(Formerly MENDELIANA)

BAG

Journal of Basic & Applied Genetics

December 2024
Volume XXXV
Issue 2
E-ISNN: 1852-6233

Journal of the Argentine Society of Genetics
Revista de la Sociedad Argentina de Genética
www.sag.org.ar/jbag
Buenos Aires, Argentina

DOI: 10.35407/bag.2024.35.02

Journal of the
Argentine Society
of Genetics

BAG | Journal of Basic & Applied Genetics

V. XXXV - No. 2

December 2024

Editorial Board
Comité Editorial

Diseño y maquetación
Lic. Mauro Salerno
maurosalerno92@gmail.com

Corrección de estilo
Dra. Gabriela Leofanti
gabrielaleofanti@gmail.com

Comité de colaboradores

Dra. María Mercedes Ibañez
Facultad de Agronomía y Veterinaria,
Universidad Nacional de Río Cuarto. Río Cuarto, Córdoba, Argentina
mibanez@ayv.unrc.edu

Dr. Daniel Maizon
Estación Experimental Agropecuaria "Guillermo Covas", Instituto Nacional de Tecnología Agropecuaria, y Universidad Nacional de La Pampa. Anguil, La Pampa, Argentina
maizon.daniel@inta.gob.ar

Dra. María Andrea Tomas
Estación Experimental Agropecuaria Rafaela,
Instituto Nacional de Tecnología Agropecuaria; Consejo Nacional de Investigaciones Científicas y Técnicas; Universidad Nacional de Rafaela. Rafaela, Santa Fe, Argentina
tomas.maria@inta.gob.ar

Imágen de tapa:
Nativa al abrigo de la sierra
Autora: **Gabriela Agustina Leofanti**

NOTA DEL EDITOR GENERAL

En los últimos años, la Inteligencia Artificial (IA) ha revolucionado las metodologías de investigación científica y redacción de artículos científicos, así como la gestión editorial de revistas científicas y libros académicos. Esta tecnología -que permite que las computadoras simulen la inteligencia y las capacidades humanas de resolución de problemas- involucra el desarrollo de algoritmos modelados a partir de los procesos de toma de decisiones del cerebro humano, que pueden "aprender" de los datos disponibles y realizar clasificaciones y predicciones.

Los beneficios del uso de la IA son claros, entre ellos, el aumento de la eficiencia de los procesos de indagación, análisis e interpretación de resultados, y toma de decisiones. Sin embargo, en la comunidad científica internacional han surgido preocupaciones por el empleo de modelos de lenguaje grandes (*Large Language Models o LLMs*) -como *ChatGPT*, *Google Bard*, o *Bing*- en las actividades que le son propias. Estas preocupaciones se centran principalmente en aspectos éticos y en la autenticidad e integridad de las publicaciones científicas debido al posible uso fraudulento o malicioso de las herramientas. De hecho, se han planteado desafíos para evitar los perjuicios que pueden derivarse del uso de la IA tales como refuerzo de sesgos, falta de privacidad de los datos (particularmente importante en la investigación con seres humanos), perpetuación de inexactitudes, y potencial para reducir el pensamiento crítico por exceso de confianza en las herramientas. La reducción del pensamiento crítico, en particular entre los investigadores más jóvenes que son más propensos a usar la IA por su familiarización con las tecnologías, puede tener consecuencias indeseables para el avance del conocimiento científico y sus aplicaciones.

Por lo antedicho, ha surgido la necesidad de desarrollar guías o protocolos para el uso de la IA en investigación científica, redacción científica, y gestión editorial, a fin de asegurar la aplicación responsable y ética. En relación a este punto, se han alcanzado algunos consensos mínimos en la comunidad internacional, a saber: (a) la IA utilizada para el desarrollo de un trabajo o la elaboración de un manuscrito no puede citarse como autor porque sólo los seres humanos pueden responsabilizarse del contenido de un artículo científico; (b) como entidades que no son legales, los modelos de IA no pueden hacer afirmaciones sobre conflictos de interés ni manejar derechos de autor o licencias de uso; (c) editores y revisores son responsables de las evaluaciones, opiniones y decisiones sobre los manuscritos que gestionan. En definitiva, los distintos actores del sistema científico son completamente responsables de sus acciones -incluyendo las realizadas por IA- y, por lo tanto, de cualquier incumplimiento ético.

Es fundamental, entonces, que los usuarios de la IA elijan las herramientas por los beneficios que de ellas pudieran obtenerse, pero en pleno conocimiento de las limitaciones y posibles perjuicios que pueden derivarse del uso inapropiado. Así mismo, la comunidad internacional debe seguir trabajando en el desarrollo de guías y protocolos que aseguren el uso responsable y ético de la IA en el campo científico.

Elsa L. Camadro

NOTE FROM THE GENERAL EDITOR

In the last years, Artificial Intelligence (AI) has revolutionized the methodologies of scientific research and the writing of scientific papers, as well as the editorial management of scientific journals and academic books. This technology -which allows computers to simulate the human intelligence and capacities for problem resolution- involves the development of algorithms modelled from the processes of decision making of the human brain that can "learn" from the available data to make classifications and predictions.

The benefits of using AI are clear, among them, the increasing efficiency of the processes of investigation, analysis and interpretation of results, and decision making. In the international scientific community, notwithstanding, concerns have been raised on the use of Large Language Models (LLMs) -such as *ChatGPT, Google Bard,* or *Bing*- in the activities that are their own. These concerns are principally centered in ethical aspects, and in the authenticity and integrity of the scientific publications due to the possible fraudulent or malicious use of the tools. In fact, challenges have been posed to avoid the damages that can be derived from the use of AI such as reinforcement of biases, lack of data privacy (particularly important in research with human beings), perpetuation of inaccuracies, and the potential to reduce critical thinking due to overconfidence in the tools. The reduction in critical thinking, particularly among the youngest researchers who are more prone to use AI due to their familiarization with technologies, might have undesirable consequences on the advancement of the scientific knowledge and its applications.

For the previously mentioned, the need has arisen to develop guides or protocols for the use of AI in scientific research and writing, and in editorial management to insure an ethical and responsible application. In relation to this point, minimum consensus have been reached by the international community, namely: (a) AI used for the development of a work or the elaboration of a manuscript cannot be cited as an author because only human beings can be responsible for the content of a scientific article; (b) not being legal entities, AI models cannot make statements regarding conflicts of interest or manage copyrights or licenses of use; (c) editors and reviewers are responsible for the evaluations, opinions, and decisions on the manuscripts handled by them. In short, the various actors of the scientific system are fully responsible of their actions -including those performed with AI- and, thus, of any unethical breach.

It is fundamental, then, that users of AI choose the tools according to the benefits that could be obtained from them, but in full knowledge of the limitations and possible damages that might be derived from their improper use. Likewise, it is a must that the international community continue the work on the development of guides and protocols to insure the responsible and ethical use of AI in the scientific field.

Elsa L. Camadro

BIBLIOGRAFÍA SUGERIDA
SUGGESTED BIBLIOGRAPHY

1. Sociedades Científicas y Editoriales
   Scientific Societies and Publishing Houses

   Committee on Publication Ethics (COPE)
   https://publicationethics.org/search?t=artificial+intelligence

   Journal of the American Medical Association (JAMA)
   https://jamanetwork.com/

   World Association of Medical Editors (WAME)
   https://www.wame.org/resources/ethics-resources/

2. Autores
   Authors

   BaHamman A. 2023. Balancing Innovation and Integrity: The role of Artificial Intelligence in Research and Scientific Writing. Nature and Science Sleep 15: 1153-1156

   Spinak E. 2023. https.//blog.scielo.org/es/2023/08/30/inteligencia-artificial-y-comunicación-en-investigaciones/

Contents
Contenidos

July-December 2024
Julio-Diciembre 2024

BAG | Journal of Basic & Applied Genetics

Vol XXXV (2); 13-25; July-December 2024

Journal of the Argentine Society of Genetics

POPULATION EXPANSION OF *Prosopis alba* GRISEB. (LEGUMINOSAE) IN SOUTHERN SOUTH-AMERICA: PHYLOGEOGRAPHICAL AND ECOLOGICAL APPROACH BASED ON CPDNA SEQUENCES

EXPANSIÓN POBLACIONAL DE *Prosopis alba* GRISEB. (LEGUMINOSAE) EN EL SUR DE SUDAMÉRICA: APROXIMACIÓN FILOGEOGRÁFICA Y ECOLÓGICA BASADA EN SECUENCIAS DE ADNCP

Bessega C.[1,2*], Pometti C.[1,2], Saidman B.O.[1,2], Fortunato R.[3,4], Santoro C.M.[5,6], Mcrostie V.[6,7,8], Vilardi J.C.[1,2]

[1] Departamento Ecología, Genética y Evolución (EGE), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires (UBA), Buenos Aires, Argentina.

[2] Instituto de Ecología, Genética y Evolución (IEGEBA), CONICET-UBA, Buenos Aires, Argentina.

[3] Instituto de Botánica Darwinion (CONICET/ANCEFYN) CONICET, Buenos Aires, Argentina.

[4] ESIlyCA, Universidad de Morón, Morón, Argentina

[5] Instituto de Alta Investigación, Universidad de Tarapacá, Arica, Chile

[6] Núcleo Milenio AFOREST, NCS2022_024, Chile

[7] Facultad de Ciencias Sociales, Escuela de Antropología, Pontificia Universidad Católica de Chile, Santiago, Chile

[8] Centro del Desierto de Atacama UC, Santiago, Chile

Corresponding author:
Cecilia Bessega
cecib@ege.fcen.uba.ar

ORCID 0000-0002-8575-1828

Cite this article as:
Bessega C., Pometti C., Saidman B.O., Fortunato R., Santoro C.M., Mcrostie V., Vilardi J.C. 2024. POPULATION EXPANSION OF *Prosopis alba Griseb.* (LEGUMINOSAE) IN SOUTHERN SOUTH-AMERICA: PHYLOGEOGRAPHICAL AND ECOLOGICAL APPROACH BASED ON cpDNA SEQUENCES. BAG. Journal of Basic and Applied Genetics XXXV (2): 13-25.

Received: 08/23/2023
Revised version received: 07/08/2024
Accepted: 07/31/2024

General Editor: Elsa Camadro

DOI: 10.35407/bag.2024.35.02.01
ISSN online version: 1852-6233

Available online at
www.sag.org.ar/jbag

ABSTRACT

Genealogical relationships among DNA lineages considering their current geographic distributions are useful to infer historical events that have shaped the contemporary distributions of species and their genetic variation. In this study we analyzed the variation of the *nadhF-rpl32* intergenic spacer in *Prosopis alba* Griseb. samples (*algarrobo*) collected in Chile, Argentina and Bolivia in order to contribute to our understanding of the evolutionary history of this species in southern South-America. We assessed the influence of environmental conditions on the demographic history of them by using a Bayesian ecological clustering (BPEC) approach and simulations based on the theory of coalescence. The results obtained allowed us to identify nine haplotypes. The Tajima (*TD*= -1.35) and Fu (*Fs*= -2.36) tests were non-significant, suggesting absence of selection. On the other hand, the disparity between sequences or raggedness (*rg*=0.021) was also non-significant, compatible with the population expansion. The coalescent analysis using MCMC indicated that the best fit demographic model was the linear growth one, with a time to the most recent common ancestor, for the haplotypes sampled in the present analysis, TMRA=0.0072, that is, roughly 7,000 generations. The BPEC analysis identified two clusters whose distribution partially overlaps in the Atacama Desert (Chile) and allows us to postulate that the species would have expanded to the north and west from the Chaqueña Region in Argentina. The comparison of scenarios by means of ABC (Approximate Bayesian Computation) analyses is in accordance with this result as the cases where the East cluster or the Argentinean samples were postulated as ancestral, yielded the higher posterior probabilities. The analysis performed contributed to the *P. alba* historical reconstruction throwing light on the trans Andean movement considering direction, time and natural- and human-mediated dispersal agents.

