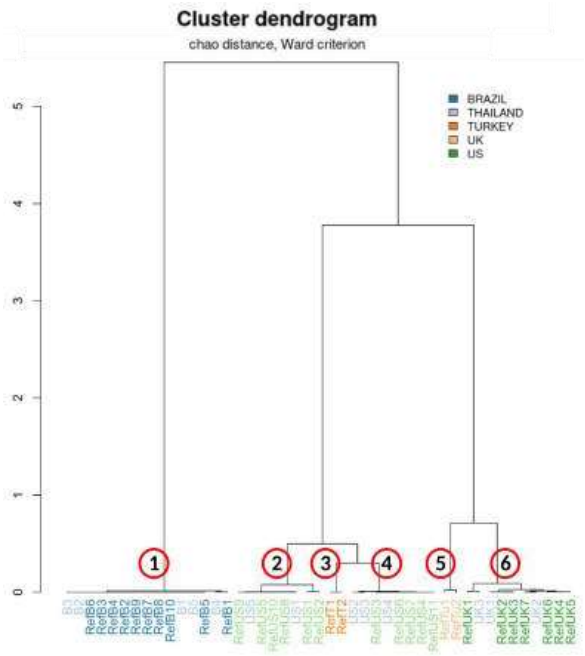


Figure 6. Final cluster tree. 1. Brazil clade; 2. US clade 1; 3. Thailand clade 3; 4. US clade 2; 5. Turkey clade and 6. UK clade.

We chose Ward.D2 as the input to the method. hclust in pvclust for hierarchical Clustering because it is the correct method that implements Ward's method (Murtagh & Legendre, 2014), as conducted in the previous analysis in this study. We also decided to apply the "binary" distance after running several combination distance metrics ("method.dist"= "euclidean", "maximum", "manhattan", "canberra", or "minkowski") with Ward's Method and found only "binary" distance produced similar cluster tree topology as seen in Figure 7 (other cluster trees were not shown). Figure 7 illustrates the consistent cluster tree, supported by multi-scale bootstrap values assigned to each node of the clades, as approximately unbiased (AU) values. All clades in Figure 7 had consistent members; for example, the US clade consisted only of US datasets, as well as four other clades formed by datasets originating from the UK, Brazil, Thailand, and Turkey. Moreover, all clades were supported by AU values (p-value ≥ 95%). Therefore, we can confidently confirm that all clades are stable and that the cluster tree is supported by statistical analysis.

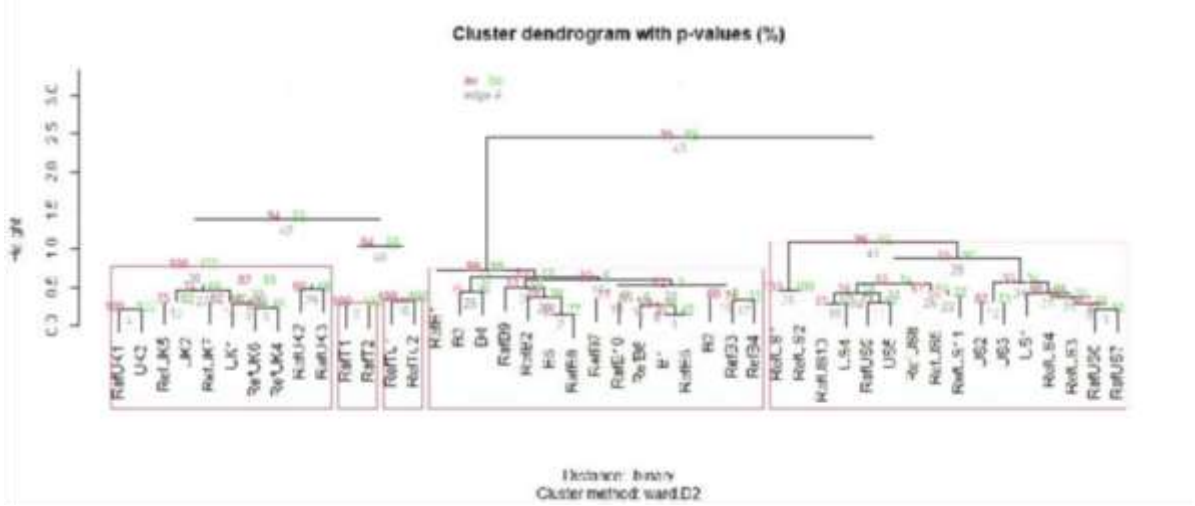


Figure 7. Cluster dendrogram after assessment using pvclust.

