

Análisis sobre la presencia de genes asociados a los patotipos diarrógenos de bacterias Escherichia Coli en muestras de aguas residuales recién tratadas según la estación climática vigente y la clasificación epidemiológica

Brandon Arias Sandoval¹, Catalina Fallas Flores¹, Marie Sofia Villalobos Martínez¹
brandon.ariassandoval@ucr.ac.cr, catalina.fallas@ucr.ac.cr, marie.villalobos@ucr.ac.cr

RESUMEN

Escherichia coli (E. coli) es una bacteria que generalmente se encuentra de manera inofensiva en la microbiota intestinal del ser humano, sin embargo, algunas de sus cepas patógenas tienen la capacidad de ocasionar diarreas. Es por tanto relevante su estudio, puesto que responde como la segunda mayor causa de muerte en niños, debido a las enfermedades diarreicas que origina. Esta bacteria es comúnmente encontrada en las heces y, por lo tanto, en las aguas residuales. Aunado a lo anterior, las plantas de tratamiento cumplen un papel fundamental ya que están diseñadas para remover o quitar la carga contaminante de las aguas residuales que provienen de una comunidad. En concordancia con lo antes mencionado, el objetivo principal del presente trabajo es determinar si la presencia de los genes asociados a los patotipos diarrógenos de esta bacteria en muestras de aguas residuales recién tratadas provenientes de sistemas de aguas negras y grises varía según la estación climática vigente o la clasificación epidemiológica. Para ello, se utilizó un modelo mixto lineal generalizado de medidas repetidas para probar dicho objetivo, en donde no se encontraron diferencias significativas de la propensión de la presencia de los genes según la estación climática vigente o la clasificación epidemiológica.

PALABRAS CLAVE: medidas repetidas, binomial, plantas de tratamiento.

INTRODUCCIÓN

De acuerdo a la Organización Mundial de la Salud (2018), la bacteria Escherichia coli (E. coli) pertenece a la familia Enterobacteriaceae y que generalmente se encuentra de manera inofensiva en la microbiota intestinal del ser humano según la información proporcionada por la Organización Mundial de la Salud (2018). A pesar de que esta bacteria raramente induce enfermedades, algunas de sus cepas patógenas tienen la capacidad de ocasionar diarreas o enfermedades extraintestinales tanto en personas sanas como en aquellas personas con inmunodeficiencias.

Dada la relevancia en la salud pública ante estas infecciones, según la Organización Mundial de la Salud (2017) las enfermedades diarreicas se posicionan como la segunda mayor causa de muerte de niños menores de cinco años, y ocasionan la muerte de 525 000 niños cada año. En Costa Rica, desde 1998, de igual forma, se considera a la enfermedad diarreica aguda (EDA) como la segunda causa de morbilidad (estado enfermo, de discapacidad o mala salud) y

¹ Estudiantes de bachillerato en Estadística, Universidad de Costa Rica.

mortalidad en la niñez y en los adultos mayores. La infección puede ser transmitida por medio de la ingesta de agua o alimentos contaminados, y también de una persona a otra debido a una higiene deficiente, como por ejemplo, un inadecuado lavado de manos, un factor contribuyente de esta propagación.

La bacteria del E.coli es comúnmente encontrada en muestras de heces y por lo tanto, en aguas residuales. En particular, la presencia de esta bacteria en aguas residuales, ha cobrado relevancia en el ámbito de la investigación epidemiológica, con el objetivo de realizar una constante vigilancia en este amplio sector poblacional en el cual se puede dar presencia de estos patógenos. Este seguimiento constante permite identificar y comprender la propagación de la bacteria para así evaluar las posibles implicaciones en la salud pública. Es fundamental destacar que las Plantas de Tratamiento de Agua Residual (PTAR), son un conjunto de obras de infraestructura, diseñadas hidráulica y sanitariamente para poder remover o quitar la carga contaminante de las aguas residuales que provienen de una comunidad o institución determinada. Estas son de gran importancia ya que minimizan diversos problemas causados por el mal manejo de las aguas residuales que en consecuencia afectan el entorno (Gómez, 2016). La vigilancia de las plantas de tratamiento implica la recolección de muestras y su posterior análisis para determinar la concentración de las cepas presentes. Este monitoreo garantiza prevenir los posibles riesgos asociados a la presencia de la bacteria *Escherichia coli*.