Key words: *algarrobo*, Atacama Desert, Bayesian analysis, coalescent models, *nadhF-rpL32* intergenic spacer, population genetics

RESUMEN

Las relaciones genealógicas entre linajes de ADN considerando sus distribuciones geográficas actuales son útiles para inferir eventos históricos que han dado forma a las distribuciones contemporáneas de las especies y su variación genética. En el presente trabajo se analizó la variación del espaciador intergénico *nadhF-rpl32* en muestras de *Prosopis alba* Griseb. (*algarrobo*) coleccionadas en Chile, Argentina y Bolivia para contribuir a nuestra comprensión de la historia evolutiva de esta especie en el sur de Sudamérica. Hemos evaluado la influencia de las condiciones ambientales sobre la historia demográfica de esta especie mediante la utilización de un enfoque de agrupamiento ecológico Bayesiano (BPEC) y simulaciones basadas en la teoría de la coalescencia. Los resultados obtenidos permitieron identificar nueve haplotipos. Las pruebas de Tajima (*TD*= -1.35) y Fu (*Fs*= -2.36) no fueron significativas, sugiriendo ausencia de selección. Por otra parte, la disparidad entre secuencias (*raggedness*, *rg*=0.021) tampoco fue significativa, compatible con la expansión poblacional. El análisis coalescente utilizando MCMC indicó que el modelo demográfico de mejor ajuste fue el de crecimiento lineal, con un tiempo hasta el ancestro común más reciente, para los haplotipos muestreados en el presente análisis, TMRA=0,0072, es decir, aproximadamente 7.000 generaciones. El análisis BPEC permitió identificar dos grupos cuya distribución se superpone parcialmente en el Desierto de Atacama (Chile) y permite postular que la especie se habría expandido hacia el norte y el oeste desde la Región Chaqueña en Argentina. La comparación de escenarios mediante análisis ABC (Cómputos Bayesianos Aproximados) concuerda con este resultado ya que los casos en donde las muestras del grupo del Este o de Argentina fueron postuladas como ancestrales arrojaron las mayores probabilidades posteriores. El análisis realizado contribuyó en la reconstrucción histórica de *P. alba* y en el esclarecimiento del movimiento trasandino considerando la dirección, el tiempo y los agentes de dispersión naturales y humanos.

Palabras clave: *algarrobo*, análisis Bayesiano, Desierto de Atacama, espaciador intergénico *nadhF-rpL32*, modelos coalescentes

INTRODUCTION

Examining population genetic structure over historical, spatial, and temporal scales and its relationship with environmental changes is crucial for understanding species distributions and adaptations to the ongoing climatic changes (Rico *et al.,* 2021). Phylogeography (Avise *et al.,* 1987) is a widely applied discipline that seeks to integrate the genealogical relationships among DNA lineages (sequences) with their current geographic distributions to infer historical events that have shaped the contemporary distributions of species and their genetic variation. More recently, Manolopoulou *et al.* (2016) proposed the use of Bayesian Phylogeographic and Ecological Clustering (BPEC) to combine DNA sequence genealogies with geographical distribution, environmental data, and phenotypic measurements to reveal geographic structuring of genetic clusters consistent with migration events.

Prosopis (Leguminosae, Mimosoideae) is a genus composed of 44 species belonging to five sections (*Prosopis, Anonychium, Strombocarpa, Monilicarpa* and *Algarobia*) well represented in arid and semi-arid regions of the world (Burkart, 1976). Recently, based on phylogenetic results, a proposal to disaggregate the genus was performed and the names for the sections and species were proposed to be changed (Hughes *et al.,* 2022). Here we still use the classification of *Prosopis* according to Burkart (1976), to consider the nomenclatural stability of taxonomic groups: Arts. 14.1, 34 and 56 from the International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) (Turland *et al.*, 2018).

Several *Algarobia* species are both ecologically and economically multipurpose trees as they can grow on sandy soils and contribute to stabilize dunes, combat desertification, and reforest degraded areas. This section (*algarrobos*) includes most species of economic importance that have been widely introduced in arid and semiarid regions, either with negative or positive effects on local populations (Burkart, 1976; Roig, 1993; Barros, 2010). *Algarrobos* are very appreciated in hot and arid areas as they provide shade, wood and edible fruits. The wood, very hard and highly caloric, is useful as firewood, charcoal and for furniture. Legume (pod) properties are highly variable among species and some of them are used as human food and forage (Burkart, 1976; Roig, 1993; Cony, 1996; Capparelli, 2007; Pometti *et al.,* 2009).

Prosopis alba Griseb. is one of the most important species of *Algarobia* from an economic point of view. It has been claimed that it is currently distributed in plains of subtropical Argentina, Uruguay, Paraguay, Bolivia, Perú and Chile (Burkart, 1976). In Chile, isolated natural and planted *Prosopis* populations can be found up to 3,000 masl throughout the western Andean slope, being typically confined to discrete zones of groundwater discharge and/or on riverbanks of perennial/ephemeral watersheds that flows into the Pampa del Tamarugal basin and the oases in the Salar de Atacama basin. Several authors (McRostie *et al.,* 2017; Bessega *et al.,* 2021) also recorded its presence in the Atacama Desert. Although this desert is one of the most hyper-arid on earth (Dunai *et al.,* 2005) some species of *Prosopis* can grow in it.

The distribution and cultural significance of Chilean *algarrobos* support the assumption that they are native to the Atacama Desert. AMS dating of *algarrobo* endocarps throughout several archaeological and paleoecological sites of the Atacama Desert estimates the presence of these trees around 4,000 years before present, suggesting that humans could have acted as vectors, at least for some of these species (McRostie *et al.,* 2017). Additionally, the considerable exo-morphological similarities between the trees from Atacama and *P. alba* specimens found in Salta Province, in the north of Argentina, have led Palacios and Brizuela (2005) to suggest that these species were introduced from Argentina during pre-Hispanic times. These authors stated that the presence and similarities between *algarrobos* from different localities in the Americas are a consequence of their cultivation, and this appears to be a case for domestication that, as with other American crops, did not persist with the arrival of European settlers.

Non-coding fragments of the chloroplast genome are the most appropriate markers for phylogeographical studies due to their uniparental inheritance and their capacity to detect processes of neutral evolution (Avise, 2000). By comparing DNA sequences, evolutionary relationships, levels of variation and geographical sub-structuring within and between groups of populations may be derived (Avise *et al.,* 1987). In particular the variation of the *nadhF-rpl32* intergenic spacer has been shown to be very useful in elucidating the origin and diversification of populations of a related species to *P. alba: P. chilensis* (Aguilar *et al.,* 2020). Here we analyzed sequence variation of the same cpDNA region in *P. alba* samples collected in Chile, Argentina, and Bolivia in order to contribute to our understanding of the evolutionary history of this species in southern South-America. We assessed the influence of environmental conditions on the demographic history of these species by using a BPEC approach and simulations based on the theory of coalescence. Considering the extreme characteristics of the Atacama Desert and previous works that discuss the endemism of *P. alba* in Chile, our work is based on the following two hypotheses: (1) Trans Andean ("Chilean") *P. alba* populations have been colonized by gene flow from "Argentinean" populations, and (2) the distribution of *P. alba* populations is at least partially affected by environmental factors. The phylogeographic ecological analysis may throw light in reference to the origin of *P. alba* populations in the Atacama Desert in Chile.

MATERIALS AND METHODS

Plant material, DNA extraction, amplification and sequencing

The samples of *Prosopis alba* were collected in 12 sites: two in Argentina, three in Bolivia, and seven in Chile. Within each sampling site, 1–6 individuals were collected, and their geographical coordinates were recorded with a handheld GPS (Table 1, Fig. 1A). Taxonomic identification of the specimens was based on Burkart (1976) and carried out by Reneé Fortunato. Herbarium vouchers of the material analyzed here are listed in Bessega *et al.* (2016, 2018, 2021) and deposited at BAB herbarium, INTA, Hurlingham, Buenos Aires, Argentina.

Total genomic DNA was extracted from the leaves of each plant using the DNA easy Plant mini kit (QIAGEN Inc., Valencia, CA, USA) following the instructions of the manufacturer. The non-coding chloroplast region *ndhF-rpL32* was selected for the present study as this region showed the greatest variation among several surveyed loci (*trnQ-rps16, trnH-psbA, rpl32R-ndhF, rpl32F-trnL,*

Table 1. Sampling sites, geographic coordinates, number of sequences (n) and *ndhF-rpl32* haplotypes of *Prosopis alba* from Chile, Argentina and Bolivia.

Sampling site	Country	Code	n	Latitude (S)	Longitude (W)	Haplotypes (N° of individuals)
Zapiga	Chile	ZAPI	1	19° 38’ 3.64’’	69° 57’ 7.51’’	H1 (1)
Tiliviche	Chile	TILI	1	19° 33’ 14.3’’	69° 57’ 11.5’’	H1 (1)
Valle Viejo	Chile	VVJO	3	20° 46’ 8.4’’	69° 34’ 6.79’’	H1 (2) H2 (1)
Quillagua Sur	Chile	QUIS	1	21° 39’ 5.86’’	69° 32’ 2.54’’	H3 (1)
Chiu Chiu	Chile	CHIU	4	22° 19’ 3.81’’	68° 39’ 0.65’’	H1 (2) H5 (1) H6 (1)
Yaye	Chile	YAYE	6	22° 55’ 6.72’’	68° 12’ 7.3’’	H1 (6)
Camar	Chile	CAMA	5	23° 24’ 3.53’’	67° 57’ 4.27’’	H1 (2) H3 (1) H4 (2)
Huajchilla	Bolivia	HUAJ	2	16° 37’ 26.35’’	68° 3’ 17.46’’	H7 (1) H8 (1)
Mecapaca	Bolivia	MECA	2	16° 40’ 2.89’’	68° 0’ 15.42’’	H1 (1) H4 (1)
Tahuapalca	Bolivia	TAHU	1	16° 43’ 16.7’’	67° 52’ 36.57’’	H1 (1)
Campo Duran	Argentina	CDUR	2	22° 10’ 56.64’’	63° 42’ 49.68’’	H1 (1) H4 (1)
Fernandez-Forres	Argentina	FFOR	1	27° 54’ 24’’	63° 54’ 13’’	H9 (1)

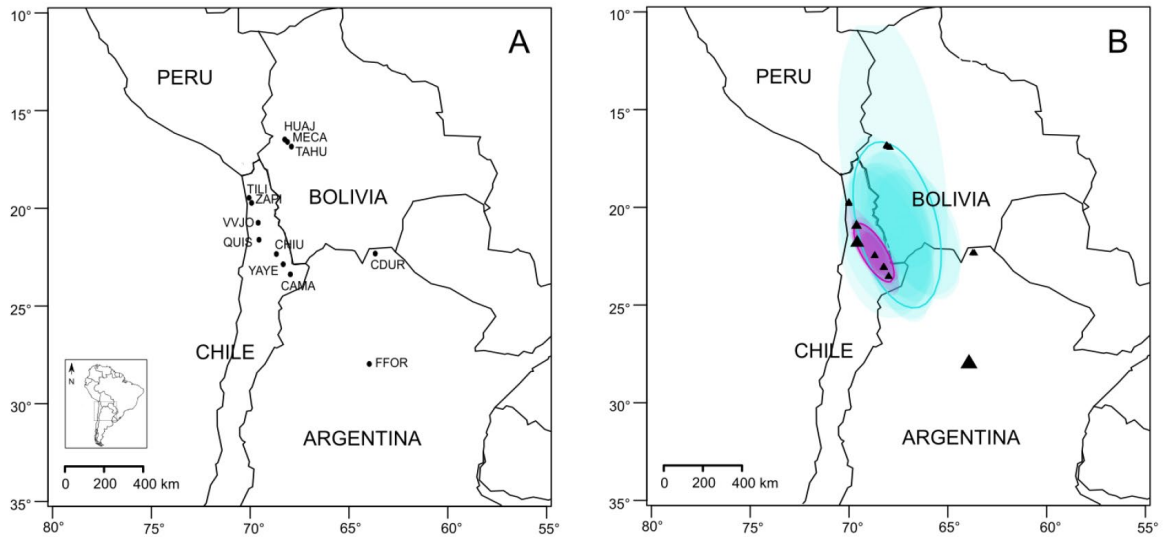

Figure 1. A: Location of the *Prosopis* alba sampling sites from Chile, Bolivia and Argentina in South America; **B**: Results from Bayesian phylogeographic and ecological clustering analysis (BPEC) based on *ndhF-rpl32* sequence data. The simulated contour areas are centered for each population cluster, and the shaded areas show the radius of 50% concentration contours around it. Note: light blue = cluster 1, violet = cluster 2. The triangles sizes represent the root probability of each sample. TILI: Tiliviche, ZAPI: Zapiga, VVJO: Valle Viejo, QUIS: Quillagua Sur, CHIU: Chiu-Chiu, YAYE: Yaye, CAMA: Camar, HUAJ: Huajchilla, MECA: Mecapaca, TAHU: Tahuapalca, CDUR: Campo Duran, FFOR: Fernandez-Forres.

trnD-trnT; Shaw *et al.,* 2007). The *ndhF-rpL32* intergenic region was amplified in 29 *P. alba* individuals, with the primers described by Shaw *et al.* (2007) and successfully used at intraspecific level before in *P. chilens*is (Aguilar *et al.,* 2020). The PCR amplifications were carried out in a 50 ul reaction volume containing 10–30 ng of DNA, 0.6 uM of each primer, 0.2 mM of dNTPs, 0.3 U of Taq DNA polymerase (Invitrogen, Foster City, CA, USA), and 1.5 mM of MgCl2. The amplifications were carried out in a T100 thermal cycler (Life Science Research, BioRad) with a cycling profile of initial denaturation at 94° C for 5 min followed by 35 cycles of 45 s at 94° (denaturation), 45 s at 55° (annealing) and 45 s at 72° C (extension), and a final extension step of 10 min at 72°.

