



Universidad Nacional Autónoma de México
Facultad de Ciencias
Genómica computacional
“Estructura genética poblacional de aislados de
***Listeria monocytogenes* y su asociación con la**
fuelle de alimento”



Casarin Perea Rodrigo; Cureño Sanchez Misael; Gonzalez Servín Luis Israel; Juárez-García, G. Lizbeth.

Resumen

Los patógenos transmitidos por los alimentos son de relevancia para la salud humana por la cantidad de enfermedades y muertes que causan. *Listeria monocytogenes* es una de las bacterias de mayor preocupación debido a su propagación, resistencia a condiciones como temperatura, pH y salinidad, así como los padecimientos que origina. El propósito de este trabajo es conocer la estructura poblacional de aislados de *L. monocytogenes* de diversos alimentos y determinar si hay una relación entre poblaciones genéticamente distintas y la fuente de alimento de la que provienen. Se utilizaron 504 aislados de *L. monocytogenes* provenientes de 19 países y de alimentos clasificados en Frutas/Vegetales, Mariscos, Carne, Quesos y Otros lácteos. Mediante PCA se identificaron 10 poblaciones genéticamente distintas. Se encontró que la población 2 está sobrerrepresentada en Quesos y Frutas/Vegetales, las poblaciones 5, 8 y 9 en Mariscos, y la población 10 en Carne. Mientras que, las poblaciones 1, 6 y 10 no se observaron en Frutas/Vegetales ni las poblaciones 3, 5 y 8 en Queso.

Palabras clave: Genómica, PCA, patógeno, transmisión, listeriosis.

Introducción

Los patógenos transmitidos por los alimentos están causando un gran número de enfermedades con efectos significativos en la salud humana. Los alimentos contaminados causan 600 millones de casos de enfermedades y 420 mil muertes al año alrededor del mundo. Estos patógenos pueden ser hongos, virus, pero principalmente bacterias, siendo *Listeria monocytogenes* de las más comunes junto con *Escherichia coli*, *Salmonella spp.*, *Shigella spp.*, *Clostridium spp.*, *Campylobacter jejuni*, *Cronobacter sakazakii* y *Staphylococcus aureus* (Bintsis, 2017).

L. monocytogenes es una bacteria grampositiva que tiene 13 serotipos, incluidos 1/2a, 1/2b, 1/2c, 3a, 3b, 3c, 4a, 4ab, 4b, 4c, 4d, 4e y 7; entre ellos, los serotipos 1/2a, 1/2b y 4b se han asociado con la gran mayoría de las infecciones transmitidas por alimentos. La longitud total media del genoma es de 2,974 Mb. Es un patógeno intracelular invasivo que depende de muchos factores de adhesión e invasión que facilitan la colonización del tracto gastrointestinal (TGI) y el tránsito a través de la barrera intestinal. Además, otros factores de virulencia, como la citolisina (listeriolisina O), la proteína de polimerización de actina ActA y las fosfolipasas, son importantes para la supervivencia intracelular del patógeno y la propagación de la infección (Sibanda y Buys, 2022).

Puede causar listeriosis, una infección grave, relativamente rara, pero con una alta tasa de mortalidad (20-30 %) que tiene algunos síntomas similares a una gripe, como dolor de cabeza, rigidez en el cuello, fiebre, dolores musculares, y además confusión, pérdida del equilibrio. Además, puede causar meningitis (inflamación de la membrana que rodea la médula espinal y el cerebro), gastroenteritis (inflamación de las membranas mucosas del estómago y el intestino) y septicemia

(propagación sistémica de bacterias y toxinas en la sangre) como resultado de una infección (Bintsis, 2017).