We have demonstrated a simple yet robust method for clustering kombucha according to its origin. We have successfully repurposed a collection of statistical and analytical techniques, as well as bioinformatic analysis, typically used in molecular ecology studies. This metabarcoding-based classification method has proven to be robust and straightforward for assigning a product origin label. This method also relies on metabarcoding data derived from the products' microbiome genetic background (DNA). Therefore, it is suitable for any products manufactured using fermentation or that utilise microbes for processing.

Conclusion

This study has successfully established a simple method for product origin assignation based on metabarcoding analysis. Among tested data analytical methods, the combination of Ward's Method and Chao distance is the most robust for grouping and clustering analysis. We have also assessed the uncertainty of the clustering analysis using multi-scale bootstrap resampling and confirmed the result with confidence (p-value ≥ 95%).

Acknowledgments

IBN: Designed the study, conducted data analysis, and wrote the initial draft of the manuscript. DAN and ARS: reviewed the manuscript. All authors agree to the final version of the manuscript. Universitas Gadjah Mada funded this study for IBN (Grant Number: 559/UN1.P.III/Dit-Lit/PT.01.03/2022 and 6530/UN1.P1/PT.01.03/2024).

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

Alotaibi, K. (2015). *Non-metric multidimensional scaling for distance-based privacy-preserving data mining* [Doctoral Thesis, University of East Anglia]. https://ueaeprints.uea.ac.uk/id/eprint/52228

Aung, T., S Eun, J.-B. (2022). Impact of time and temperature on the physicochemical, microbiological, and nutraceutical properties of laver kombucha (*Porphyra dentata*) during fermentation. *LWT*, *154*, 112643. https://doi.org/10.1016/j.lwt.2021.112643

Baker, M. T., Lu, P., Parrella, J. A., S Leggette, H. R. (2022). Consumer acceptance toward functional foods: A scoping review. *International Journal of Environmental Research and Public Health*, *19*(3), 1217. https://doi.org/10.3390/ijerph19031217

Chiu, C.-H., S Chao, A. (2014). Distance-based functional diversity measures and their decomposition: A framework based on Hill numbers. *PLoS ONE*, *9*(7), e100014. https://doi.org/10.1371/journal.pone.0100014

Dash, S., Shakyawar, S. K., Sharma, M., S Kaushik, S. (2019). Big data in healthcare: Management, analysis and future prospects. *Journal of Big Data*, *6*(1), 54. https://doi.org/10.1186/s40537-019-0217-0

Dechakhamphu, A., Wongchum, N., Chumroenphat, T., Tanomtong, A., Pinlaor, S., S Siriamornpun, S. (2023). In vitro and in vivo evaluation for antioxidant and antidiabetic properties of *Cyperus rotundus* L. Kombucha. *Foods*, *12*(22), 4059. https://doi.org/10.3390/foods12224059

Gower, J. C. (2014). Principal coordinates analysis. In R. S. Kenett, N. T. Longford, W. W. Piegorsch, S F. Ruggeri (Eds.), *Wiley StatsRef: Statistics Reference Online* (1st ed.). Wiley. https://doi.org/10.1002/9781118445112.stat05670

Jolliffe, I. T., S Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Júnior, J. C. D. S., Meireles Mafaldo, Í., De Lima Brito, I., S Tribuzy De Magalhães Cordeiro, A. M. (2022). Kombucha: Formulation, chemical composition, and therapeutic potentialities. *Current Research in Food Science*, *5*, 360–365. https://doi.org/10.1016/j.crfs.2022.01.023

Kaashyap, M., Cohen, M., S Mantri, N. (2021). Microbial diversity and characteristics of kombucha as revealed by metagenomic and physicochemical analysis. *Nutrients*, *13*(12), 4446. https://doi.org/10.3390/nu13124446