Sequencing was performed by Macrogen Inc. (Seoul, South Korea, http://dna.macrogen.com). Electropherograms were visualized and edited using BioEdit software (Hall, 1999). Sequences were aligned with the multiple alignment option of BioEdit (Hall, 1999) and were adjusted manually. Gaps were not coded, and all the sequences are available upon request (cecib@ege.fcen.uba.ar), and accessible in the Genebank public database.

Data analysis

Genetic diversity, structure, and haplotype network

Considering that an isolation pattern due to physical barriers is expected mainly due to the Andean mountains, diversity among sequences was first quantified in each country by the mean number of Haplotypes (*H*), nucleotide diversity (π) (Nei and Li, 1979), haplotype diversity (*Hd*), the unique allele proportion and private alleles, using *strataG* package (Archer *et al.,* 2016) of R ver. 4.2.1 software (R Development Core Team, 2022).

Genetic structure was evaluated by analysis of molecular variance (AMOVA) using the function *poppr.amova* of the *poppr* package (Kamvar *et al.,* 2014, 2015). The genetic differentiation (Φ_{ST}) was estimated considering countries and BPEC clusters. The significance levels of Φ statistics were obtained with the function *randtest.amova* of the *ade4* package (Chessel *et al.,* 2004; Dray and Dufour, 2007; Dray *et al.,* 2007; Bougeard and Dray, 2018) of R software, with 2,000 permutations. To estimate the relationships between haplotypes, a minimum spanning network was constructed using the function *haploNet* from *pegas* package (Paradis, 2010) using R software.

Correlation between genetic and geographic distance

Pairwise Nei's (1978) genetic distances between populations (*D*) were estimated with the function *dist.genpop* of the package *adegenet* (Jombart, 2008; Jombart and Ahmed, 2011). The possible existence of isolation by distance (IBD) was analyzed by comparing pairwise population geographical distance with nucleotide divergence matrices by Mantel test (with 2,000 permutations) using the package *ade4* (Dray *et al.,* 2007).

Bayesian phylogeographical and ecological clustering (BPEC)

To verify the geographical structure of the populations and the most likely ancestral geographical locations, we analyzed the distribution of the *ndhF-rpL32* intergenic sequence variability by means of a BPEC approach, including geographical and environmental variables as potentially explanatory factors, using the package *BPEC* (Manolopoulou *et al.,* 2016) of R software. The analyses were conducted considering as covariates the geographical coordinates and the following seven environmental variables: (1) altitude, (2) mean temperature of the warmest quarter, (3) mean temperature of coldest quarter, (4) precipitation of driest quarter, (5) precipitation of the warmest quarter, (6) average spring normalized difference vegetation index (*NDVI*), and (7) average summer *NDVI*. *NDVI* data were obtained from NOP (NASA Earth Observations) from 200 m resolution images (https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MOD_NDVI_M). Environmental climatic data were taken from WorldClim v.2 (https://www.worldclim.org/data/worldclim21.html).

The running conditions were 100,000 iterations, saving 10,000 posterior samples, with a parsimony relaxation parameter, *ds*= 0. To avoid bias owing to a scale effect, environmental variables were scaled to mean= 0 and variance= 1. The probable geographical distribution of the clusters identified was plotted with the function *bpec.contourPlot* in R software. For the resulting clusters, genetic diversity parameters, population structure and haplotype network were also estimated in the same way as described above for countries.

Population demographic history

In order to evaluate the population demographic history, we used several proxies: a) Tajima´s *D* (*TD*) (Tajima, 1989) and Fu's F_S (Fu, 1997) statistical tests; b) mismatch distribution plot and raggedness index (Harpending, 1994); c) coalescent Monte Carlo simulations of Markov chains; and d) Approximate Bayesian Computational (ABC) analysis (Beaumont *et al.*, 2002).

TD and F_S were obtained with the functions *tajimasD* and *fusFS* using the *strataG* package (Archer *et al.,* 2016) while the mismatch distribution plot was produced with the function *MMD* of the *pegas* package (Paradis, 2010) using R software. The statistic *rg* was calculated with the specific script described in Vilardi *et al.* (2021). The statistical significance of *rg*, *TD* and F_S was determined by comparison with the neutral expected distribution obtained from coalescent model simulations generated with the function *coal_model* of the *coala* package (Staab and Metzler, 2016) using R software.

The coalescent analysis of the *nadF-rpl32* fragment sequences to determine the most probable time-dependent demographic model was conducted by Monte Carlo simulations of Markov chains (MCMC)

with the *coalescentMCMC* package (Paradis, 2015). Four demographic models (constant, exponential, step, and linear growth) were compared simulating 15,000 phylogenetic trees, from which the last 10,000 were kept (burnin=5,000). Models were compared according to their deviance information criterion (*DIC*) (Spiegelhalter *et al.*, 2002). The time to the most recent common ancestor (TMRCA) was estimated using the function *branching.times* from *ape* package from R software (Paradis *et al.*, 2004; Paradis and Schliep, 2018) assuming ultrametricity.

The ABC analysis procedure (Beaumont *et al.*, 2002) was performed with DIYABC v2.1 (Cornuet *et al.*, 2014) and based on the cpDNA dataset. Five different scenarios were tested with DIYABC considering the samples from the different countries as ancestral possibilities (scenarios 1, 2, and 3; Fig. 2A) and the ones from the BPEC clusters retrieved (scenarios 1 and 2; Fig. 2B). Priors for the different parameters were adjusted after performing 10,000 short simulations (Table S1).

Table S1. Parameters used as prior settings for DIYABC analysis.

Parameter	Distribution	Scenario	
		Minimum	Maximum
By countries			
Effective population size			
N1	Uniform	100	30000
N2	Uniform	100	30000
N3	Uniform	100	30000
Time of event			
t1	Uniform	1	25000
By clusters			
Effective population size			
N1	Uniform	100	30000
N2	Uniform	100	30000
Time of event			
t1	Uniform	1	25000

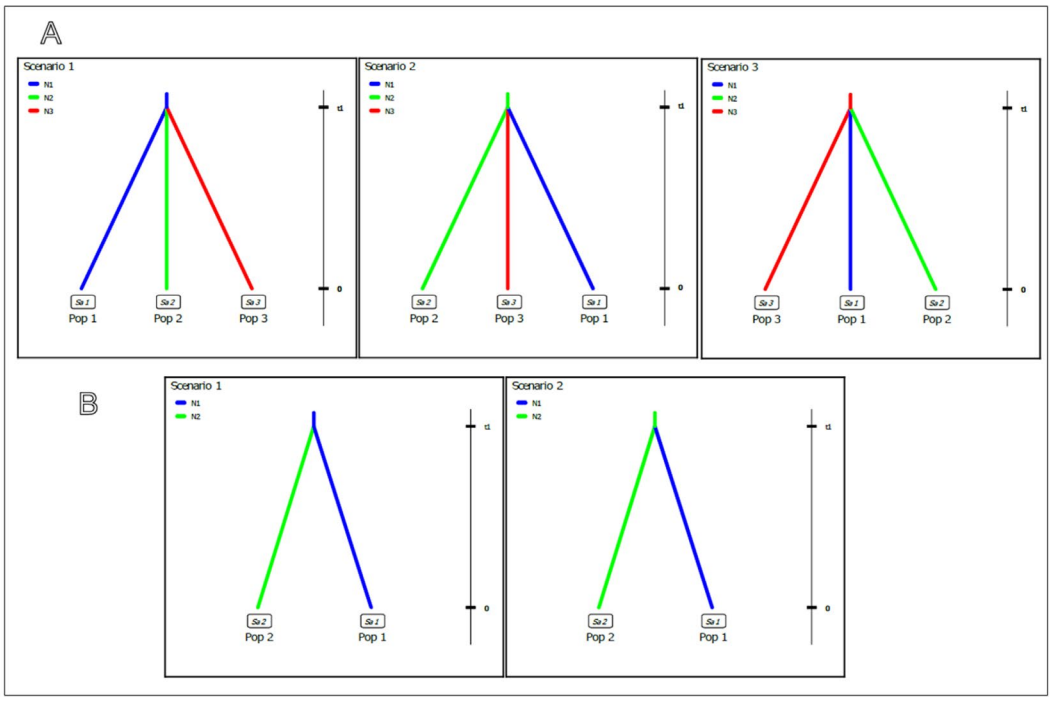

Figure 2. Five evolutionary scenarios (Sc) considered for comparison by ABC approach based on *ndhF-rpl32* sequence data. **A:** Sc1, Sc2 and Sc3 considering as ancestral the samples from different countries. Pop 1= Argentina, Pop 2= Bolivia and Pop 3= Chile. **B:** Sc1 and Sc2 considering as ancestral the samples from the different BPEC clusters. Pop 1= Cluster 1 (East) and Pop 2= Cluster 2 (West).

RESULTS

Genetic diversity, structure, and haplotype network by countries

The cpDNA *nadhF-rpl32* intergenic region was amplified and sequenced in all the *Prosopis alba* samples from Chile, Bolivia and Argentina allowing the recognition of 9 haplotypes (Table 1).Genetic variation was analyzed by country (Table 2). Haplotype diversity (*Hd*) was high in Bolivia and Argentina (0.9-1), and medium in Chile (0.55). The nucleotide diversity (p) was low in all the countries varying among 0.004 and 0.005. In reference to the private alleles, Chile was the country with the highest number of private alleles (4) followed by Bolivia (2) and Argentina (1). As the sample size was very different among countries, the unique allele proportion gives a better idea of the country diversity, being higher in Argentina (1) followed by Bolivia (0.6) and finally by Chile (0.14).

The differentiation among countries (Table 3) was estimated by the Φ_{ST} index that resulted low and not significant (Φ_{ST}=0.014, *P*=0.315). In agreement with this, the AMOVA analysis indicated that most of the variation occurs within countries (98.6%) and only 1.4% of the variation can be detected among countries.

The relationships among the cpDNA haplotypes were represented with a median-Joining network (Fig. 3A). The network shows that the two most frequent alleles (H4 and H1) differ in a single mutation, while the other haplotypes arose by the accumulation of several mutations. The network does not show a geographical association, there are haplotypes that are present in the three countries (H1 and H4) but there are also others that are only present in Chile (H3, H2, H5 and H6), in Bolivia (H7 and H8) and in Argentina (H9).

Correlation between genetic and geographic distance

The correlation between genetic and geographic distances was negative and non-significant (*r*=0.056, *P*=0.3), indicating that genetic differentiation between populations did not fit a spatial pattern as the expected by the isolation-by-distance model.

Bayesian phylogeographical and ecological clustering (BPEC)

The phylogeographical clustering determined that the most probable number of migration events was 1 (*P*=1) yielding two clusters (Fig. 1B) with high posterior probabilities for all the haplotypes (*PP*>0.999) to belong to cluster 1. Each individual was assigned unequivocally to its cluster (*P*=1). The network central haplotypes (H4, H1 and H9) exhibited the highest root node posterior probabilities (*PP*=0.173, 0.170 and 0.155, respectively) being H4 the most probable ancestral one.