El objetivo principal de este presente trabajo es determinar si la presencia de ocho genes asociados a los patotipos diarrógenos de bacterias *Escherichia Coli* en muestras de aguas residuales recién tratadas provenientes de sistemas de aguas negras y grises varía según la estación climática que esté vigente, y de igual manera para la clasificación epidemiológica.

METODOLOGÍA

Para este estudio se procedió a definir y utilizar como variable respuesta (ecuación 2) la presencia de alguno de los ocho genes de virulencia asociados a los patotipos diarrógenos de bacterias *Escherichia Coli* en muestras de aguas residuales recién tratadas provenientes de sistemas de aguas negras y grises del año 2013. Se contó con 118 muestras recolectadas que corresponden a cinco plantas de tratamiento ubicadas dentro del Gran Área Metropolitana de Costa Rica, las cuales son: Montelindo, André Chalé, Rincón Verde, Los Reyes y Las Palmas. Dichas muestras se tomaron a la entrada y a la salida de las plantas en dos momentos del año: estación seca y estación lluviosa. En cada periodo se recolectaron 3 muestras de diferentes días (ver anexo 1).

Las variables que se utilizaron fueron las siguientes: nombre de la planta de tratamiento (*wwtp*), día del año en el que se realizó el muestreo (*día*), concentración de *Escherichia Coli* en la muestra en millones (*coli*), indicación de presencia de los genes en la muestra al salir de la planta (*any out*), indicación de presencia de los genes en la muestra al entrar a la planta (*any int*), estación en que se tomó la muestra (*sampling month1*) y clasificación de la semana epidemiológica de acuerdo a los casos de diarrea (*epidemiological class*).

Primero se realizó un análisis descriptivo de los datos, acto seguido se ajustó el modelo para la hipótesis de investigación que se evaluó durante el análisis estadístico, la cual fue

determinar si la presencia de ocho genes asociados a los patotipos diarrógenos de bacterias Escherichia Coli en muestras de aguas residuales recién tratadas provenientes de sistemas de aguas negras y grises varía según la estación climática que esté vigente y la clasificación epidemiológica.

Seguidamente se revisó la significancia de las variables independientes mediante pruebas Chi Cuadrado con un 95% de confianza, además, se calculó la asociación entre las variables predictoras para verificar el supuesto de multicolinealidad. El modelo lineal generalizado mixto se ajustó con cien mil iteraciones y se utilizó el método de optimización de BOBYQA (Bound Optimization BY Quadratic Approximation), este método es un algoritmo de aproximación cuadrática el cual funciona para que el modelo converja de manera adecuada, dicho modelo fue el siguiente:

$$\text{Log}\left(\frac{\pi_{i,j,f,k,D,C}}{1-\pi_{i,j,k,D,C}}\right) = (\beta_0 + \beta_{0,i}) + (\gamma_i + \gamma)D + \alpha_j + \tau_f + \lambda C + \theta_k \quad (1)$$

$$Y \sim \text{Binomial}(\pi_{i,j,k,D,C}) \quad (2)$$

Donde:

1. La i representa la planta de tratamiento ($i : 1, 2, 3, 4, 5$).
2. La j representa la estación climática ($j : 1, 2$).
3. La f representa la clasificación epidemiológica ($f : 1, 2, 3, 4$).
4. La k representa la presencia de genes en la muestra al entrar a la planta ($k: 1,0$).
5. β_0 = Intercepción general.
6. $\beta_{0,i}$ = Efecto aleatorio de la i -ésima planta de tratamiento sobre el intercepción.
7. γ_i = Efecto aleatorio de la i -ésima planta de tratamiento sobre la pendiente.
8. γ = Coeficiente asociado al efecto del día.
9. α_j = Coeficiente asociado al efecto de la estación climática, con la restricción: $\alpha_2 = -(\alpha_1)$.
10. τ_f = Coeficiente asociado al efecto de la clasificación epidemiológica, con la restricción: $\tau_4 = -(\tau_1 + \tau_2 + \tau_3)$.
11. λ = Coeficiente asociado al efecto de la concentración de Escherichia Coli.
12. θ_k = Coeficiente asociado al efecto de la presencia de los genes en la muestra al entrar a la planta, con la restricción: $\theta_2 = -\theta_1$.

