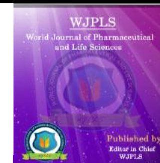
World Journal of Pharmaceutical and Life Sciences
WJPLS
www.wjpls.org
SJIF Impact Factor: 6.129



HAPLOTYPE ANALYSIS OF 16S RIBOSOMAL RNA ISOLATED FROM CUTTING BOARDS AT DIFFERENT BOVINE MEAT RETAIL STORES

Ayman Sabry*[1,2], Abdullah Abdulhafith Aljaddawi[3] and Salah Abo Aba[3]

[1]Biology Department, Faculty of Science, Taif University, Taif, KSA.
[2]Cell Biology Dept, Institute of Genetics and Biotechnology, National Research Centre, Dokki, Giza, Egypt.
[3]Biology Department, Faculty of Sciences King Abdulaziz University.

Corresponding Author: Ayman Sabry
Biology Department, Faculty of Science, Taif University, Taif, KSA.

Article Received on 05/09/2022 **Article Revised on 26/09/2022** **Article Accepted on 16/10/2022**

ABSTRACT

This work was carried out to present a comprehensive workflow of haplotype analysis using R for 16s rRNA sequences. Eleven bacterial samples were isolated from cutting boards at four different bovine meat retail stores. Phylogenetic analysis revealed presence of 8 different bacterial strains. *Macrococcus caseolyticus* was found in samples from the four locations. The length of sequences ranged from 1071 for *M.caseolyticus* in location VI to 1280 also for *M. caseolyticus* but in location II. The GC% content averaged 52.8, ranging from 51 to 55 % with no outlier. In this data set no common haplotype was noticed, meaning each haplotype was unique. The smallest distance between the haplotype pairs was 41 nucleotides between H7–H8, and 42 nucleotides between H1–H2, where the largest distance was 931 nucleotides between H10-H11. The haplotype difference (H_d) was 0.98 ($\pm$0.05) and the nucleotide difference (π) was 0.27. The average number of nucleotide differences (K) was 264 and the total variance of k (free recombination), V(k) was 106. Four conserved regions were detected. These 4 regions occupied the distance from nucleotide 701 to 1265. All 4 conserved regions were significant Based on results presented in this work, this haplotype analysis framework is found to be as efficient and straightforward.

KEYWORDS: Phylogenetic, Haplotype analysis, 16s rRNA, DNA sequence.

INTRODUCTION

Bovine meat (beef) is one of the main protein resources in human nutrition The nutritive benefits of bovine meat are countless (Thomson & Kotula, 1959; Bauchart et al., 2007; Udenigwe & Howard, 2013). Through the last 50 years, meat production has risen dramatically all over the world, and the whole production has more than quadrupled since 1961 (Weis, 2013). The world's largest meat producing continent is Asia, contributing about 40-45% of global meat production (Ritchie, 2019). The major producers of beef and buffalo meat are USA, Brazil, and China, followed by Argentina, Australia, and India. During the period form 1961 to 2018, the production of bovine meat increased to about 40,000 tons in KSA (Ritchie, 2019). Within time, the global demand for bovine meat is keeping increase. KSA is the largest consumer of bovine meat. Various meat products, (e.g. sausages, beef burgers, luncheons, etc.) are became very common food in KSA (Reyad et al., 2016). Therefore, research on bovine meat and meat products is significantly ensuring public health, and global as well as regional economy.

Meat could be contaminated through any point of the supply chain including production, packaging, transport, and storage (Rather et al., 2017). Accordingly, ensuring the safety of meat and meat products has been at the vanguard of public concerns. Many food-borne diseases are one of the major sources of anxiety and stress in human lives (Wilcock et al., 2004). Cutting boards - specially - are usually made of porous materials which made it less adequate for thorough cleaning and elevated opportunity of adherence of bacteria and establishment of resident microbiota (Faille & Carpentier, 2009; Ksontini et al., 2013; Stellato et al., 2015, 2016).