Se encuentra en ambientes como materia vegetal en descomposición, aguas residuales, agua y suelo, y puede sobrevivir a temperaturas extremas (1 a 45 °C) y concentraciones de sal. Es portado asintóticamente por numerosas especies animales. Además es psicrotrofo, es decir, capaz de crecer en condiciones de refrigeración o temperaturas inferiores a 10 °C. Su capacidad para formar biopelículas, permite que contamine y persista en una variedad de sustratos. Puede propagarse rápidamente en edificios húmedos, goteando de tuberías o techos sobre los alimentos. Una vez que la bacteria ingresa a una fábrica de procesamiento de alimentos, puede vivir allí durante años, contaminando a veces los productos alimenticios. Se ha encontrado en una variedad de alimentos productos lácteos (queso, leche pasteurizada), mariscos (camarones, pescados ahumados), embutidos, aves de corral, carne (chicharrones finos, lenguas de cerdo, jamón de pavo) y vegetales (maíz, arroz, col, papa, etc.) (Bechtel y Gibbons, 2021).

Por otra parte, un polimorfismo de un solo nucleótido (SNP) es una variación de un único nucleótido que se produce en una posición específica del genoma y representa la forma más básica de variación genética, la cual contribuye a la diversidad genética dentro de una población. Las diferencias en los patrones de SNP entre individuos y poblaciones se pueden utilizar para estudiar relaciones genéticas, genética de poblaciones y evolución de una especie, para identificar variantes genéticas comunes asociadas con condiciones específicas (Gauch *et al.*, 2018).

El análisis de componentes principales (PCA) es un método estadístico comúnmente utilizado en genética y genómica, particularmente con datos de polimorfismo de un solo nucleótido (SNP), ayuda a reducir la dimensionalidad de los datos manteniendo la mayor variabilidad posible. Esto es particularmente útil cuando se trata de grandes conjuntos de datos, como los generados en estudios genómicos. Este método transforma los datos originales de SNP en un nuevo conjunto de variables no correlacionadas llamadas componentes principales (PC). Cada componente principal es una combinación lineal de las variables del SNP originales y están ordenados por la cantidad de varianza que explican. Los primeros componentes principales capturan la mayor parte de la variabilidad de los datos (Abraham y Inouye, 2014)..

Los componentes principales pueden interpretarse en el contexto de la variación genética entre individuos. Las cargas de cada SNP en los componentes principales indican la contribución de ese SNP a la variabilidad general capturada por el componente principal correspondiente. Los resultados del PCA se pueden visualizar trazando individuos en el espacio definido por los primeros dos o tres componentes principales. Los grupos de individuos pueden indicar una estructura o relación genética subyacente. Esto ayuda a identificar verdaderas asociaciones genéticas al tener en cuenta la subestructura de la población (Gauch *et al.*, 2018).

Pregunta de investigación

¿Existe una relación entre las poblaciones genéticamente distintas de *Listeria monocytogenes* y la fuente de alimento de la que proviene?

Objetivos

- Conocer la estructura poblacional de aislados de *L. monocytogenes* de diversas fuentes de alimentos
- Determinar si hay una relación entre poblaciones genéticamente distintas de *L. monocytogenes* y la fuente de alimento de la que proviene

Hipótesis

Al utilizar análisis de componentes principales para las secuencias de aislados de *L. monocytogenes* se determinará si estos pertenecen a una población homogénea o a una población que contiene subgrupos genéticamente distintos. Además, al comparar los subgrupos, se determinará si estos se distribuyen uniformemente en las fuentes de alimento.

Método

Este trabajo se basa en Bechtel y Gibbons (2021) donde se analizaron 504 genomas de cepas aisladas, no clonalmente idénticas de "queso", "carne cruda escurrida", "entorno de procesamiento de mariscos", "aguacates hass", "helado", "lechuga", "leche", "queso crema para untar", "papa", "carne cruda", "leche cruda", "verduras crudas cortadas", "carne al por menor", "salami", "camarones", "ambiente de matadero" y "salmón ahumado" los cuales se clasificaron dentro de 5 grupos: Frutas/Vegetales, Mariscos, Carne, Quesos y Otros lácteos. Todos estos provenientes de 19 países (Canadá = 24, Estados Unidos = 272, México = 10, Brasil = 2, Chile = 12, Uruguay = 5, Dinamarca = 1, Francia = 7, Alemania = 1, Grecia = 1, Irlanda = 3, Italia = 64, Noruega = 3, Polonia = 2, Suiza = 10, Reino Unido = 62, Corea del Sur = 1, Australia = 1 y Nueva Zelanda = 19).