The posterior distribution of the covariates showed significant differences between the two clusters for six out of seven environment variables (average summer *NDVI*, average spring *NDVI*, mean temperature of warmest quarter, mean temperature of coldest quarter, precipitation of driest quarter, precipitation of warmest quarter; $P<2x10^{-16}$). Only the altitude differentiation between the clusters was no significant (*t*=-3.43, P=0.00059). Considering the sample here analyzed, the most likely ancestral location was Fernández-Forres (FFOR) in Argentina (*P*=0.155, cluster 1, Fig. 1B).

Genetic diversity, structure and haplotype network by BPEC clusters

When genetic variability distribution was analyzed considering the clusters retrieved by BPEC (Table 2), the cluster 2 (from the west) exhibited less variation

Table 2. Molecular diversity indices considering the sampling sites (by country, A) and the genetic groups retrieved by (by cluster, B) of *Prosopis alba* from Chile, Argentina and Bolivia. H= number of haplotypes/alleles, π = nucleotide diversity, Hd = haplotype diversity, UA = unique alleles proportion, PA= private alleles, SD = standard deviation

	Grouping	Location	Diversity				
			Number of Haplotypes (H)	Nucleotide diversity (π) (SD)	Haplotype diversity (Hd) (SD)	Unique alleles proportion (UA)	Private alleles (PA)
A	BY COUNTRY						
		CHILE	6	0.00406 (0.00124)	0.55714 (0.124)	0.14	4
		BOLIVIA	4	0.00580 (0.00206)	0.900 (0.161)	0.60	2
		ARGENTINA	3	0.00446 (0.00157)	1.000 (0.272)	1	1
B	BY CLUSTERS						
		CLUSTER 1 (EAST)	7	0.00553 (0.00165)	0.800 (0.083)	0.33	6
		CLUSTER 2 (WEST)	3	0.00275 (0.00103)	0.385 (0.149)	0.07	2
	GLOBAL		9	0.00439 (0.00109)	0.648 (0.096)	0.21	-

according to the haplotype diversity (*Hd*= 0.38 vs. 0.80), private alleles (2 vs. 6) and unique allele proportion (0.07 vs. 0.33).

The differentiation between the BPEC clusters (Table 3) was borderline significant (Φ_{ST} =0.08, P=0.053) and much higher than the differentiation among countries shown above. The partition of the variation was consistent in demonstrating high variation within clusters and low variation, about 8%, among clusters.

In the median-joining network analyzed according to the clusters recovered by *BPEC* (Fig. 3B), H2 and H3 haplotypes were only found in cluster 2, H1 was shared by both clusters and the rest of the haplotypes were located exclusively in cluster 1.

Population demographic history

The first proxy to infer the demographic history of *P. alba* and to detect past population growth was based on Tajima's *D* and Fu's F_S statistical test. Both indices were negative but not significant as they fall inside the $CI_{95\%}$ generated by the coalescent simulations (*TD*=-1.351 [-1.727–1.995], F_S=-2.326 [-3.831–4.622]).

A second approach to evaluate the demographic history was based on the mismatch distribution plot (Fig. 4) and the raggedness index. The mismatch distribution plot did not show a multimodal and ragged shape revealing that the population is not in demographic equilibrium. According to this, the raggedness index estimated under

Table 3. Analysis of molecular variance (AMOVA) and genetic differentiation (ΦST) based on *ndhF-rpl32* sequences of individuals of *Prosopis alba* from Chile, Argentina and Bolivia. Grouping by country (A) and by the retrieved clusters (B).

Grouping	df	Sum of squares	Estimated variance	%	Genetic differentiation
A. By Country					
Among countries	2	4.253	0.027	1.389	Φ_{ST} = 0.014
Within countries	26	50.781	1.953	98.61	(P= 0.315)
Total	28	55.035	1.980	100	
B. Considering BPEC clusters					
Among clusters	1	4.368	0.172	8.397	Φ_{ST} = 0.084
Within clusters	27	50.667	1.877	91.603	(P= 0.053)
Total	28	55.034	2.049	100	

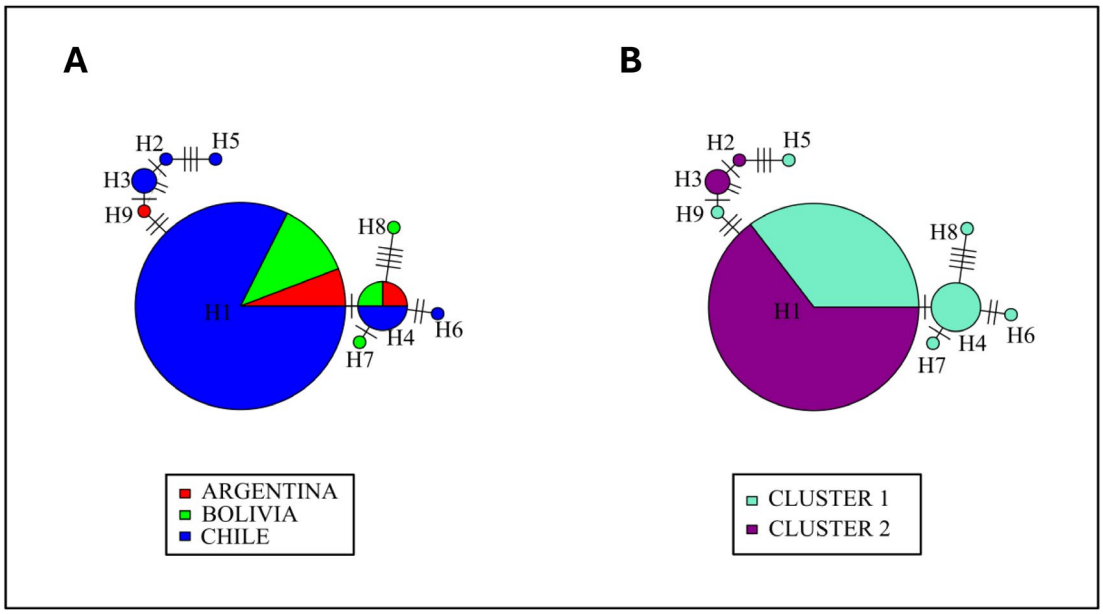

Figure 3. Haplotype network of the *Prosopis alba* based on *ndhF-rpl32* haplotypes. **A:** circles are colored according to the country of haplotype; **B:** circles are colored according to the genetic clusters obtained by BPEC. The size of the circles is proportional to the frequencies of each haplotype across all populations. Lines represent the mutational steps between haplotype sequences.

the demographic expansion model was non-significant (*rg*=0.021 [0.017–0.342]).

The third approach performed was based on coalescent trees simulated by MCMC procedures. The best fitting model was the linear one (*DIC*=-132.01). According to this, model theta (θ) grew from 1.15 x 10^{-16} to 1.98 x 10^{-2} and the time to the most recent ancestor (TMRA) was 0.0072 units of time before present (Fig. 5).

The same analysis was also applied considering only the samples included by BPEC in the cluster 2. The fittest model was the linear one (DIC=-445.09), with θ growing from 6.8 x 10^{-3} to 8.27 x 10^{-3} and a TMRA=0.0042.

Finally, ABC demographic inference analysis showed that among scenarios 1, 2 and 3, the scenario 1, that assumes the Argentinean populations as the ancestral ones, is the most likely scenario that would underline the observed genetic population divergence considering the direct and the logistic approach (Fig. 6A,B) with a posterior probability [$CI_{95\%}$] of 0.362 [0.0000-0.7832] and 0.3887 [0.3550-0.4223] respectively. When comparing scenarios 1 and 2 considering the BPEC retrieved clusters, the scenario 1 that considers the cluster 1 (east one, light blue) as ancestral, yielded a higher posterior probability [$CI_{95\%}$] for the direct (0.512 [0.0739-0.9501]) and the logistic regression (0.5655 [0.5111-0.6199]) approaches, suggesting that it is the best supported one (Fig. 6C,D).

DISCUSSION

In this study we analyzed from a phylogeographic standpoint samples from different *P. alba* populations from southern South-America in order to contribute to elucidate the evolutionary history of this important forest resource in arid and semiarid regions. We analyzed the genetic diversity, population structure, demographic history and Bayesian phylogeographical and ecological clustering (BPEC) based on the *nadhF-rpl32* intergenic spacer. We found evidence of two genetic lineages and signatures of demographic expansions.

Genetic diversity was analyzed considering the country distribution of the populations and the BPEC retrieved clusters. Genetic diversity in Chile and in cluster 2, the western one, was lower than in the remaining countries and cluster according to the haplotype diversity (*Hd*) and unique allele proportions estimators. Genetic diversity is expected to be lower in recent colonized areas in comparison with the original sites and the high presence of private haplotypes can be considered an indicator of long-term residency given the low mutation rate of organellar DNA (Wares *et al.*, 2002; Torre *et al.*, 2022). Additionally, the combination of high haplotype diversity and low nucleotide diversity, as observed in Argentina and Bolivia, can be signature of a rapid demographic expansion from a small effective population size (Avise, 2000; Joshi *et al.*, 2013). In accordance with this, we found signals of demographic expansion based on the different considered proxies. First, the negative Tajima's D (*TD*) and Fu's (*Fs*) estimates showed an excess of low frequencies polymorphisms relative to expectation, suggesting size expansion. Second, the mismatch distribution plot was unimodal and consequently the smoothness of the distribution (*rg* index) resulted no significant, and third, the coalescent analyses indicated that the linear model was the one that best fits the demographical populational expansion. The genetic diversity parameters and the signals of population expansion obtained may be interpreted in accordance with our first hypothesis that suggests that the Chilean populations have been colonized by populations from the east through trans-Andean migration. Finally, the comparison of several divergent scenarios by means of ABC analyses is in accordance with this result as the cases where the east cluster samples and the

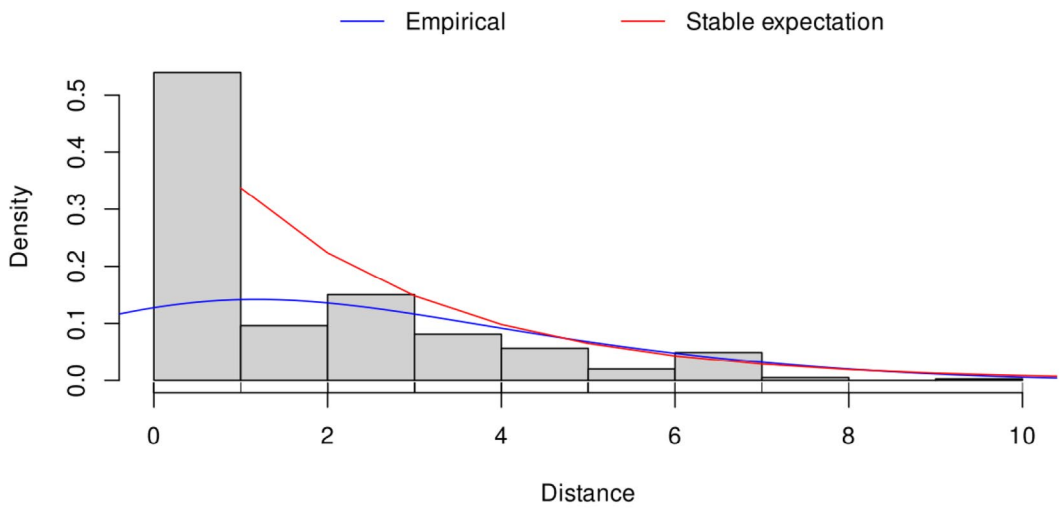

Figure 4. Pairwise mismatch distribution plot obtained based on the *ndhf-rpl32* sequences from *Prosopis alba*.