Several microorganisms could grow on meat and meat products given suitable condition for microorganisms proliferation. Various microorganisms including bacteria are responsible for meat deterioration, food poisoning, and carcass contamination that have drawback effects on economy of both farmers and marketing, not to mention meat quality. Multiple infectious diseases, namely, anthrax, clostridiosis, leptospirosis, Q-fever listeriosis, erysipelas, glanders,

brucellosis, salmonellosis, streptococcosis, and yersiniosis may develop as result of improper handling of meat and meat products. Microorganisms such as Brochothrix thermosphacta and Pseudomonas spp. can spoil meat and meat products if they grow to a certain extent (Koutsoumanis et al., 2006).

One definition of haplotype is an adjacent subset of data contained in DNA sequence (Hancock & Di-Rienzo, 2008). An illustration of a haplotype by this definition grounded on molecular observation, is actual inherited sequence (Glusman et al., 2014). With the advances in sequencing technology, haplotypes, that are spreading over one gene or more in a particular genomic region of interest not genes could be considered as the basic united element of inheritance (Lloyd et al., 2016), thus the usage DNA data in the conservation has converted from genetic conservation to genomic conservation (Allendorf et al., 2010; Funk et al., 2012). Elucidation of haplotype structure is one of many good categories of haplotype analysis (Hoehe, 2003). Moreover, comparative analysis of haplotype sequences could provide a great deal of information about within and between species genetic diversity (Toparslan et al., 2020). Genomic data analysis is usually time consuming and missy job, due to the fact it involves multiple software programs (i.e. MEGA, DnaSP, splitsTree, TASSEL and Arlequin, Excoffier et. al 2006, Librado et. al 2009). Each one of these software programs requires specified input and output file format. In this work we are presenting a comprehensive framework for 16s rRNA sequence, which requires only single input file. Therefore, the genomic data analysis job could be carried out in one single software program.

Current work was carried out to fulfill the following objectives: 1) To isolate various bacterial strains from different retail stores located in Makkah, Saudi Arabia. 2) To present a comparhesive workflow of haplotype analysis using R for 16s rRNA sequences.

MATERIALS AND METHODS

Collection of samples and isolation of bacteria

Bacterial strains were isolated from cutting boards of four different retail stores located in Makkah, Saudi Arabia. Bacterial isolates were grown on nutrient agar (NA) plates and incubated at 37 °C for 24 h. Purified bacterial strains were grown in nutrient broth (NB) and preserved in 20% glycerol solution at -20 °C for subsequent use.

Identification of bacteria

Biochemical characterization

In order to determine various biochemical characteristics of bacteria, gram staining, oxidase and catalase tests were performed according to Vincent and Humphrey (1970), Shekhawat et al. (1992), and Hayward (1960).

Molecular characterization

Bacterial genomic DNA was extracted using the QIAGEN Kit, Germany as described in manufacturer's instructions. The 16S rRNA gene was amplified using the universal primers, 27F (5´-AGAGTTTGATCCTGGCTCAG-3´) and 1492R (5´-AAGGAGGTGATCCAGCCGCA-3´). Amplified products were purified, and Sanger sequenced at Macrogen, South Korea. The sequences were then manually edited by SnapGene Viewer software version 3.3.3.

Bioinformatics analysis

All sequences from different locations were subjected to ncbi BLAST search tool http://blast.ncbi.nlm.nih.gov to detect non-chance sequence similarity. BLAST search was restricted to 16s ribosomal RNA sequences, where models (XM/XP) as well as uncultured/environmental samples were also filtered out, such that more reliable results would be attained. Each individual sequence was solely blast, where blast hit with the lowest expect-value (which indicate number of non-chance alignments) was picked. In order to ensure that blast outputs were governed by expected-value (aka e-value), Blast algorithm parameter was decreased such the expected thresh- old was set to more stringent value of 1e−6. Alignment of the sequence was carried out using msa package (Bodenhofer et al., 2015).

Exploratory data and phylogenetic analyses were carried out under R Project for Statistical Computing (R Core Team, 2022). Where Exploratory data analysis was done using Seqinr (Charif & Lobry, 2007) R package. Phylogenetic analysis was carried out by ape package (Paradis et al., 2004). Reconstruction of the phylogenetic tree was done using Neighbor joint method (Nei, 1987).