Se realizaron únicamente los pasos a seguir para llevar a cabo la descarga de las secuencias, su alineamiento con el programa MAFFT y el PCA con la herramienta TASSEL. Posteriormente, tomando los agrupamientos (a través del PCA) de *L. monocytogenes* basado en su similitud genética se realizó una gráfica de abundancia de cada población en los 5 grupos de alimentos, y un análisis de χ^2 para determinar estadísticamente si hay o no asociación entre las poblaciones genéticamente distintas y el tipo de alimento. A continuación se describen estos pasos y se encuentran en el Anexo 1.

○ Alineamiento

Para realizar los alineamientos se utilizó la herramienta MAFFT (Multiple Alignment using Fast Fourier Transform). Este es un programa que realiza alineaciones múltiples de secuencias biológicas. Los métodos de alineación que utiliza, son métodos basados en la transformada de Fourier y heurísticas (Barraza-Garza, *et al.*, 2013).

La transformada de Fourier es una fórmula matemática que transforma una función en el dominio del tiempo en su representación en el dominio de la frecuencia. Al representar estas secuencias en el espacio de frecuencias, es más sencillo identificar las diferencias entre una secuencia y otra.

La fórmula de la transformada de Fourier es la siguiente:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \cdot e^{-i\omega t} dt$$

Donde:

- $F(\omega)$ es la Transformada de Fourier de $f(t)$ en el dominio de la frecuencia.
- ω es la frecuencia angular.
- $e^{-i\omega t}$ es una función exponencial compleja.

MAFFT implementa esta fórmula para representar las secuencias biológicas en el espacio de la frecuencia y de este modo realiza la alineación múltiple.

- PCA

La herramienta TASSEL (Trait Analysis by aSSociation, Evolution and Linkage) principalmente realiza análisis estadísticos y genéticos de datos en variación genética, especialmente en estudios de asociación de genoma completo y el identificador de loci de caracteres cuantitativos.