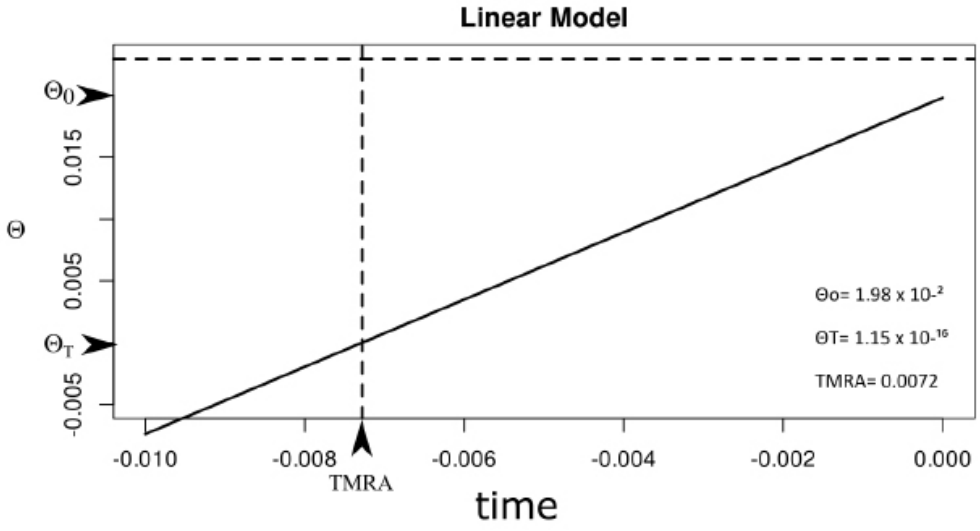

Figure 5. Expected variation of θ through time for the linear growing model. The predicted line is compared with the constant model (horizontal dashed line) and the vertical dashed line represents the time to the most recent common ancestor (*TMRCA*).

Argentinean ones were postulated as ancestral, yielded the higher posterior probabilities. This hypothesis is partially compatible with the proposal made by Palacios and Brizuela (2005). These authors suggested that morphological resemblance between some species of *algarrobo* trees found in Quilmes (Tucumán), Tolombón (Salta), Belén, Bolsón de Fiambalá and Chaschuil river (Catamarca) in Argentina and in Quillagua (Río Loa) and Canchones (Pampa del Tamarugal) in Chile, could be due to the introduction from east to west in the midst of a process of domestication in the Atacama Desert during pre-Hispanic times. This possibility was also suggested by McRostie *et al.* (2017) that provided evidence that the *algarrobo* species were not native to the Atacama Desert of Chile. It was suggested that in Atacama, the spread of species of section *Algarobia* was contemporaneous with the introduction of other crops by humans (McRostie, 2014; McRostie *et al.*, 2017). Based on AMS radiocarbon dates from archaeobotanical and palaeoecological records and settlement patterns in the Atacama Desert, it was suggested that the introduction of these trees occurred late in the Holocene (ca. 3,000 yr BP or later), and that the most likely vectors were humans. Moreover, Ugalde *et al.* (2021) analyzed the presence and abundance of *Prosopis* species in Atacama considering different cultural periods over the last 13,000 years and conclude that legume trees (i.e. chañar and algarrobo), probably introduced later in the Holocene, became a key resource from the Formative Period in the Pampa del Tamarugal, Loa and Salar de Atacama basins, linked with the Chaco region and the Northwest of Argentina. Seeds may have been circulated as part of intense mobility systems (caravans), which were rooted in traditional ways of life (McRostie *et al.*, 2022).

This aspect could be discussed considering the genetic differentiation estimations obtained in *P. alba* samples from different countries based on the *ndhF-rpL32* intergenic sequences. Our phylogeographic survey showed that only 1.4% of the genetic differentiation was found among countries. This percentage -which represents low estimates of genetic differentiation- could be indicative of a recent differentiation between the populations, compatible with the relative recent colonization suggested by Palacios and Brizuela (2005) and McRostie *et al.* (2017). The lack of detection of a clear spatial pattern goes in the same direction as no significant correlation was detected between the genetic and geographic distance matrices. However, the differentiation reaches up to 8% and the Φ_{ST} estimate results borderline significant when BPEC groups are considered. This result should be taken into account with greater certainty since climatic variables were considered for obtaining the clusters through BPEC. It is possible to consider that environmental variables such as temperature, precipitation and *NDVI* may partially explain the structure detected in *P. alba.* The Atacama area (that mainly overlaps with cluster 2, west) is one of the most arid deserts on Earth that has its own unique climatic characteristics. Two main causes may be contributing to explain the extreme aridity of the region. The cold water of the Humboldt Current, running parallel to the Chilean and southern Peruvian coasts and preventing precipitation in the coastal areas; and the occurrence of the Andes cord, that produces a rain-shadow effect, blocking moisture from the Amazon basin. As expected by the second hypothesis proposed, the environmental differences may be contributing to explain the geographical patterns of genetic variation retrieved.

The lack of association observed in the median joining network of haplotypes considering the countries (Fig. 3A) can also be interpreted considering the environmental variables and association expected in BPEC network under such scenario. However, in the network recovered by BPEC analysis (Fig. 3B) inconsistencies were also found. The eastern cluster (cluster 1) presents more exclusive haplotypes, suggesting indirectly that cluster 2 (from west) might be considered as derived. Indeed, our BPEC analysis lends some weight to this conclusion, since the most likely ancestral location detected, considering the sampling here conducted, was in the Chaqueña region in Argentina. Gene flow from Argentina to Chile has also been suggested for *P. chilensis*, a species related to *P. alba*, based on cpDNA sequence data, since Bolivian and Argentinean haplotypes were found in Chilean samples. A possible colonization (or introduction) of *P. chilensis* from the Bolivian Chaco and Argentinian Monte to the Chilean Matorral was proposed (Aguilar *et al.*, 2020). The complex pattern here found in *P. alba* is somehow inconsistent with a single migration event, as the haplotypes H2 and H3 are absent in the eastern populations, however this may be attributed to genetic drift or sampling error.

The low level of genetic differentiation among countries and clusters can be interpreted as signals of recent differentiation events although the low substitution rate of cpDNA and the long reproductive cycles of *Prosopis* species, should be considered. For woody plants with long reproductive cycle, their cpDNA mutation rate is much lower than in other species (Gaut *et al.*, 1996; Yan *et al.*, 2018). In oaks (*Quercus* spp.), for example, where seedlings need 3-4 years to produce their first acorns, the substitution rate resulted very low (0.19-0.96 x 10^{-9} s/s/y), even below the average values reported for non-coding regions in other angiosperm lineages (1.2-1.7 x 10^{-9} s/s/y). Additionally, in *Cercidium*, the average substitution rate reported is also much lower (0.318 x 10^{-9} s/s/y) than the average values generally reported for non-coding regions of the chloroplast genome, but also consistent with the notion that woody taxa should have slower rates of molecular evolution (Qi *et al.*, 2012).

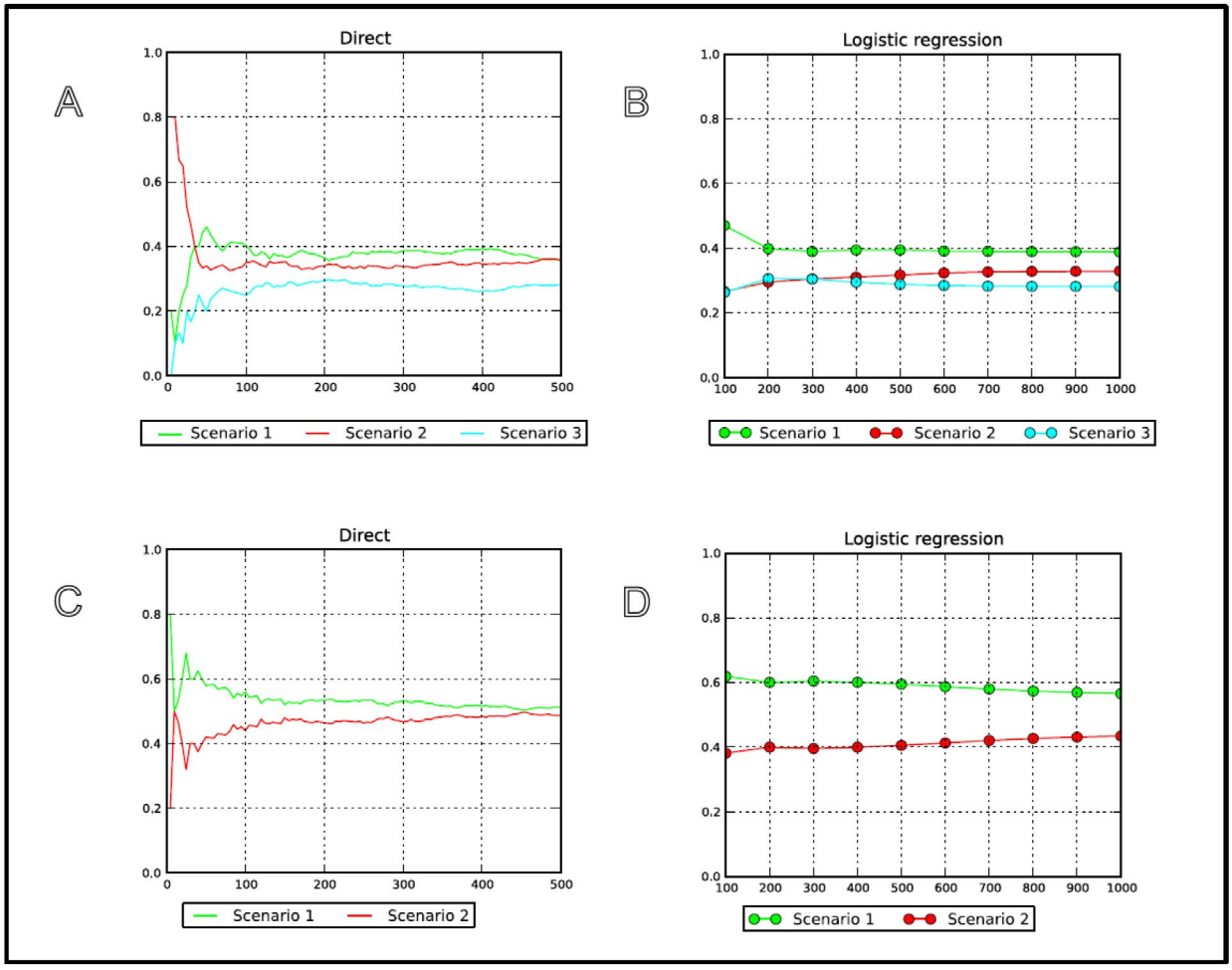

Figure 6. Plots showing fitness of competing scenarios based on direct estimates and logistic regression simulated in DIYABC based on *ndhF-rpl32* sequence data. **A and B:** comparison between scenarios 1, 2 and 3 by country. **C and D:** comparison among scenarios 1 and 2 by BPEC clusters.

The global nucleotide diversity estimated here for *ndhF-rpL32* was π =0.004, which may imply a divergence time among haplotypes of 4 to 6.2 Mya if we consider the mean substitution rate in oak and *Cercidium* one respectively. These time estimates are consistent with that obtained using the same cpDNA region in *P. chilensis* populations from Argentina and Chile, where the haplotype divergence in the phylogroups detected in Argentina was proposed in 5.22 Mya and was associated with the Paranaean Sea marine transgression (Aguilar *et al.,* 2020).

Luebert (2011) suggests that disjunct elements in arid and semi-arid areas on both sides of The Andes occurred through the mountain uprising, which created a barrier to the dispersal, promoting differentiation of species on both slopes. The timeline of the Andean uprising is currently a matter of controversy, but it seems to be agreed that the current elevation is early Pliocene (Luebert, 2011). The lineages distributed in the Desert of Atacama, the Chaco and the tropical and/or Mediterranean Andes could be explained considering that Andean areas were colonized after the vicariance from the basal areas of one or both sides of the mountain range. Long distance for the latter cases, via trans-Andean corridors, could not be ruled out *a priori*. Indeed, in *Prosopis* it was suggested that successive vicariance events split the ancestral wide area, and long-distance dispersal episodes led to recolonizations from North to South America, and vice versa (Bessega *et al.,* 2006). However, the estimated population expansion times estimated are much lower. The coalescent MCMC analysis yielded a TMRA value (Θ= 0.007) that suggest that 7,000 generations have occurred since the most recent common ancestor for the haplotypes sampled in the present analysis. As *Prosopis* trees may have long overlapping generations from near 50 to ~100 years, and it was seen and described that the first bloom and pods production occurs near the third and fourth year of growth (Castillo and Tarnowski, 2011), populations may have been expanding since about 28,000 years bp. When the analysis is restricted to the sequences belonging to cluster 2 (west), TMRCA estimate drops to near a half (Θ= 0.004), suggesting that the expansion time in Chilean territory would have started around 16,000 years bp.

Human-mediated migration may have played a substantial contribution to current distribution of *Prosopis* populations on both sides of The Andes given the traditional link between this resource and the original communities (Pasiecznik *et al.,* 2001; Giovannetti *et al.,* 2008; Rivera and Dodd, 2013; Uribe *et al.,* 2020). However, based on the present results, *P. alba* may have started to disperse to the north and west from the Chaqueña Region in Argentina much earlier, and the movement may have been mediated by natural trans-Andean dispersers such as guanacos, small mammals and/or birds.