Pegas R package (Paradis 2010) was used to analyze the haplotype diversity (Hd), the average number of nucleotide differences, the nucleotide diversity (π). The polymorphic site (S), the singleton variable sites (SP), and the parsimony informative sites (PIP).

RESULTS AND DISCUSSION

A total of 11 bacterial isolates were attained from cutting boards out of 4 different retail stores located in Makkah city. These 11 bacterial isolates were subjected to BLAST query https://blast.ncbi.nlm.nih.gov/Blast.cgi. BLAST Query sequences were evaluated and compared with the database sequences.

The output of BLAST query for the 11 bacterial isolates is presented in Table (1). For this query, E-value was less than zero for all the 11 isolates, which indicated these alignments are not attained by chance, therefore, E-value excluded from table (1). The query coverage percentage almost reached 100% except for *Macrococcus caseolyticus* (M.cas) in the second location. The identity percentage ranged from 95 to 99%.

Eight different strains were isolates from cutting boards at the four studied locations (Table 1). M. cas was fond

in all 4 locations, no other strain was repeated in any of the four locations.

Sequences length and percentage of GC content are presented in Figure (1). The length of sequences ranged from 1071 for *M.caseolyticus* in location VI to 1280 also for *M. caseolyticus* but in location II. The GC% content averaged 52.8, ranged from 51 to 55 % with no outlier.

Frequencies of DNA bases of all 11 sequences isolated from cutting boards are presented in Figure (2). Some differences in base frequencies of thiamine and Adenine were noticed for *M. caseolyticus* in location I. The average content of T, A, C, and G was 21, 26, 23 and 30%. The total content of A+T was 47 % much lower than that of C+G 53%. However, percentage of GC content ranged from 51 for *Klebsiella pneumonia* and *Macrococcus caseolyticus* at location B to 55 for *Klebsiella pneumonia* at location I.

The 16s rRNA with length 1329 base pair belonging to 11 isolates was aligned with the ClustalW method using the msa package (Bodenhofer et al., 2015). Figure (3) shows the neighbor-joining phylogenetic tree of 11 isolates from cutting boards. The branch lengths of the tree have been colored to represent the genetic distance. A different color was set to each type of nucleotide. Thus, color changes on the plot have shown nucleotide differences among samples. The distance scale has shown 1% genetic variation per nucleotide substitution. Each row of the multiple sequence alignments corresponding to the names at the end of the tree represents a complete sequence.

Phylogenetic tree is composed of two huge clades. The first clade was made of 2 clusters, the first cluster comprised five different species from the four locations. The second cluster comprised *M.ase* from locations I and II, this cluster had the longest branch length. The second clade has the shortest branch length and comprised only 3 species form locations II, III and VI.

Haplotypes Extraction

Haplotype sequences were attained from the DNA matrix through comparing each row with rows and each column with columns. When nucleotides were compared, the same nucleotides between haplotypes were expressed with dots while different nucleotides were directly written. The number of haplotypes per species was extracted using the haplotypes package and short R commands. In this data set no common haplotype was noticed, meaning each haplotype was unique (i.e. 11 haplotype resulted from 11 isolates, Table 2).

Distance matrix of Haplotype and heat map

Haplotype distance matrix was estimated based on the haplotype sequence matrix (11x1094) using the dist.hamming() function. The distance matrix shows the difference in the number of nucleotides between the two haplotypes. In Table (3) the smallest difference between the haplotype pairs was 41 nucleotides between H7–H8, and 42 nucleotides between H1–H2, where the largest distance was 931 nucleotides between H10-H11.

A heat map was constructed with phylogenetic trees from the haplo- type sequences matrix using the heat map () function (R Core Team, 2022). The haplotype distance matrix was extracted using R code published by Toparslan et al. (2020) from the haplotype sequence matrix (11x1094). The matrix obtained was a symmetric version of Hamming distance matrix which is used to construct the heat map. The Heat Map is fully compatible with the haplotype distance matrix shown in Table (3).