	1532: 202883367	1533: 202883397	1534: 204059060	1535: 204059180	1536: 204059957	1537: 204764426	1538: 205427981	1539: 206492925	1540: 206493021	1541: 207114208	1542: 210235075	1543: 210235180	1544: 210235726	1545: 210235939	1546: 210236675	1547: 212497348	1548: 212893509	1549: 215393158	1550: 215393255	1551: 215393350	1552: 216310255	1553: 216310366	1554: 216812238	1555: 216812289	1556: 217610295	1557: 217610450	1558: 218367682	1559: 220857872
33-16	R	G	C	C	G	C	A	N	T	T	C	C	C	T	A	T	T	N	C	C	T	C	G	G	T	G	T	N
38-11	G	A	C	C	G	C	A	T	T	N	C	C	G	T	C	G	C	G	C	G	T	C	G	A	T	G	C	G
4226	G	A	C	C	G	A	A	T	Y	N	C	C	C	T	A	T	T	T	C	C	T	C	A	A	N	G	T	A
4722	G	G	C	C	G	C	N	T	C	T	C	C	C	T	A	T	N	G	C	C	T	C	N	A	T	G	N	G
A188	G	A	C	C	G	C	C	T	T	T	C	T	C	G	C	T	T	G	C	C	G	C	A	A	T	G	C	G
A214N	R	G	C	C	G	A	C	T	Y	C	C	C	C	T	A	T	T	G	C	G	T	C	A	A	N	G	C	G
A239	G	A	C	C	G	C	A	T	Y	C	C	C	C	T	A	T	T	G	C	C	T	C	A	A	T	G	C	G
A272	R	G	C	C	G	C	C	T	C	C	G	C	C	T	A	T	T	G	C	C	G	T	C	A	A	T	G	C
A441-5	R	G	C	C	C	C	A	T	C	C	C	C	C	T	A	T	C	G	C	C	C	G	A	A	T	G	C	G
A554	R	G	C	C	G	C	C	T	C	T	G	T	C	T	A	T	C	G	C	G	T	C	A	A	N	G	C	G
A556	R	G	C	C	G	C	C	T	C	T	G	T	C	G	C	T	T	G	C	C	T	T	G	G	T	G	C	G
A6	G	A	T	T	C	C	A	T	C	T	C	C	C	T	C	T	C	G	C	C	C	C	A	A	T	G	C	G
A619	G	A	C	C	G	C	A	T	C	T	C	C	C	T	A	T	C	G	C	G	C	C	A	A	T	G	T	G
A632	G	G	C	C	G	A	C	T	C	C	C	C	C	T	A	T	T	G	C	G	T	C	A	A	T	G	T	G
A634	G	G	C	C	G	A	C	T	C	C	C	C	C	T	A	T	T	G	C	G	T	C	A	A	T	G	T	G
A635	G	G	C	C	G	A	C	T	C	C	C	C	C	T	A	T	T	G	C	G	T	C	A	A	T	G	T	G
A641	G	A	C	C	G	A	A	T	Y	T	C	C	G	T	C	T	C	G	C	C	C	A	A	N	A	T	G	
A654	R	G	C	C	G	A	A	T	T	T	G	T	C	G	C	G	T	N	C	G	T	C	A	A	N	G	T	G
A659	G	A	C	C	G	C	A	T	T	C	C	C	C	T	A	T	T	G	C	G	T	A	A	T	G	C	A	
A661	R	G	C	C	G	C	A	T	C	N	N	C	C	T	A	T	C	G	C	N	C	A	A	T	G	C	G	
A679	G	G	C	C	G	A	C	T	Y	C	C	C	N	T	A	T	T	G	C	G	T	C	A	A	T	G	T	A
A680	G	G	C	C	G	A	C	T	N	C	C	C	C	T	A	T	T	G	C	G	T	C	A	A	T	G	T	A
A682	G	G	C	C	G	C	A	T	T	T	C	C	T	C	G	C	T	N	T	C	C	A	A	T	G	C	A	
AB28A	G	A	C	C	G	C	A	T	T	N	C	C	G	T	C	G	C	G	C	G	T	C	G	A	T	G	C	G
B10	R	G	C	C	G	C	A	T	Y	T	C	C	C	T	A	T	T	G	C	G	T	C	G	A	T	G	T	A
B103	R	G	C	C	G	C	A	T	T	T	C	C	C	T	A	T	T	G	C	G	T	C	G	A	T	G	C	G
B104	R	G	C	C	G	C	C	T	C	C	C	C	C	T	A	T	T	G	C	G	T	C	G	G	T	G	C	G
B105	G	G	C	C	G	A	C	T	T	T	G	T	C	G	C	T	T	G	C	C	T	T	G	G	T	G	C	G
B109	G	G	C	C	G	A	C	T	Y	C	C	C	C	T	A	T	T	G	C	C	T	C	A	A	T	G	T	A

Figura 1. Ejemplo de visualización al utilizar la herramienta TASSEL. Se observan los colores: amarillo (representa la secuencia normal), verde (las alteraciones existentes), blanco (espacios vacíos) y azul (recortes fijos). Además se pueden observar una serie de números. Cada fila es una secuencia de ADN. Y en la parte de arriba nos dice en qué parte de la secuencia nos encontramos.

Por último, TASSEL también permite hacer PCA. Esto lo hace analizando los componentes principales de las secuencias. Puede usar diversos métodos aunque el más efectivo es el de utilizar matrices de covarianza.

Los datos de SNP se representan como una matriz donde cada fila corresponde a un individuo y cada columna corresponde a un locus de SNP. Las entradas en la matriz normalmente representan la presencia o ausencia de un alelo particular (p. ej., 0, 1 o 2 para el número de copias de un alelo específico). Además, antes de aplicar PCA, es necesario estandarizar los datos de SNP para garantizar que cada SNP contribuya por igual al análisis. La estandarización implica restar la media y dividir por la desviación estándar de cada SNP.