Although the analysis here performed is only based on the cpDNA *nadhF-rpl32* intergenic region and can be consider weak, it contributes in the *P. alba* (*algarrobo)* historical reconstruction throwing light on the trans-Andean movement, considering direction and time. A more intensive sampling of the South-American populations in combination with more cpDNA and nuclear genotyping is needed to give a more complete picture of the phylogeography of *P. alba*. Up to our knowledge, this is the first phylogeographic study in *P. alba* which gives strong evidence of the introduction of *P. alba* in the Atacama Desert pointing the Chaqueña region in Argentina as the source area.

REFERENCES

Aguilar D.L., Acosta M.C., Baranzelli M.C., Sérsic A.N., Delatorre-Herrera J., Verga A., Cosacov A. (2020) Ecophylogeography of the disjunct South American xerophytic tree species *Prosopis chilensis* (Fabaceae). Biol. J. of Linn. Soc. 129(4): 793-809. doi: 10.1093/biolinnean/blaa006

Archer F.I., Adams P.E., Schneiders, B.B. (2016) *stratag*: An r package for manipulating, summarizing and analyzing population genetic data. Mol. Ecol. Resour. 17: 5–11.

Avise J.C. (2000) Phylogeography. The history and formation of species. Cambridge: Harvard University Press, Cambridge, Massachusetts, United States.

Avise J.C., Arnold J., Ball R.M., Bermingham E., Lamb T., Neigel J.E., Reeb C.A., Saunders N.C. (1987) Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics. Annu. Rev. Ecol. Syst. 18:489-522. doi: 10.1146/annurev.es.18.110187.002421

Barros S. (2010) El género *Prosopis*, valioso recurso forestal de las zonas áridas y semiáridas de América, Asia y Africa. Ciencia e Investigación Forestal 16(1): 91–128.

Beaumont M.A., Zhang W., Balding D.J. (2002) Approximate Bayesian computation in population genetics. Genetics 162: 2025–2035. doi: 10.1093/genetics/162.4.2025

Bessega C., Pometti C.L., Ewens M., Saidman B.O., Vilardi J.C. (2016) Fine-scale spatial genetic structure analysis in two Argentine populations of *Prosopis alba* (Mimosoideae) with different levels of ecological disturbance. Eur. J. For. Res. 135(3): 495–505.

Bessega C., Pometti C., Fortunato R., Greene F., Santoro C.M., McRostie V. (2021) Genetic studies of various *Prosopis* species (Leguminosae, Section *Algarobia*) co-occurring in oases of the Atacama Desert (northern Chile). Ecol. Evol. 11(5): 2375–2390. doi: 10.1002/ece3.7212

Bessega C., Pometti C., López R.P., Larrea-Alcázar D., Fortunato R.H., Saidman B.O., Vilardi J.C. (2018) Genetic diversity and differentiation among *Prosopis alba* (Leguminosae) populations from dry valleys of Bolivia with different levels of human disturbance and altitude. Ecol. Evol. 8(22): 11309–11321. doi: 10.1002/ece3.4610

Bessega C., Vilardi J.C., Saidman B.O. (2006) Genetic relationships among American species of the genus *Prosopis* (Mimosoideae, Leguminosae) inferred from ITS sequences: evidence for long-distance dispersal. J. Biogeogr. 33: 1905–1915.

Bougeard S., Dray S. (2018) Supervised multiblock analysis in R with the ade4 package. J. Stat. Softw. 86(1): 1–17. doi: 10.18637/ jss.v086.i01

Burkart A. (1976) A monograph of the genus *Prosopis* (Leguminosae subfam. Mimosoideae). J. Arnold Arboretum 57(3–4): 217–249; 450–525.

Capparelli A. (2007) Los productos alimenticios derivados de *Prosopis chilensis* (Mol.) Stuntz y *P. flexuosa* DC., Fabaceae, en la vida cotidiana de los habitantes del NOA y su paralelismo con el algarrobo europeo. Kurtz. 33(1): 1–19.

Castillo E.M., Tarnowski E. (2011) Cultivo experimental de *Prosopis chilensis* en el Pedemonte de las Yungas. https://inta.gob.ar/sites/default/files/script-tmp-prosopis_chilensis.pdf

Chessel D., Dufour A.B., Thioulouse J. (2004) The ade4 package – I: One-table methods. R News 4(1): 5–10.

Cony M.A. (1996) Genetic variability in *Prosopis flexuosa* D.C. a native tree of Monte phytogeography province, Argentina. For. Ecol. Manag. 87: 41–49.

Cornuet J.M., Pudlo P., Veyssier J., Dehne-Garcia A., Gautier M., Leblois R., Marin J.M., Estoup A. (2014) DIYABC v2.0: A software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. Bioinformatics 30: 1187–1189.

Dray S., Dufour A. (2007) The ade4 package: Implementing the duality diagram for ecologists. J. Stat. Softw. 22: 1–20. doi: 10.18637/ jss.v022.i04

Dray S., Dufour A., Chessel D. (2007) The *ade4* package – II: two-table and K-table methods. R News 7: 47–52.

Dunai T.J., Gonzalez Lopez G.A., Juez-Larre J. (2005) Oligocene–Miocene age of aridity in the Atacama Desert revealed by exposure dating of erosion-sensitive landforms. Geology 33(4): 321–324. doi: 10.1130/G21184.1

Fu Y.X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147: 915–925.

Gaut B.S., Morton B.R., McCaig B.C., Clegg M.T. (1996) Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. P. Natl. Acad. Sci. (PNAS), USA. 93:10274–10279. doi: 10.1073/pnas.93.19.10274

Giovannetti M., Lema V., Bartoli C., Capparelli A. (2008) Starch grain characterization of *Prosopis chilensis* (Mol.) Stuntz and *P. flexuosa* DC, and the analysis of their archaeological remains in Andean South America. J. Archaeol. Sci. 35: 2973-2985. doi: 10.1016/j.jas.2008.06.009

Hall T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acid. S. 41: 95–98.

Harpending H.C. (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. Hum. Biol. 66: 591–600.

Hughes C.E., Ringelberg J.J., Lewis G.P., Catalano S.A. (2022) Disintegration of the genus *Prosopis* L. (Leguminosae, Caesalpinioideae, Mimosoid clade). PhytoKeys 205: 147-189. doi: 103897. phytokeys.205.75379

Jombart T. (2008) Adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405. doi:10.1093/bioinforma tics/btn129

Jombart T., Ahmed I. (2011) Adegenet 1.3-1: New tools for the analysis of genome wide SNP data. Bioinformatics 27(21): 3070–3071.

Joshi J., Salar R.K., Banerjee P., Upasna S., Tantia M.S., Vijh R.K. (2013) Genetic Variation and Phylogenetic Relationships of Indian Buffaloes of Uttar Pradesh. Asian-Australas. J. Anim. Sci. 26 (9):1229-1236. doi: 10.5713/ajas.2012.12669

Kamvar Z.N., Brooks J.C., Grünwald N.J. (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front. Genetics 6: 208.

Kamvar Z.N., Tabima J.F., Grünwald N.J. (2014) *Poppr*: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2: e281.

Luebert F. (2011) Hacia una fotogeografía histórica del Desierto de Atacama. Rev. Geogr. Norte Gd. 50: 105-133.

Manolopoulou I., Hille A., Emerson B. (2016) BPEC: an R package for Bayesian phylogeographic and ecological clustering. arXiv preprint arXiv:1604.01617.

McRostie V.B. (2014) Arboricultura y Silvopastoralismo en el Período Formativo (1.400 a.c.-500 D.C.) de la cuenca del Salar de Atacama. Chungara (Arica) 46(4): 543-557. doi: 10.4067/S0717-73562014000400002

McRostie V.B., Babot P., Calás E., Gayó E., Gallardo F., Godoy-Aguirre C., Labarca R., Latorre C., Núñez L., Ojeda K., Santoro C.M., Valenzuela D. (2022) Silvopastoralism and the shaping of forest patches in the Atacama Desert during the Formative Period (ca. 3000–1500 years BP). The Holocene 32(12): 1492–1502. doi: 10.1177/09596836221122636

McRostie V.B., Gayó E.M., Santoro C.M., De Pol-Holz R., Latorre C. (2017) The pre-Columbian introduction and dispersal of Algarrobo (*Prosopis*, Section Algarobia) in the Atacama Desert of northern Chile. PLoS ONE 12(7): e0181759. doi: 10.1371/journal.pone.0181759

Nei M. (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 23: 341–369.

Nei M., Li W.H. (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA 76: 5269–5273.

Palacios R.A., Brizuela M.M. (2005) Fabaceae, parte 13. Subfam. II. Mimosoideae, parte 4. Tribu VI. Mimoseae, parte B. *Prosopis* L. Flora Fanerogámica Argentina, 92: 3–25.

Paradis E. (2010) *pegas*: an R package for population genetics with an integrated–modular approach. Bioinformatics 26: 419–420.

Paradis E. (2015) *coalescentMCMC*: MCMC algorithms for the coalescent. R package version 0.4-1. https://CRAN.R-project.org/package=coalescentMCMC (accessed March 2023).

Paradis E., Schliep K. (2018) *ape 5.0*: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35: 526–528.

Paradis E., Claude J., Strimmer K. (2004) *APE*: analyses of phylogenetics and evolution in R language. Bioinformatics 20: 289-290.

Pasiecznik N., Felker P., Harris P.J.C., Harsh L.N., Cruz J.C., Tewari K., Cadoret K., Maldonado L.J. (2001) The *Prosopis juliflora-Prosopis pallida* complex: a monograph. HDRA, Coventry, UK.

Pometti C.L., Pizzo B., Brunetti M., Macchioni N., Ewens M., Saidman B.O. (2009) Argentinean native wood species: Physical and mechanical characterization of some *Prosopis* species and *Acacia aroma* (Leguminosae, Mimosoideae). Bioresource Technol. 100(6):1999-2004.

Qi X., Chen Ch., Comes H.P., Sakaguchi Sh., Liu Y., Tanaka N., Sakio H., Qiu Y. (2012) Molecular data and ecological niche modelling reveal a highly dynamic evolutionary history of the East Asian Tertiary relict *Cercidiphyllum* (Cercidiphyllaceae). New Phytol. doi: 10.1111/j.1469-8137.2012.04242.x

R Development Core Team (2022) R: A Language and Environment for Statistical Computing. R Foundation for Statistical computing. http://www.r-project.org/ (accessed January 2023)

Rico Y., León-Tapia M.Á., Zurita-Solís M., Rodríguez-Gómez F., Vásquez-Morales S.G. (2021) Influence of Pleistocene climatic oscillations on the phylogeography and demographic history of

endemic vulnerable trees (section Magnolia) of the Tropical Montane Cloud Forest in Mexico. PeerJ 9:e12181. doi: 10.7717/peerj.12181

Rivera M., Dodd J. (2013) Domesticando el Desierto: medio ambiente y ocupaciones humanas en Ramaditas, Desierto de Atacama. Diálogo Andino 41: 45-60. doi: 10.4067/S0719-26812013000100004

Roig F.A. (1993) Aportes a la Etnobotánica del Género *Prosopis*. In: Iadiza (Ed.), Contribuciones Mendocinas a la quinta Reunión Regional para América Latina y el Caribe de la Red de Forestación del CIID. Conservación y mejoramiento de especies del género *Prosopis*. Unidades de Botánica y Fisiología Vegetal (IADIZA) Editores, Mendoza, pp. 99-119.

Shaw J., Edgar B., Lickey E.B., Schilling E.E., Small R. (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in Angiosperms: the tortoise and the hare III. Am. J. Bot. 94: 275–288.

Spiegelhalter D.J., Best N.G., Carlin B.R., van der Linde A. (2002) Bayesian measures of model complexity and fit. J. R. Stat. Soc., Series B 64: 583–616.

Staab P.R., Metzler D. (2016) *coala*: an R framework for coalescent simulation. Bioinformatics 32: 1903–1904.

Tajima F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–595.