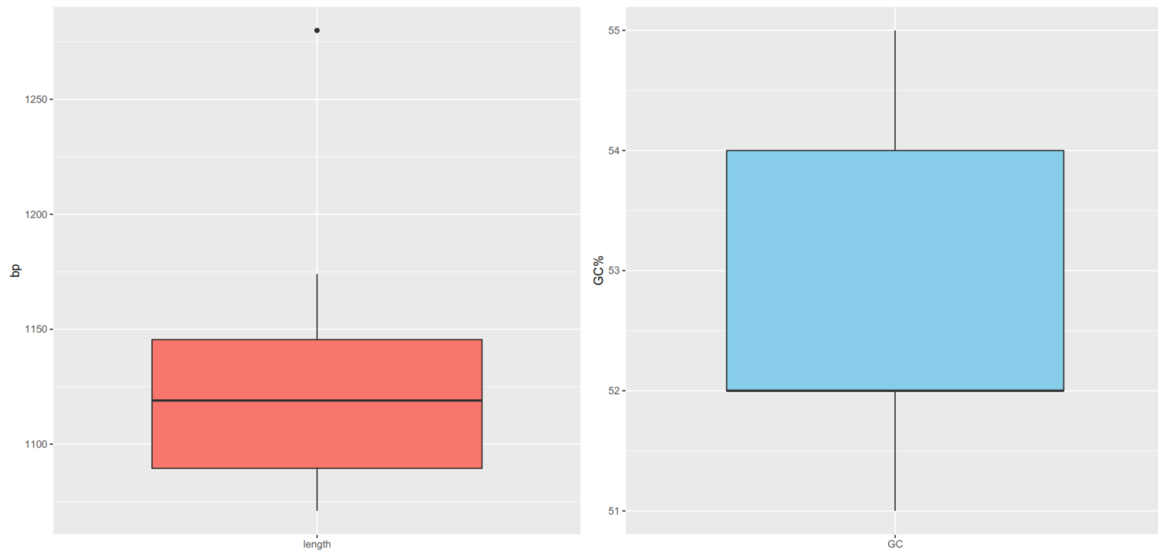
Reconstruction of phylogenetic tree was carried out using ggtree and ggplot2 packages. Tree calculation was estimated by the neighbor-joining method supported by ape package (Paradis et al., 2004).

The phylogenetic relationship between haplotype was demonstrated using the bootstrap method (Fig 6). The distance was estimated by the Hamming distance method of nucleotide differences between the two sequences (Paradis, 2010). The confidence interval was defined as strong for 85%.

Haplotype polymorphism & diversity cutting boards isolates, alignment of the 11 sequences resulted in 1324 sites, of which 337 sites were with gapes (Table 4). Invariable (monomorphic) sites were 212 sites. Variable (polymorphic) sites were 775 sites of which 482 singleton variable sites and 293 parsimony informative sites, and the total number of mutations was 1108.

Results of nucleotides diversity are show in table (6). The haplotype difference (H_d) was 0.98 (±0.05) and the nucleotide difference (π) was 0.27. The average number of nucleotide differences (K) was 264 and the total variance of k (free recombination), V(k) was 106.

For the 11 isolates 4 conserved regions were detected (table 7). These 4 regions occupied the distance from nucleotide 701 to 1265. For all the 4 conserved regions the value of heterozygosity was almost 1, with all the 4 regions were significant.



(a) Sequences lengths **(b) GC content**

Figure 1: Box plot of 16s Sequences length (a) and percentages of GC content isolated (b)