- Abundancia relativa de *L. monocytogenes*

A partir del material complementario de Bechtel y Gibbons (2021) se obtuvieron los datos de la combinación de cada población con cada tipo de alimento, se calcularon sus frecuencias y se graficaron en Excel (Anexo 2).

○ Análisis de χ^2

La prueba de chi-cuadrado es una prueba estadística que se utiliza para determinar si hay una asociación significativa entre dos variables categóricas (Connelly, 2019).

Se establecieron las hipótesis: H0: No hay asociación significativa entre las dos variables categóricas, es decir, los aislados de las poblaciones se distribuyen uniformemente en las fuentes de alimento. vs. Ha: Hay una asociación significativa entre las dos variables categóricas, es decir, los aislados de al menos una de las poblaciones no se distribuyen uniformemente en las fuentes de alimento.

A partir del material complementario de Bechtel y Gibbons (2021) se obtuvieron los datos de la combinación de cada población con cada tipo de alimento y se realizó una tabla de contingencia (Anexo 2).

Se estableció la significancia estadística de $\alpha = 0,05$. Para calcular los grados de libertad se usó la fórmula: $gl = (r-1)*(c-1)$, donde c es el número de columnas, y r es el número de filas.

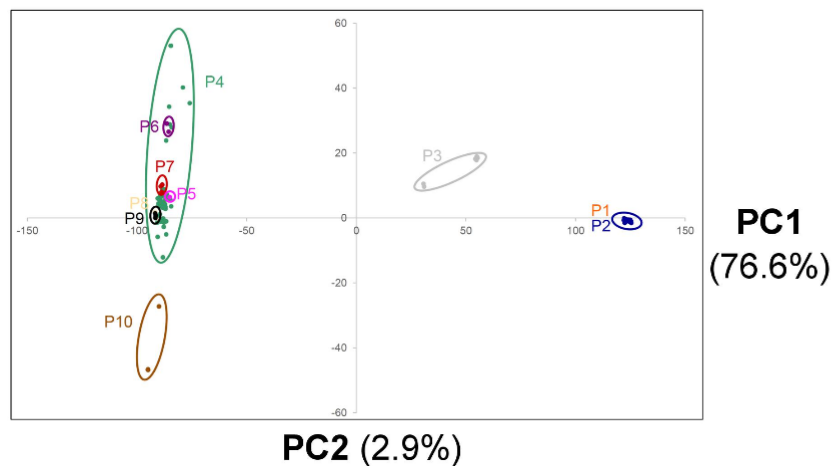
Se siguió la regla de decisión: Si el estadístico de χ^2 calculado es mayor que el valor χ^2 crítico, o si el valor de p es menor que α , se rechaza la hipótesis nula.

Finalmente, se realizó dicha prueba en R (Anexo 3).

Resultados y discusión

Con el PCA se definieron 10 poblaciones genéticamente distintas (P1-P10). Se obtuvo que el PC1 explicó el 76,6% de la varianza y separó las poblaciones P1, P2, P3 y P6 (Fig. 1A), PC2 explicó el 2,9% de la varianza y separó las poblaciones P4, P5, P6 y P7 (Fig. 1B), y PC3 explicó el 2,4% de la varianza y separó la población P9.