Torre S., Sebastiani F., Burbui G., Pecori F., Pepori A.L., Passeri I., Ghelardini L., Selvaggi A., Santini A. (2022) Novel Insights Into Refugia at the Southern Margin of the Distribution Range of the Endangered Species *Ulmus laevis*. Front. Plant Sci. 13: 826158. doi:10.3389/fpls.2022.826158

Turland N.J., Wiersema J.H., Barrie F.R., Greuter W., Hawksworth D.L., Herendeen P.S., Knapp S., Kusber W.-H., Li D.-Z., Marhold K., May T.W., McNeill J., Monro A.M., Prado J., Price M.J., Smith G.F. (2018) International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017. Koeltz Botanical Books, Germany. doi: 10.12705/Code.2018

Ugalde P.C., McRostie V.B., Gayo E.M., García M., Latorre C., Santoro C.M. (2021) 13,000 years of sociocultural plant use in the Atacama Desert of northern Chile. Veg. Hist. Archaeobot. 30: 213–230.

Uribe M., Angelo D., Capriles J., Castro V., De Porras M.E., García M., Gayo E., González J., Herrera M.J., Izaurieta R., Maldonado A., Mandakovic V., McRostie V., Razeto J., Santana F., Santoro C., Valenzuela J., Vidal A. (2020) El Formativo en Tarapacá (3000-1000 aP): Arqueología, naturaleza y cultura en la Pampa del Tamarugal, Desierto de Atacama, norte de Chile. Lat. Am. Antiq. 31(1): 81–102.

Vilardi J.C., Freilij D., Ferreyra L.I., Gómez-Cendra P. (2021) Ecological phylogeography and coalescent models suggest a linear population expansion of *Anastrepha fraterculus* (Diptera: Tephritidae) in southern South America, Biol. J. Linn. Soc. 133(3): 779–795. doi: 10.1093/biolinnean/blab029

Wares J.P., Goldwater D.S., Kong B.Y., Cunningham C.W. (2002) Refuting a controversial case of a human-mediated marine species introduction. Ecol. Lett. 5:577–584. doi: 10.1046/J.1461-0248.2002.00359.X

Yan M., Xiong Y., Liu R., Deng M., Songm J. (2018) The Application and Limitation of Universal Chloroplast Markers in Discriminating East Asian Evergreen Oaks. Front. Plant Sci. 9:569. doi:10.3389/fpls.2018.00569

FINANCIAL SUPPORT

This research was funded by Agencia Nacional de Investigación y Desarrollo, ANID, Chile (projetc FONDECYT 11180805 given to V McR), Núcleo Milenio, AFOREST (projetc NCS2022-24 given to V McR), Agencia Nacional de Promoción Científica y Tecnológica, ANPCYT, Argentina (project PICT 2020-1402 given to CB), and Universidad de Buenos Aires, Argentina, UBA (project UBACYT 20020190200106BA given to CB).

AUTHOR CONTRIBUTION

CB: data, conceptualization, formal analysis, funding acquisition, investigation, writing original draft; **CP:** formal analysis, methodology; **RF:** data curation; **BOS:** conceptualization, formal analysis, investigation, writing original draft; **CS:** funding acquisition, supervision; **V McR:** funding acquisition, investigation, supervision; **JCV:** conceptualization, formal analysis, investigation, writing original draft.

BAG | Journal of Basic & Applied Genetics

Vol XXXV (2); 27-33; July-December 2024

Journal of the Argentine Society of Genetics

VARIANTE DEL GEN *KMT5B* Y LA DISCAPACIDAD DEL DESARROLLO INTELECTUAL AUTOSÓMICO DOMINANTE 51. REPORTE DE UN CASO Y CONSIDERACIONES DE LA INTERPRETACIÓN CLÍNICA

VARIANT OF THE *KMT5B* GENE AND THE DISABILITY OF AUTOSOMAL DOMINANT INTELLECTUAL DEVELOPMENT 51. CASE REPORT AND CONSIDERATIONS OF THE CLINICAL INTERPRETATION

Ratti S.G.[1,2,3], Álvarez E.O.[2,3], Allende E.[2,3], Funes M.C.[2,3], Sacchi O.J.[4], Della Vedova M.C.[5,6], Mendoza G.[5,6], Marsá S.[5,6]

[1] Centro de Diagnóstico Genético, Laboratorios Puntanos SE, Ministerio de Salud, San Luis, Argentina.

[2] Laboratorio de Epigénesis y Neuropsicofarmacología Experimental, Facultad de Ciencias Médicas (FCM), Universidad Católica de Cuyo (UCCuyo), San Luis, Argentina.

[3] Cátedra de Genética Humana, FCM, UCCuyo, San Luis, Argentina.

[4] IMBECU-CONICET, Mendoza, Argentina.

[5] GENES, Laboratorio de Genética y Biología Molecular, San Luis, Argentina.

[6] Facultad de Química, Bioquímica y Farmacia, Universidad Nacional de San Luis, Argentina.

Corresponding author:
Silvia G. Ratti
silratti@gmail.com

ORCID 0000-0002-8802-217X

Cite this article as:
Ratti S.G., Álvarez E.O., Allende E., Funes M.C., Sacchi O.J., Vedova D.C., Mendoza G., Marsá S. 2024. VARIANT OF THE *KMT5B* GENE AND THE DISABILITY OF AUTOSOMAL DOMINANT INTELLECTUAL DEVELOPMENT 51. CASE REPORT AND CONSIDERATIONS OF THE CLINICAL INTERPRETATION. BAG. Journal of Basic and Applied Genetics XXXV (2): 27-33.

Received: 07/25/2024
Revised version received: 10/14/2024
Accepted: 11/25/2024
General Editor: Elsa Camadro
DOI: 10.35407/bag.2024.35.02.02
ISSN online version: 1852-6233

Available online at
www.sag.org.ar/jbag

ABSTRACT

The intellectual developmental disorder (IDD) is a disruption of the neurodevelopment which is characterized by a decrease in the ability for reasoning and understanding abstract and complex information. IDD can occur either isolated or in conjunction with other neurodevelopmental disorders such as autism, motor disruptions, sensorial disruptions, dreaming and feeding disruptions. The objective of the present communication is to expand the knowledge about the phenotype-genotype relationship so far scarcely described in variants of the *KMT5B* gene, and also to contribute to the understanding of the clinical implications of neurodevelopmental disorders. A clinical case is presented of a female child referred for medical consultation due to maturational delay. Her medical history included a prenatal echography diagnostic of intrauterine growth delay. During physical examination a cephalic perimeter of -2 SD and slight dysmorphias were observed as positive signs. In addition, her reddish hair color, very white skin, hypoplasia of the nails, and an irregular border café-au-lait spot on the left iliac fossa were particularly noticeable. A molecular genetic study for genes involved in autism spectrum disorder was requested. A variant of uncertain significance was detected in the *KMT5B* gene. The phenotype presented by the patient contributes to the so far scarce description of the genotype-phenotype relationship of *KMT5B* gene variants. It is proposed to extend the phenotypic characterization of the gene by including information on the alteration in epigenetic regulation and neighbor genes that share the locus of the gene under study.

Key words: autism, dominant, intellectual disability in development, *KMT5B* gen, neurodevelopment

RESUMEN

La discapacidad intelectual del desarrollo es una perturbación del neurodesarrollo que se caracteriza por una disminución en la capacidad de razonar y de comprender una información abstracta y compleja. Puede presentarse aislado o en conjunto con otros trastornos del neurodesarrollo como el autismo o alteraciones motoras, sensoriales, del sueño y/o de la alimentación. El objetivo de la presente comunicación es ampliar el conocimiento de la relación fenotipo-genotipo descrita escasamente hasta el momento sobre la variante de significado incierto detectada en el gen *KMT5B*, y aportar a la comprensión sobre las implicancias clínicas de los trastornos del neurodesarrollo. Se presenta el caso clínico de una niña que fue derivada a la consulta médica por retraso madurativo. Como antecedente, en el diagnóstico ecográfico prenatal se detectó retardo de crecimiento intrauterino. Al examen físico, como signos positivos, se destacaron el perímetro cefálico Pc – 2,5 DS y dismorfias leves, y llamó la atención el color de cabello rojizo, la piel muy blanca, hiponiquia y una mancha de color café con leche en la fosa ilíaca izquierda. Se solicitó estudio molecular génico de panel de genes para trastornos del espectro autista. Se detectó una variante de significado incierto en el gen *KMT5B*. El fenotipo que presenta la paciente contribuye a la escasa descripción hecha hasta el momento de la relación genotipo-fenotipo de variantes del gen *KMT5B*. Se propone ampliar la interpretación fenotípica incluyendo alteración en la regulación epigenética y la revisión de los genes que comparten el locus del gen en estudio.

Palabras clave: autismo, dominante, discapacidad intelectual del desarrollo, gen *KMT5B*, neurodesarrollo

INTRODUCCIÓN

La discapacidad intelectual (DI) o discapacidad intelectual del desarrollo (DID) es una de las inhabilidades del desarrollo neurológico más comunes en el mundo y se caracteriza por una disminución en la capacidad de razonar y de comprender una información abstracta y compleja. Puede presentarse aislada o conjuntamente con otros trastornos del neurodesarrollo como el autismo, trastornos motores, sensoriales, del sueño y/o de la alimentación. También puede presentarse con epilepsia o trastornos psicopatológicos como la ansiedad, depresión y desregulación emocional (Des Portes, 2020).

En la actualidad, más del 50% de la etiología puede identificarse de origen genético (Des Portes, 2020). Sin embargo, en la mayoría de los casos no se puede establecer una única relación completa entre el fenotipo y el genotipo estudiado en cada uno de los pacientes. Es de suma importancia en la consulta genética clínica hacer una semiología exhaustiva, considerando todo el fenotipo del paciente y aumentar los recursos diagnósticos y la observación del probando con elementos que pueden sumarse a la práctica clínica habitual.

Aunque se identifique la variante génica que se relaciona con la situación clínica del paciente con DI, la variabilidad de las presentaciones fenotípicas requiere también razonamiento clínico molecular. Es preciso tener en cuenta el ambiente molecular del gen alterado para poder asociar al mismo con toda la clínica de un paciente, como por ejemplo considerar cambios epigenéticos o el locus en el que se encuentra el gen, que, aunque en el terreno de lo hipotético, puede explicar el fenotipo del propósito. Cabe aclarar que el concepto de ambiente se menciona desde la perspectiva de la Teoría General de los Sistemas de Ludwing von Bertalanffy (von Bertalanffy, 1950).

El gen *KMT5B* ha sido relacionado con el trastorno del desarrollo intelectual, autosómico dominante 51 (Faundes et al., 2018), y el producto del gen con una metiltransferasa que trimetila la histona nucleosomal H4 en la lisina 20 (Schotta et al., 2004). Por lo tanto, es objetivo de la presente comunicación ampliar el conocimiento sobre la relación fenotipo-genotipo descrita escasamente hasta el momento sobre variantes del gen *KMT5B* y aportar a la comprensión sobre las implicancias clínicas de los trastornos del neurodesarrollo.

CASO CLÍNICO

Se presentó a la consulta genética una niña de seis años de edad derivada por retraso madurativo. Entre los antecedentes se destacan: embarazo inducido con tratamiento hormonal, retardo de crecimiento intrauterino detectado en el diagnóstico ecográfico durante el segundo trimestre, absceso renal de la madre que motivó amenaza de parto prematuro (sic). No se refirió tratamiento antibiótico. La niña nació por cesárea a las 36 semanas de gestación con presentación cefálica y un Apgar 8/9. Las medidas antropométricas estuvieron dentro de rangos normales.

Al momento de la consulta las medidas antropométricas fueron: peso en percentil (Pc) 50, talla Pc +3 DS y perímetro cefálico −2,5 DS. Como datos positivos de la semiología se destacaron: braquicefalia, implantación del cabello anterior alta, cabello rizado y pelirrojo, ausencia de remolino central del cabello, frente alta y bombé, hendiduras palpebrales cortas inclinadas hacia arriba, presencia de epicantos, plenitud orbitaria, hipertelorismo ocular, comisuras labiales inclinadas hacia abajo, labio superior e inferior gruesos, paladar alto, orejas con rotación posterior, forma de dedos ahusada, hiponiquia, pigmentación de la piel muy blanca y una mancha café con leche en la zona de la fosa ilíaca izquierda.