$$\beta_{0,i} \sim N(0, \sigma_{\beta_{0,i}}^2)$$

$$\gamma_i \sim N(0, \sigma_{\gamma_i}^2)$$

Mediante el modelo ajustado (ecuación 1) se pusieron a prueba distintas hipótesis con un nivel de significancia de 5%. En primer lugar, se verificó la hipótesis de no correlación entre pendientes e interceptos (ecuación 3). En segundo lugar, se verificó la hipótesis de no pendientes aleatorias (ecuación 4). Dichas hipótesis se plantearon de la siguiente forma:

$$H0: \rho = 0 \text{ y } H1: \rho \neq 0 \quad (3)$$

$$H0: \sigma_{\gamma_i}^2 = 0 \text{ y } H1: \sigma_{\gamma_i}^2 > 0 \quad (4)$$

Donde:

ρ es la correlación entre pendientes e interceptos.

$\sigma_{\gamma_i}^2$ es la varianza de los efectos aleatorios.

$\sigma_{\beta_{0,i}}^2$ es la variabilidad asociada a las plantas de tratamiento.

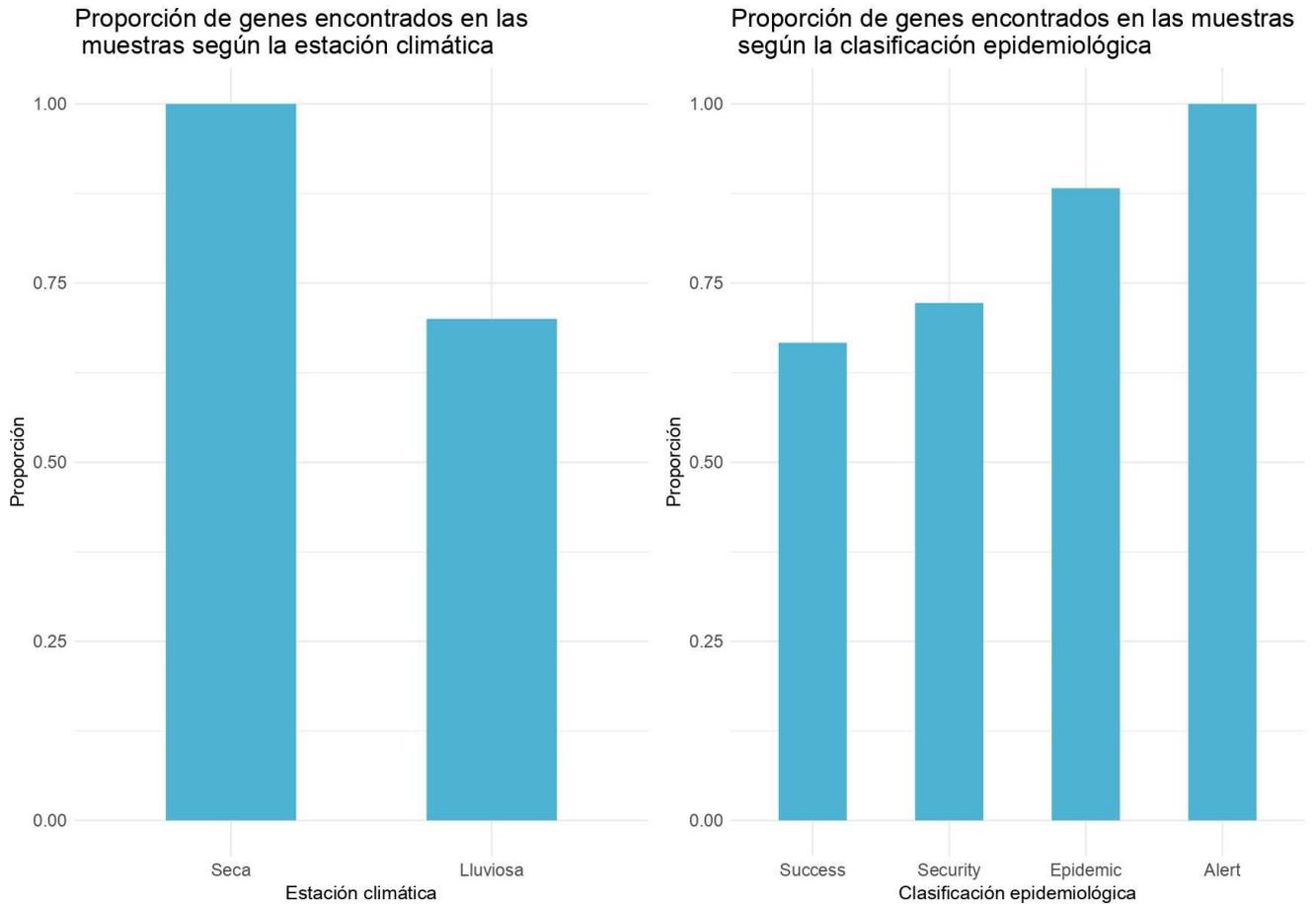
Posteriormente, se hicieron los respectivos contrastes de hipótesis para detectar diferencias entre las estaciones climáticas y también entre clasificación epidemiológica. Para el análisis estadístico de los datos y la elaboración de las distintas pruebas realizadas en el presente estudio, se utilizó el lenguaje de programación estadístico R en su versión 4.3.2, en el entorno Rstudio, para el proyecto realizado, se utilizaron los paquetes ggplot2 (Wickham. H, 2016), emmeans (Lenth. R, 2024) y lme4 (Bates. D, Maechler. M, Bolker. B & Walker. S, 2015).