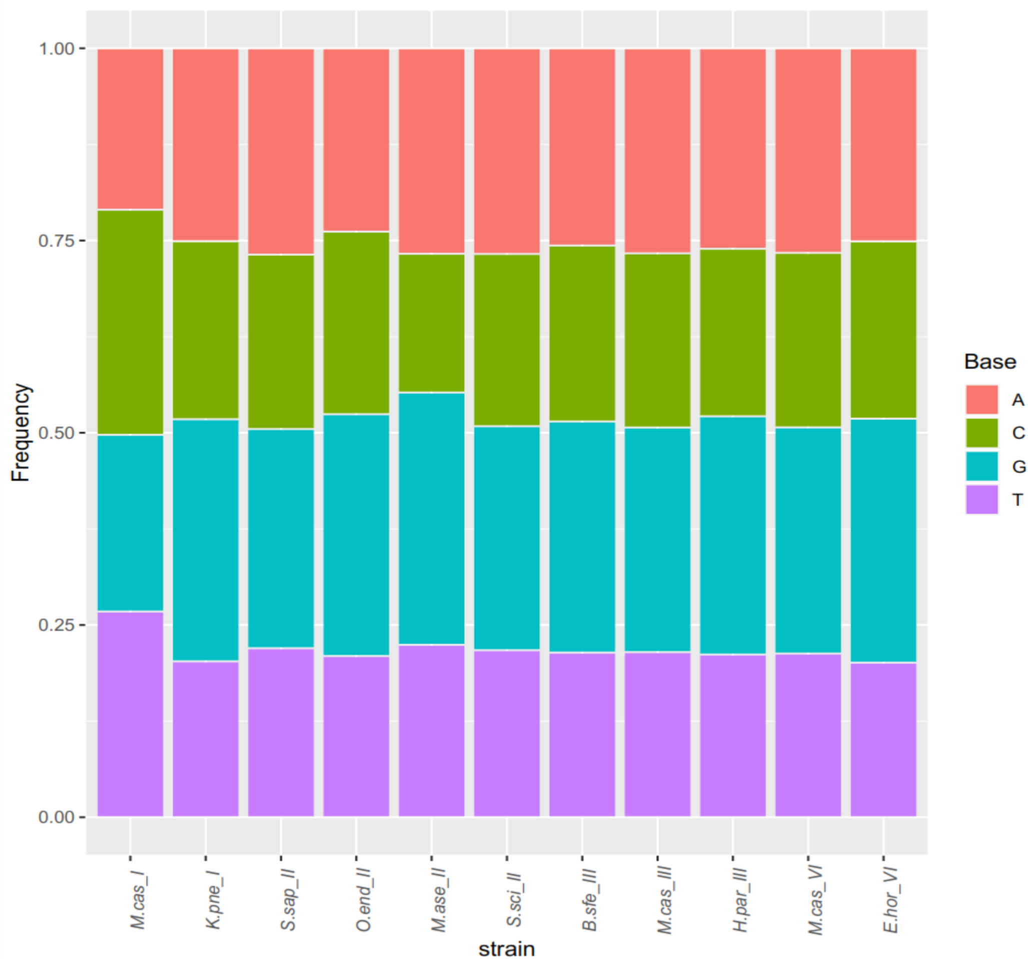


Figure 2: Base Frequencies of bacrtia strains collected from cutting boards, (latin number is an indicative of the location).

0.1

Figure 3: Neighbor-joining phylogenetic tree of 11 isolates from cutting boards, Colors both on the phylogenetic tree and on the branch scale represent genetic distance. (Latin numbers is an indicative of the location). Colors both on the phylogenetic tree and on the branch, scale represent genetic distance. Sequences have been colored by four different colors shown in the 'seq' column for each nucleotide. Color changes on the aligned sequences represent nucleotide differences.