A



B

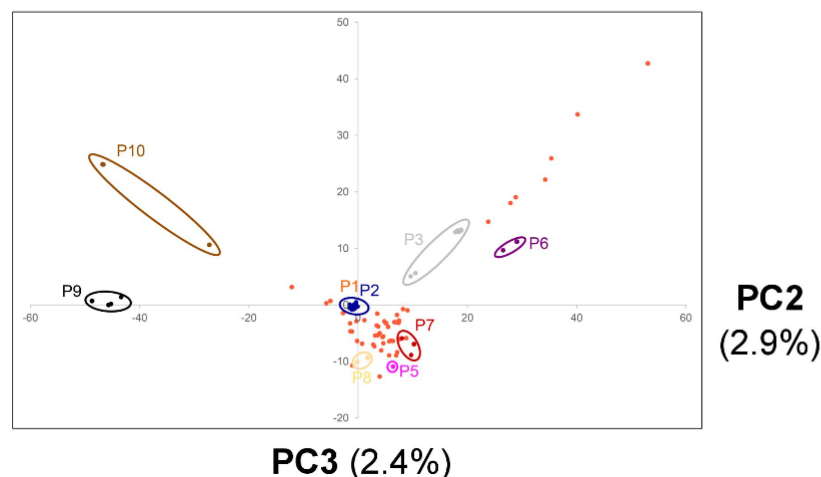


Figura 2. Estructura poblacional de 504 cepas de *L. monocytogenes* aisladas de alimentos. Análisis de componentes principales (PCA). (A) PC1 y PC2 y (B) PC2 y PC3. Modificado de Bechtel y Gibbons (2021).

Se mostró una representación excesiva de P2 en aislados de queso y frutas/verduras; P5, P8 y P9 mostraron una representación excesiva de aislados de mariscos, y P10 mostró una representación excesiva de aislados de carne. Además la P1, P6 y P10 no se presentaron en Frutas/Vegetales, ni la P3, P5 y P8 en Queso (Fig. 3).

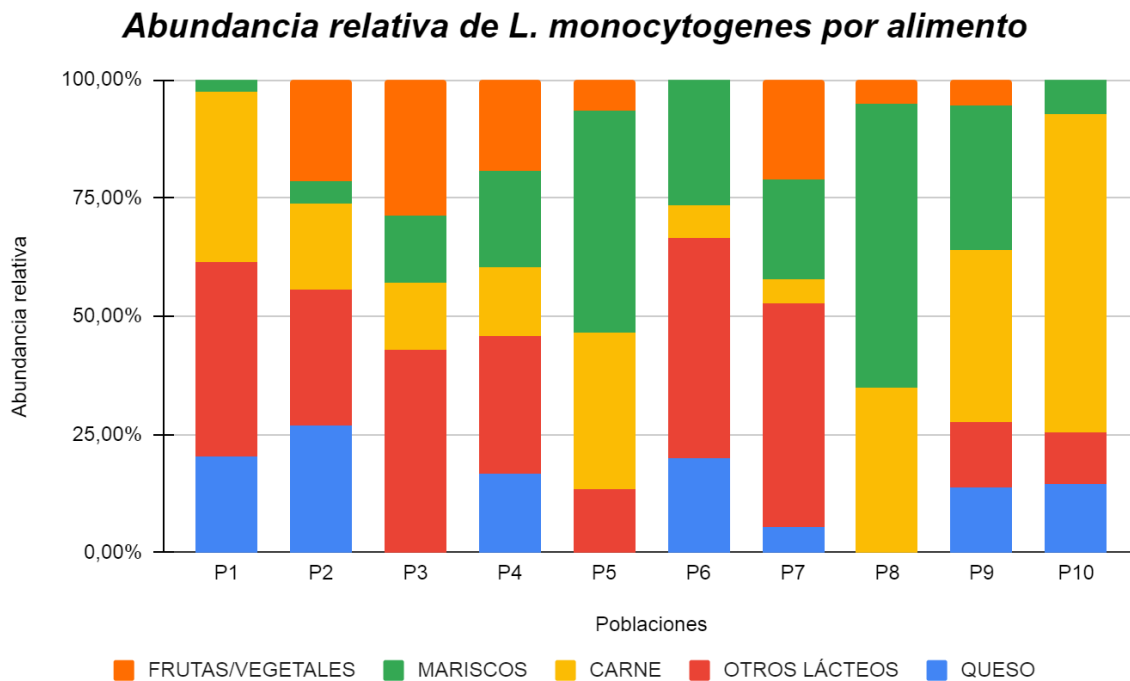


Figura 3. Abundancia relativa de *L. monocytogenes* por grupo de alimento.