RESULTADOS

En la primera etapa, el análisis descriptivo de los datos mostró que la proporción de la presencia de los genes en las muestras que son tomadas al salir de la planta es similar para cada una de las plantas de tratamiento (ver anexo 2). Al realizar las comparaciones según la estación climática, la proporción de la presencia de genes en las muestras al salir de la planta para la estación seca y la estación lluviosa es diferente, para esta última es inferior. Además, se calculó la proporción de la presencia de genes en las muestras al salir de la planta de tratamiento según la clasificación epidemiológica y se encontró que entre mayores casos de diarrea (alert) mayor es la proporción de genes en la muestra, en contraste, entre menos casos de diarrea (success), la proporción de genes en la muestra es menor (Ver Figura 1).

Figura 1

Proporción de genes encontrados al salir de la planta según estación climática y clasificación epidemiológica



En la segunda etapa del análisis se realizaron pruebas para verificar la asociación que tenían algunas variables entre sí (cuadro 1). Se estudió esta asociación entre la estación climática y la clasificación epidemiológica, también entre esta última y la presencia de genes al entrar a la planta. Ambas pruebas fueron significativas, lo que quiere decir que existe asociación entre estas variables. En contraste, la asociación entre la estación climática y la presencia de genes al entrar a la planta no fue significativa, lo que quiere decir que estas dos variables son independientes, como consecuencia de la asociación entre estas variables puede que existan problemas de multicolinealidad.

Cuadro 1

Prueba de Chi-Cuadrado para asociación de las variables

Variables	Probabilidad asociada
Estación climática - Clasificación epidemiológica	p < 0.001
Estación climática - Presencia de genes al entrar a la planta	0.3
Clasificación epidemiológica - Presencia de genes al entrar a la planta	0.02

En la tercera etapa del análisis, fueron puestas a prueba las hipótesis mediante pruebas de razón de verosimilitud (LRT). En primer lugar, se obtuvo que para la prueba de no correlación entre pendientes e interceptos (ecuación 3) no se rechazó la hipótesis nula ($p = 1$) lo cual implica que no existe una correlación entre una mayor propensión inicial de la presencia de los genes y que ésta aumente o disminuya de manera más acelerada conforme pasa el tiempo. En segundo lugar, para la prueba de no pendientes aleatorias (ecuación 4), de igual forma, no se rechazó la hipótesis nula ($p = 1$) lo que sugiere que las pendientes para cada planta de tratamiento son iguales.