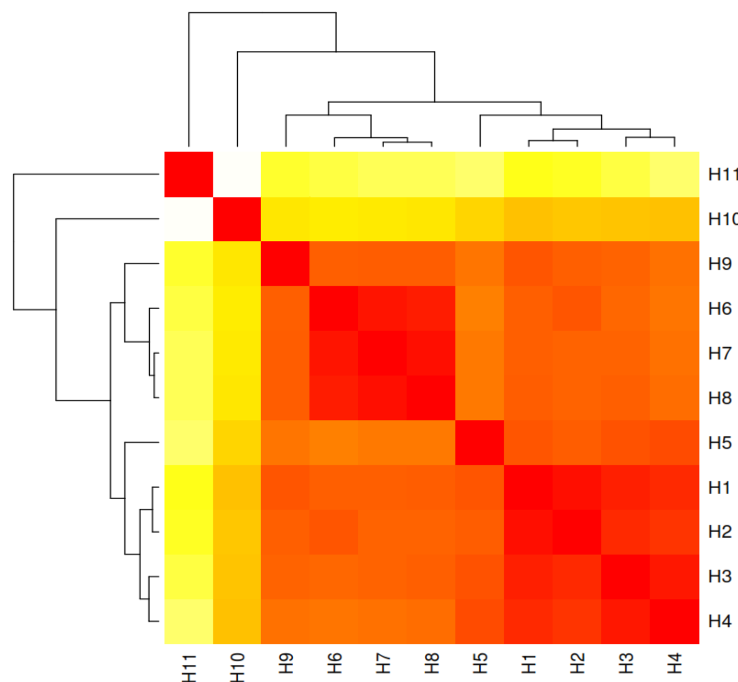


Figure 4: Heat map on the bases of number of nucleotide differences between the haplotypes. Every branch of phylogenetic tree stands for the corresponding haplotype in the matrix. The darker the color the closest relationships and vice versa.

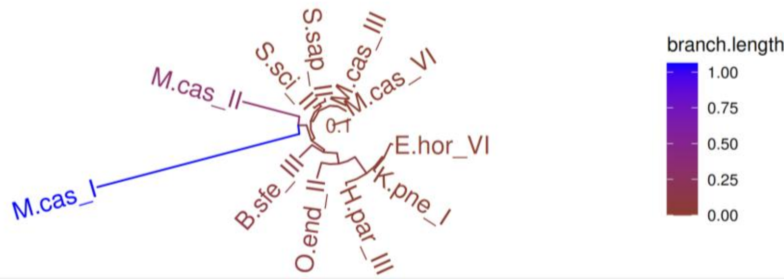


Figure 5: Colorized phylogenetic tree representing strains. The distance scale has shown 1% genetic variation per nucleotide substitution.

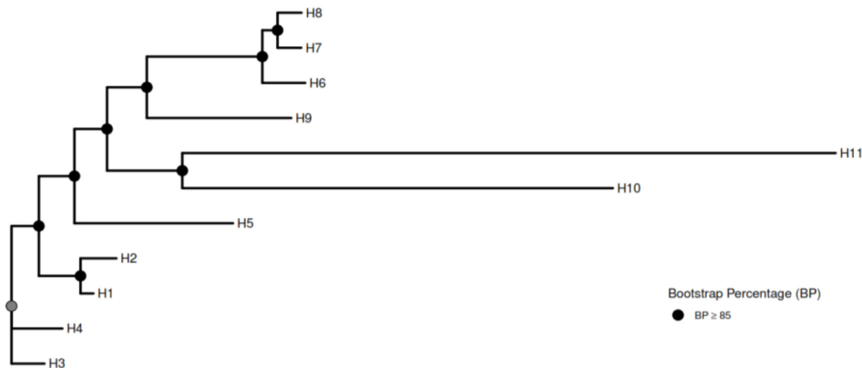


Figure 6: Neighbor-joining (NJ) tree for 16s haplotypes isolated from cutting boards. Colored internal nodes represent the bootstrap confidence level).

Table 1: BLAST quarry results for isolates of 16s ribosomal RNA sequence isolated from cutting boards in 4 different locations.

Location	Isolate	Strain	% Query coverage	% Identity
I	1	*Macrococcus caseolyticus (M.cas)*	100	99
	2	*Klebsiella pneumonia (K.pne)*	100	99
II	1	*Staphylococcus saprophyticus (S.sap)*	99	99
	2	*Ochrobactrum endophyticum (O.end)*	99	97
	3	*Macrococcus caseolyticus (M.cas)*	48	95
	4	*Staphylococcus sciuri (S.sci)*	100	99
III	1	*Bacillus sfensis (B.sfe)*	99	99
	2	*Macrococcus caseolyticus (M.cas)*	99	99
	3	*Hafnia paralvei (H. par)*	99	99
VI	1	*Macrococcus caseolyticus (M.cas)*	100	99
	2	*Enterobacter hormaechei (E.hor)*	99	99

Table 2: Number of haplotypes per species.