En la prueba de χ^2 se obtuvieron $gl = (10-1)*(5-1)=36$ y un χ^2 crítico con $\alpha = 0,05 = 26.3$. Esto indicó que, con un 95% de confianza, se rechaza la hipótesis nula de que los aislados de las 10 poblaciones se distribuyeron uniformemente entre las fuentes de alimentos, dado que χ^2 calculada = 190.16 es mayor que χ^2 crítico (26.3) y el valor de $p < 2,2e-16$ es menor a 0,05. Es decir, hay una asociación estadísticamente significativa entre el tipo de alimento y las poblaciones genéticamente distintas de *L. monocytogenes*.

Pearson's Chi-squared test

```
data: datos1
X-squared = 190.16, df = 36, p-value < 2.2e-16
```

De acuerdo con Bechtel y Gibbons (2021) existen alrededor de 40 genes presentes en la población 2 y ausentes en la mayoría de las otras poblaciones. Además se presentan diferencias en genes que codifican proteínas de la pared celular (LPxTG) y genes que codifican proteínas con la mucina (MucBP). Estos genes incluyen importantes impactos BLAST en los reguladores transcripcionales de

la familia Crp/Fnr, los reguladores transcripcionales de la familia GntR, los reguladores transcripcionales de la familia TetR/AcrR. Y si bien, estos están implicados en su virulencia y cómo ingresan al huésped, sería importante estudiar si tienen una implicación en la tendencias al tipo de sustrato en el que predominan.

Conclusión

De las 10 poblaciones principales de *Listeria monocytogenes*, se encontró que la población 2 está sobrerrepresentada en Quesos y Frutas/Vegetales, las poblaciones 5, 8 y 9 en Mariscos, y finalmente, la población 10 en Carne. Además, las poblaciones 1, 6 y 10 no se observaron en Frutas/Vegetales, ni las poblaciones 3, 5 y 8 en Queso. Esto indicó que existen asociaciones con alimentos particulares, probablemente por sus diferencias genéticas.

Referencias

- Abraham, G. y Inouye, M. (2014). Fast principal component analysis of large-scale genome-wide data. PloS one, 9(4), e93766. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0093766>
- Barraza-Garza, Guillermo, de la Rosa, Laura A., Martínez-Martínez, Alejandro, Castillo Michel, Hiram, Cotte, Marine y Alvarez-Parrilla, Emilio. (2013). La microespectroscopía de infrarrojo con transformada de Fourier (FTIRM) en el estudio de sistemas biológicos. Revista latinoamericana de química, 41(3), 125-148. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0370-59432013000300001&lng=es&tlng=es
- Bechtel, T. y Gibbons, J. (2021). Population genomic analysis of *Listeria monocytogenes* from food reveals substrate-specific genome variation. Frontiers in Microbiology, 12, 620033. <https://www.frontiersin.org/articles/10.3389/fmicb.2021.620033/full#h9>
- Bintsis, T. (2017). Foodborne pathogens. AIMS microbiology, 3(3), 529. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6604998/>
- Connelly, L. (2019). Chi-square test. Medsurg Nursing, 28(2), 127-127. <https://search.proquest.com/openview/04d2ff080887f9111b68eb7490a9630a/1?pq-origsite=gscholar&cbl=30764>
- Gauch Jr, H. G., Qian, S., Piepho, H. P., Zhou, L. y Chen, R. (2018). Effective principal components analysis of SNP data. bioRxiv, 393611. <https://www.biorxiv.org/content/10.1101/393611v1.full.pdf>
- Sibanda, T. y Buys, E. (2022). *Listeria monocytogenes* pathogenesis: The role of stress adaptation. Microorganisms, 10(8), 1522. <https://www.mdpi.com/2076-2607/10/8/1522>
- Maize Genetics | TASSEL (2007). TASSEL - Trait Analysis by aSSociation, Evolution, and Linkage <https://www.maizegenetics.net/tassel>