Seguidamente, se verificó la significancia de las variables predictoras y se obtuvo que todas fueron significativas ($p < 0.05$), a excepción de la variable *coli* ($p > 0.05$). A pesar del resultado obtenido, se decidió mantener la covariable *coli* debido a que al retirar esta del modelo la significancia de *clasificación epidemiológica* dejaba de ser significativa, así que el modelo ajustado que se utilizó es el siguiente:

$$\text{Log}\left(\frac{\pi_{i,j,f,k,D,C}}{1-\pi_{i,j,k,D,C}}\right) = (\beta_0 + \beta_{0,i}) + \gamma D + \alpha_j + \tau_f + \lambda C + \theta_k \quad (5)$$

Acto seguido se procedió a realizar las pruebas de hipótesis sobre la presencia de genes para detectar diferencias entre estaciones climáticas, en el cuadro 2 se encuentran dichas comparaciones y su probabilidad asociada. Con un 95% de confianza se encontró que no hay diferencias significativas, por ende, la propensión de la presencia de los genes asociados a los patotipos diarrógenos de bacterias *Escherichia Coli* no varía según la estación climática.

Cuadro 2

Contrastes de la presencia de genes por estación climática.

Contrastes	Probabilidad asociada
Seca - Lluviosa	0.992

Seguidamente se realizaron las pruebas de hipótesis para detectar diferencias de la presencia de genes por clasificación epidemiológica, en el cuadro 3 se muestran tales

comparaciones y su probabilidad asociada. Con un 95% de confianza se encontró que no hay diferencias significativas, por ende, la propensión de la presencia de los genes asociados a los patotipos diarrógenos de bacterias *Escherichia Coli* no varía según la clasificación epidemiológica.

Cuadro 3

Contrastes de la presencia de genes por clasificación epidemiológica.

Contrastes	Probabilidad asociada
Alert - Epidemic	1
Alert - Security	1
Alert - Success	1
Epidemic - Security	0.982
Epidemic - Success	0.990
Security - Success	0.424

CONCLUSIONES

La etiología de las diarreas es poco conocida, debido a que el sistema de salud prioriza el informe de los casos, pero no la identificación exacta del agente causal (Chacón et al., 2022). En Costa Rica los estudios asociados a las causas de la presencia de los genes en la bacteria *E. Coli* en las plantas de tratamiento no han tenido mayor auge, por lo que se carece de estudios. Según Meza (1974) se requieren nuevos y más numerosos trabajos que permitan conocer mejor las causas y variaciones regionales de las diarreas en todo el territorio costarricense.

En síntesis, se logró verificar a nivel descriptivo de los datos que las proporciones de la presencia de *E. Coli* en las plantas de tratamiento varían según la estación climática (Figura 1), más no se logró demostrar mediante el análisis estadístico la existencia de diferencias significativas entre estas, así que la propensión de la presencia de genes no difiere según la estación climática. Sin embargo, según el estudio realizado por Hannaoui et al. (2010) la presencia de *E. coli* diarreogénica es mayor en los meses más cálidos (marzo, abril y mayo) por lo que, los resultados obtenidos difieren con los encontrados por este autor.

A nivel descriptivo se aprecia que la proporción de genes encontrados según la clasificación epidemiológica (Figura 1) es ligeramente diferente para cada nivel de la clasificación, pese a esto tampoco se logró encontrar diferencias estadísticamente significativas entre la clasificación epidemiológica.

Además de ello, la Doctora Luz Chacón, quien nos brindó su conocimiento en el área, planteó una hipótesis de investigación y anticipó que no se observarían diferencias significativas

de la presencia de la bacteria en función de la estación climática, ni de la clase epidemiológica, lo cual concuerda con los resultados obtenidos.

AGRADECIMIENTOS

Agradecemos a la Doctora Luz Chacón Jiménez del Instituto de Investigaciones de la Salud (INISA) cuya contribución ha sido fundamental para el desarrollo de esta investigación, brindándonos su conocimiento en el área.