Haplotype	B.sfe	E.hor	H.par	K.pne	M.ase	M.cas	O.end	S.sap	S.sci
1	0	0	0	0	0	1	0	0	0
2	0	0	0	0	0	1	0	0	0
3	0	0	0	0	0	0	0	1	0
4	0	0	0	0	0	0	0	0	1
5	1	0	0	0	0	0	0	0	0
6	0	0	0	1	0	0	0	0	0
7	0	1	0	0	0	0	0	0	0
8	0	0	1	0	0	0	0	0	0
9	0	0	0	0	0	0	1	0	0
10	0	0	0	0	1	0	0	0	0
11	0	0	0	0	0	1	0	0	0

Table 3: Haplotype distance matrix using Hamming distance method.

	H1	H2	H3	H4	H5	H6	H7	H8	H9 H10
H2	42								
H3	90	116							
H4	117	142	72						
H5	236	252	224	208					
H6	263	235	286	322	350				
H7	263	277	275	308	326	61			
H8	260	274	269	302	330	80	41		
H9	235	262	274	312	323	262	252	258	
H10	522	545	537	529	582	648	634	624	625
H11	723	728	761	795	800	761	774	775	743 931

Table 4: Number of sites, monomorphic, polymorphic sites, and number of mutations of cutting boards.

Source	Number of sites	Monomorphic Sites	Polymorphic sites		No mutations
			Singleton	Parsimony informative	
Cutting Boards	1324	212	482	293	1108

Table 5: Number of haplotypes, haplotypes diversity and Standard deviation (SD) of haplotypes diversity for sequences isolated from cutting boards.

Source	No. of Haplotypes(H)	Haplotype(gene) diversity, (H_d)	SD Haplotype diversity
Cutting boards	11	0.98	0.05

Table 6: Nucleotide diversity, Average number of nucleotide differences (k), total variance of nucleotide differences (free mutations) sequences isolated from cutting boards.

Source	Nucleotide diversity (π)	Average no. of nucleotide differences (K)	Total variance of k (free recombination), V(k)
Cutting boards	0.27	264	106

Table 7: Conserved Regions, Conservation, Homozygosity and P-values of 16S RNA sequences isolated from cutting boards.

Start-End	Conservation	Homozygosity	P-value
701–841	0.32	0.9	0.002
870–1039	0.31	0.8	0.0001
1032–1094	0.32	1.0	0.02
1116–1265	0.34	1.0	0.0001

CONCLUSION

To the best of our knowledge, this is the first work to use comparative haplotype analysis to identify bacterial isolates based on their 16S rRNA gene which isolated from meat cutting boards in retails stores. Phylogenetic analysis of the 11 bacterial isolates revealed presence of 8 different bacterial strains. The *Macrococcus caseolyticus* was found in samples from the four studies locations.

REFERENCES

1. Allendorf, F. W., Hohenlohe, P.A., & Luikart, G. Genomics and the future of conservation genetics. Nature Reviews Genetics, 2010; 11(10): 697–709.
2. Bauchart, Caroline, Morzel, Martine, Chambon, Christophe, Mirand, Philippe Patureau, Reyne`s, Christelle, Buffie`re, Caroline, & Re´mond, Didier. Peptides reproducibly released by in vivo digestion of beef meat and trout flesh in pigs. British Journal of Nutrition, 2007; 98(6): 1187–1195.
3. Bodenhofer, U., Bonatesta, E., Horejs-Kainrath, C., & Hochre- iter, S. msa: an R package for multiple sequence alignment. Bioin- formatics, 2015; 31(24): 3997–3999.
4. Charif, D., & Lobry, R. J. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis, 2007; 207–232. of: Bastolla, U., Porto, M., Roman, H.E., & Vendruscolo, M. (eds), Structural approaches to sequence evolution: Molecules, networks, populations. Biological and Medical Physics, Biomedical Engineering. New York: Springer Verlag. ISBN: 978-3-540-35305-8.
5. Excoffier L, Heckel G. Computer programs for population genetics data analysis: a survival guide.