Laavanya, D., Shirkole, S., S Balasubramanian, P. (2021). Current challenges, applications and future perspectives of SCOBY cellulose of Kombucha fermentation. *Journal of Cleaner Production*, *295*, 126454. https://doi.org/10.1016/j.jclepro.2021.126454

Landis, E. A., Fogarty, E., Edwards, J. C., Popa, O., Eren, A. M., S Wolfe, B. E. (2022). Microbial diversity and interaction specificity in kombucha tea fermentations. *mSystems*, *7*(3), e00157-22. https://doi.org/10.1128/msystems.00157-22

Li, C., Chen, Y., S Shang, Y. (2022). A review of industrial big data for decision making in intelligent manufacturing. *Engineering Science and Technology, an International Journal*, *29*, 101021. https://doi.org/10.1016/j.jestch.2021.06.001

Martínez Leal, J., Valenzuela Suárez, L., Jayabalan, R., Huerta Oros, J., S Escalante-Aburto, A. (2018). A review on health benefits of kombucha nutritional compounds and metabolites. *CyTA - Journal of Food*, *16*(1), 390–399. https://doi.org/10.1080/19476337.2017.1410499

Murtagh, F., S Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion? *Journal of Classification*, *31*(3), 274–295. https://doi.org/10.1007/s00357-014-9161-z

Nazzaro, C., Uliano, A., Lerro, M., S Stanco, M. (2025). From claims to choices: How health information shapes consumer decisions in the functional food market. *Foods*, *14*(4), 699. https://doi.org/10.3390/foods14040699

Patin, N. V., Pitz, K. J., Baker, J. D., Chavez, F. P., S Goodwin, K. D. (2025). Markers or metagenomes: Sequencing marine eukaryotic DNA for better biodiversity surveys. *Metabarcoding and Metagenomics*, *9*, e144935. https://doi.org/10.3897/mbmg.9.144935

Pérez-Cobas, A. E., Gomez-Valero, L., S Buchrieser, C. (2020). Metagenomic approaches in microbial ecology: An update on whole-genome and marker gene sequencing analyses. *Microbial Genomics*, *6*(8). https://doi.org/10.1099/mgen.0.000409

Ricotta, C., S Podani, J. (2017). On some properties of the Bray-Curtis dissimilarity and their ecological meaning. *Ecological Complexity*, *31*, 201–205. https://doi.org/10.1016/j.ecocom.2017.07.003

Selvaraj, S., S Gurumurthy, K. (2023). An overview of probiotic health booster-kombucha tea. *Chinese Herbal Medicines*, *15*(1), 27–32. https://doi.org/10.1016/j.chmed.2022.06.010

Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. *Systematic Biology*, *51*(3), 492–508. https://doi.org/10.1080/10635150290069913

Shimodaira, H. (2004). Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. *The Annals of Statistics*, *32*(6). https://doi.org/10.1214/009053604000000823

Suzuki, R., S Shimodaira, H. (2006). Pvclust: An R package for assessing the uncertainty in hierarchical Clustering. *Bioinformatics*, *22*(12), 1540–1542. https://doi.org/10.1093/bioinformatics/btl117

Volant, S., Lechat, P., Woringer, P., Motreff, L., Campagne, P., Malabat, C., Kennedy, S., S Ghozlane, A. (2020). SHAMAN: A user-friendly website for metataxonomic analysis from raw reads to statistical analysis. *BMC Bioinformatics*, *21*(1), 345. https://doi.org/10.1186/s12859-020-03666-4

Xia, X., Dai, Y., Wu, H., Liu, X., Wang, Y., Yin, L., Wang, Z., Li, X., S Zhou, J. (2019). Kombucha fermentation enhances the health-promoting properties of soymilk beverage. *Journal of Functional Foods*, *62*, 103549. https://doi.org/10.1016/j.jff.2019.103549

Wang, W., Chen, Z., S Kuang, J. (2025). Artificial intelligence-driven recommendations and functional food purchases: Understanding consumer decision-making. *Foods*, *14*(6), 976. https://doi.org/10.3390/foods14060976

Wang, B., Rutherfurd-Markwick, K., Zhang, X.-X., S Mutukumira, A. N. (2022). Kombucha: Production and microbiological research. *Foods*, *11*(21), 3456. https://doi.org/10.3390/foods11213456