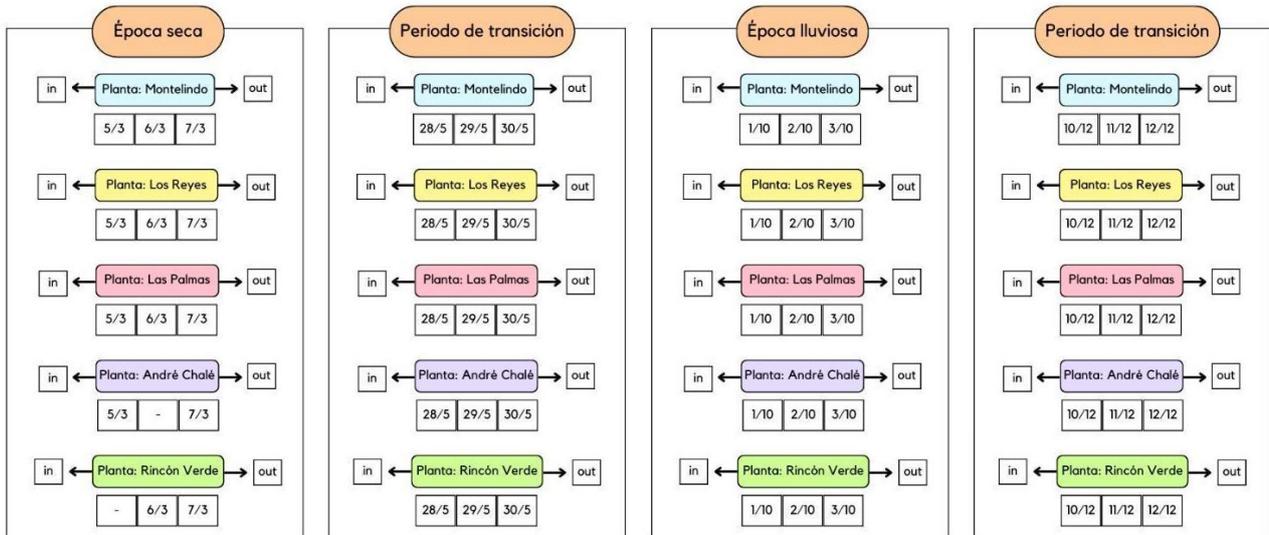
BIBLIOGRAFÍA

- Bates. D, Maechler. M, Bolker. B, Walker. S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01
- Chacón-Jiménez, L., Hall Loría, K., Rivera Navarro, P. C., Reyes Lizano, L., Achí Araya, R., & Barrantes Jiménez, K. (2022). *Circulación de genes de virulencia asociados a Escherichia coli diarrogénica en aguas residuales del Gran Área Metropolitana de Costa Rica*. *Población y Salud en Mesoamérica*, 19(2), 330-354. https://www.scielo.sa.cr/scielo.php?script=sci_arttext&pid=S1659-02012022000100330
- Gomes, T. A., Elias, W. P., Scaletsky, I. C., Guth, B. E., Rodrigues, J. F., Piazza, R. M., ... & Martinez, M. B. (2016). *Diarrheagenic escherichia coli*. *Brazilian journal of microbiology*, 47, 3-30. <https://www.scielo.br/j/bjm/a/NRsgPypwdTXGKCCmMSJs9TR/?lang=en&format=html>
- Hannaoui, E., Villalobos, L., Martínez, R., Maldonado, A., Hagel, I., & Bastardo, J. (2010). *Escherichia coli diarrogénica asociada a casos de diarrea aguda en niños de Cumaná, Venezuela*. *Investigación Clínica*, 51(4), 489-500. <https://ve.scielo.org/pdf/ic/v51n4/art06.pdf>
- Lenth, R. V. (2024, 6 abril). Estimated Marginal Means, aka Least-Squares Means [R package emmeans version 1.10.1]. <https://cran.r-project.org/web/packages/emmeans/index.html>
- Meza, R. (1974). *Algunos aspectos epidemiológicos de las diarreas en Costa Rica*. *Boletín de la Oficina Sanitaria Panamericana (OSP)*; 76 (5), mayo 1974. <https://iris.paho.org/bitstream/handle/10665.2/10762/v76n5p406.pdf;sequence=1>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. <https://ggplot2.tidyverse.org>
- World Health Organization. (2017). *Diarrhoeal disease*. <https://www.who.int/newsroom/factsheets/detail/diarrhoeal-diseas>
- World Health Organization. (2018). *E. Coli*. <https://www.who.int/es/news-room/factsheets/detail/e-coli>

ANEXOS

Anexo 1:
Figura 2

Esquema del diseño de los datos.



Anexo 2:
Figura 3.

Proporción de genes en la muestra al salir de la planta, dadas las diferentes plantas de tratamiento.