Nat Rev Genet. 2006; 7(10): 745–758. https://doi.org/10.1038/nrg1904 PMID: 16924258.
6. Funk, W. Chris, McKay, John K., Hohenlohe, Paul A., & Allen- dorf, Fred W. Harnessing genomics for delineating conservation units. Trends in Ecology & Evolution, 2012; 27(9): 489–496.
7. Glusman, G., Cox, H.C., & Roach, J.C. Whole-genome hap- lotyping approaches and genomic medicine. Genome Medicine, 2014; 6: 73.
8. Hancock, A. M, & Di-Rienzo, A. Detecting the Genetic Signature of Natural Selection in Human Populations: Models, Methods, and Data. Annu Rev Anthropol, 2008; 37: 197–217.
9. Hayward, A.C. A method for characterizing Pseudomonas solanacearum. Nature, 1960; 186: 405–406.
10. Hoehe, M. R. Haplotypes and the systematic analysis of genetic variation in genes and genomes. Pharmacogenomics, 2003; 4(5): 547–570.
11. Koutsoumanis, K., Stamatiou, A., Skandamis, P., & Nychas, G.J. Development of a microbial model for the combined effect of tem- perature and pH on spoilage of ground meat, and validation of the model under dynamic temperature conditions. Appl Environ Microbiol., 2006; 72(1): 124–134.
12. Ksontini, H., Kachouri, F., & Hamdi, M. Dairy biofilm: impact of microbial community on raw milk quality. J Food Qual., 2013; 34: 282–290.
13. Librado, P., & Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics, 2009; 25:, 25, 1451–1452.
14. Lloyd, S.S., Steele, E.J., & Dawkins, R. L. Analysis of Hap- lotype Sequences. Chap. 12 of: Kulski, Jerzy K (ed), Next Generation Sequencing. Rijeka: IntechOpen, 2016.
15. Nei, M. 1987. Molecular Evolutionary Genetics. New York: Columbia Univ. Press.
16. Paradis, E. pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics, 2010; 26: 419–420.
17. Paradis, E., Claude, J., & Strimmer, J. APE: analyses of phylogenetics and evolution in R language. Bioinformatics, 2004; 20: 289–290.
18. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2022.
19. Rather, I.A., Koh, W.Y., Paek, W.K., & Lim, J. The Sources of Chemical Contaminants in Food and Their Health Implications. Front. Pharmacol, 2017; 8. doi: 10.3389/fphar.2017.00830.
20. Reyad, R., Shawish, N. A., & Al-Humam. Contamination of beef products with staphylococcal classical enterotoxins in Egypt and Saudi Arabia. Kontamination von Rindfleischerzeugnissen mit klassischem, 2016; 1–6.
21. Ritchie, H. 2019. Meat and Dairy Production. Published online at Our- WorldInData.org. ’https://ourworldindata.org/meat-production’.
22. Shekhawat, G.S., S.K., Chakrabarti, & A.V., Gadevar. 1992. Potato Bacterial Wilt in India. India:. Tech. rept. Central Potato Re- search Institute.
23. Stellato, G., LaStoria, A., Cirillo, T., & D., Ercolini. Bac- terial biogeographical patterns in a cooking centre for hospital foodservice. In.t J. Food Microbiol, 2015; 193: 99–108.
24. Stellato, G, LaStoria, A., DeFilippis, F, Borriello, G., Villani, F., & Ercolini, D. Overlap of spoilage-associated microbiota between meat and the meat processing environment in small-scale and large-scale retail distributions. Appl Environ Microbiol, 2016; 82: 4045–4054.
25. Thomson, J.E., & Kotula, A.W. Contamination of the air sac areas of chicken carcasses and its relationship to scalding and method of killing. Poultry Science, 1959; 38: 1433-1437.
26. Toparslan, E., K., Karabag, & Bilge, U. A workflow with R: Phylogenetic analyses and visualizations using mitochondrial cytochrome b gene sequences. PLoS ONE, 2020; 15(12): e0243927.
27. Udenigwe, C.C., & Howard, A. Meat proteome as source of functional biopeptides. Food Research International, 2013; 54: 1021–1032.
28. Vincent, J.M., & Humphrey, B. Taxonimically significant group antigens in rhizobium. Journal of General Microbiology, 1970; 63: 379–382.
29. Weis, T. The meat of the global food crisis. Journal of Peasant Studies, 2013; 40: 65–85.
30. Wilcock, A., Pun, M., Khanona, J., & Aung, M. Consumer attitudes, knowledge and behaviour: a review of food safety issues. Trends in Food Science & Technology, 2004; 15(2): 56–66.