El desarrollo psicomotor de la niña fue sedestación a los 18 meses y marcha a los 26 meses. Presentó trastornos del lenguaje con glosolalia, no le gustaban los ruidos fuertes y no se pudo integrar a sus compañeros de clase. La niña mostró en la consulta dominancia manual izquierda, que se constató solicitándole que realice un dibujo en forma libre y espontánea.

Con respecto al genograma, se consideraron como datos relevantes: un primo paterno de la propósito con trastorno del espectro autista (TEA), y un primo segundo paterno de la propósito también con TEA. Ambos familiares de la misma friatría (Figura 1). En las evaluaciones diagnósticas complementarias previas, la audición y la visión con estudio de fondo de ojo fueron normales.

MATERIALES Y MÉTODOS

Se solicitó estudio genético molecular de panel de genes para TEA y tomografía computada de cerebro (TAC) (Figura 2).

El ADN genómico fue extraído a partir de leucocitos de sangre periférica. La librería de ADN se preparó mediante el método de enriquecimiento, seguido de la captura de las regiones objetivo con los kits xGen Exome Research Panel v2, xGenCNV Backbone Panel y xGen Human mtDNA Research Panel. La secuenciación se realizó en la plataforma Illumina (NextSeq o Novaseq). La región objetivo incluyó el ADN codificante (CDS), los sitios de *splicing* y el genoma mitocondrial completo. La cobertura vertical promedio esperada fue 100x, la cobertura horizontal fue mayor o igual al 97,5% a 10x y mayor o igual al 95% a 20x. El análisis bioinformático

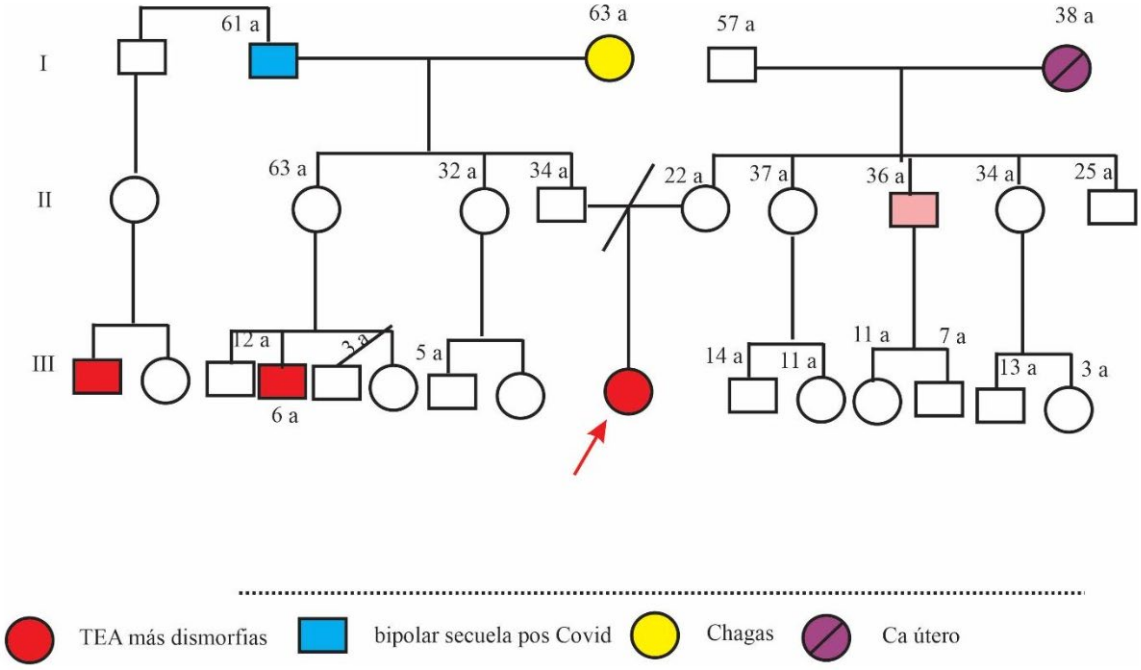

Figura 1. Genograma de paciente de seis años.

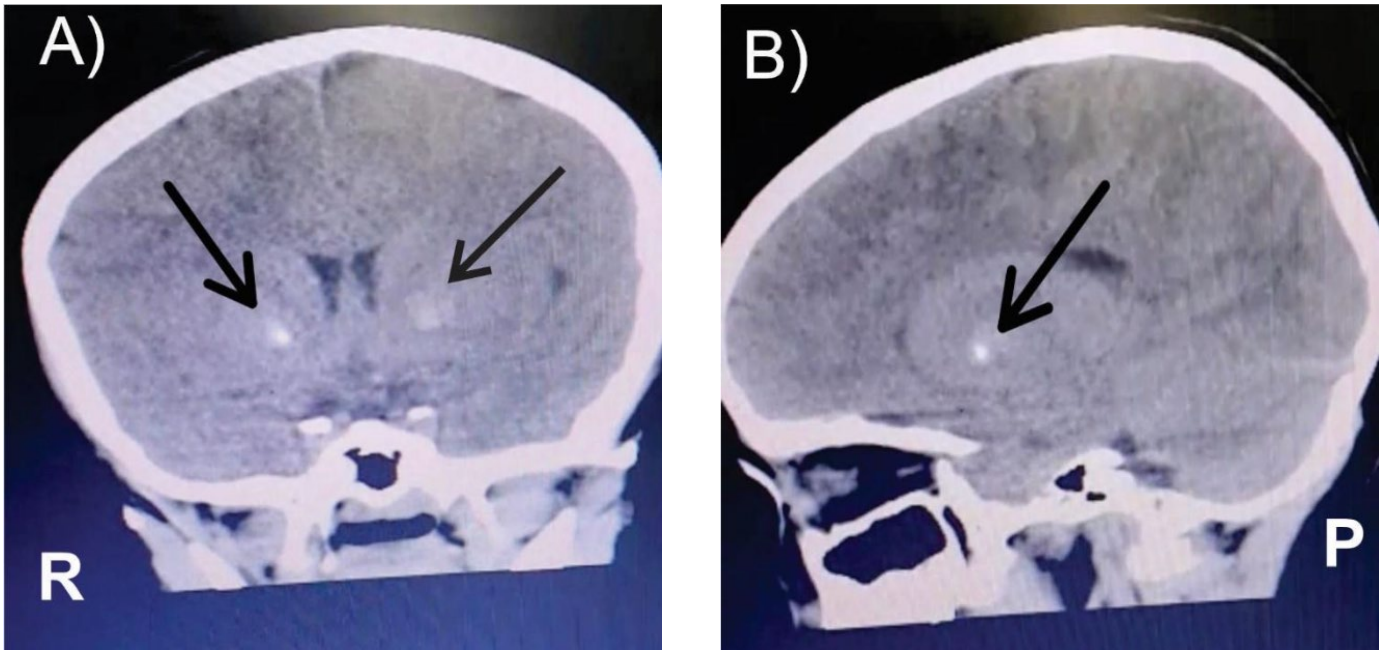

Figura 2. Tomografía Axial Computada (TAC) de niña de seis años sin contraste: (A) corte coronal y (B) corte sagital. En (A) las flechas señalan las zonas neuronales de interés. Imagen hiperdensa a nivel talámico bilateral de distribución simétrica y bordes poco definidos. En (B) la flecha señala zona hiperdensa única por debajo de la porción anterior de cuerpo calloso, de forma redondeada y bordes definidos.

se realizó utilizando la plataforma *Dragen Enrichment* (Illumina, Inc.) y Emedgene, basada en la versión GRCh38 del genoma humano.

Como información adicional complementaria se solicitó a la paciente la realización de puño y letra de un dibujo, temática libre. El bosquejo, cuya expresión refleja el estado interno emocional y espontáneo, fue analizado posteriormente por aproximaciones psicobiométricas establecidas (Jung, 1977; Békei, 1984; Di Scala y Pistoia, 1996).

Se solicitó el consentimiento informado a los padres para la publicación de este caso clínico.

RESULTADOS

En el estudio de la TAC se encontraron calcificaciones supratentoriales en los núcleos de la base bilaterales y simétricas (Figura 2 A y B).

El estudio molecular informó la variante c.1163A>C, p.(Lys388Thr), identificada en heterocigosis en el gen *KMT5B*, que resulta en la sustitución de un aminoácido en la proteína codificada y, por lo tanto, cumple con las especificaciones para calificar los criterios del *American College of Medical Genetics* (ACMG). Esta variante está ausente en las bases de datos de población gnomAD, 1000Genomes, ABraOM y en la base ClinVar. Actualmente está clasificada como una variante de significado incierto (VUS por sus siglas en inglés) (Tabla 1).

El genograma de la paciente se muestra en la Figura 1 y los dibujos realizados durante la consulta se muestran en la Figura 3 A-D.

DISCUSIÓN

Las características fenotípicas de esta paciente se ajustan a la escasa descripción que se encuentra en la literatura sobre la expresión del gen *KMT5B* (Iossifov et al., 2014). Ya se discutió en una publicación anterior (Ratti et al., 2024) que la clasificación de la variante de "significado incierto" podría ser imprecisa ya que las bases de datos que se utilizan para su clasificación no tienen representatividad de nuestras poblaciones. No se han publicado hasta el momento pruebas funcionales del gen. Se ha descrito con respecto a los trastornos de deficiencia intelectual de origen genético que la mayoría de las variantes son *de novo*, aunque estas mutaciones ocurrirían tres veces más frecuente en el entorno paterno que en el materno (Iossifov et al., 2014). El genograma muestra dos miembros afectados con TEA, de la misma friatría que nuestra paciente, procedente de la rama familiar paterna (Figura 1). Se ha solicitado el estudio molecular del gen *KMT5B* a los padres de la niña, pero hasta el momento no se dispone de los resultados, por lo que ambas posibilidades pueden ser correctas en nuestro caso.

Otras características fenotípicas que no se han descrito, están presentes en nuestra paciente. La pigmentación de la piel extremadamente clara, el cabello rojizo y la hiponiquia recuerda que la epidermis, las uñas, el pelo y el sistema nervioso tienen un origen embrionario en común ectodérmico. La mayoría de los hallazgos genéticos con respecto al color rojizo del cabello se ha estudiado en poblaciones europeas desde la perspectiva de la genética poblacional (Aiger et al., 2017). No se ha encontrado descripción del color del cabello rojizo desde un enfoque de la genética médica. Ningún otro miembro que se conozca de esta familia presenta o presentó el cabello rojizo y el color de la piel que se ha referido. Una explicación podría ser razonable si se considera que en el análisis molecular solicitado no se estudiaron todos los genes descritos hasta el momento en el locus 11q13.3. En este locus además del gen *KMT5B* se encuentran otros como el gen *KRN1* y *KRTAP5-9* asociados a la queratina (MacKinnon et al., 1991). También se encuentra el gen *FGF4* del que se describió que juega un papel en el bucle de retroalimentación con el gen *Sonic Hedgehog* que se expresa tempranamente en el desarrollo en una cascada de señalización que hasta el momento se describe para el desarrollo de las extremidades (Zuñiga et al., 1999). No necesariamente estos genes deberían presentar variantes, al no haberse secuenciado el locus completo no es posible saberlo, pero una alternativa en la interpretación del fenotipo de nuestra paciente es que una variante en un gen, en nuestro caso el gen *KMT5B* podría modificar un regulador génico y/o epigenético y alterar, parcialmente, la expresión de genes que comparten el mismo locus. El gen *KMT5B,* al expresar para una metiltransferasa que trimetila específicamente la histona nucleosomal H4 en la lisina 20, K20 (Schotta et al., 2004), permite especular un cambio epigenético que pudiera comprometer la expresión de más de un gen en el mismo locus. Aunque en la literatura científica no se ha encontrado, se puede discurrir razonablemente que algún gen de expresión en el desarrollo del ectodermo contemplaría las características fenotípicas de nuestra paciente.

Por otro lado, las calcificaciones supratentoriales descritas en la TAC sumadas a la mancha café con leche, la pigmentación de la piel, cabello e hiponiquia, podrían categorizarse como una facomatosis no descrita hasta el momento. Aunque esta última categorización parece más lejana.

Tabla 1. Resultados del estudio molecular en el gen *KMT5B* de la paciente de seis años.

Gen	Coordenada genómica	Transcripto	Alteración en la secuencia nucleotídica	Alteración en la secuencia aminoacídica	dbSNP	Clasificación
KMT5B	chr11:68166993	NM_017635.5	c.1163A>C	p.(Lys388Thr)	rs-	Significado incierto

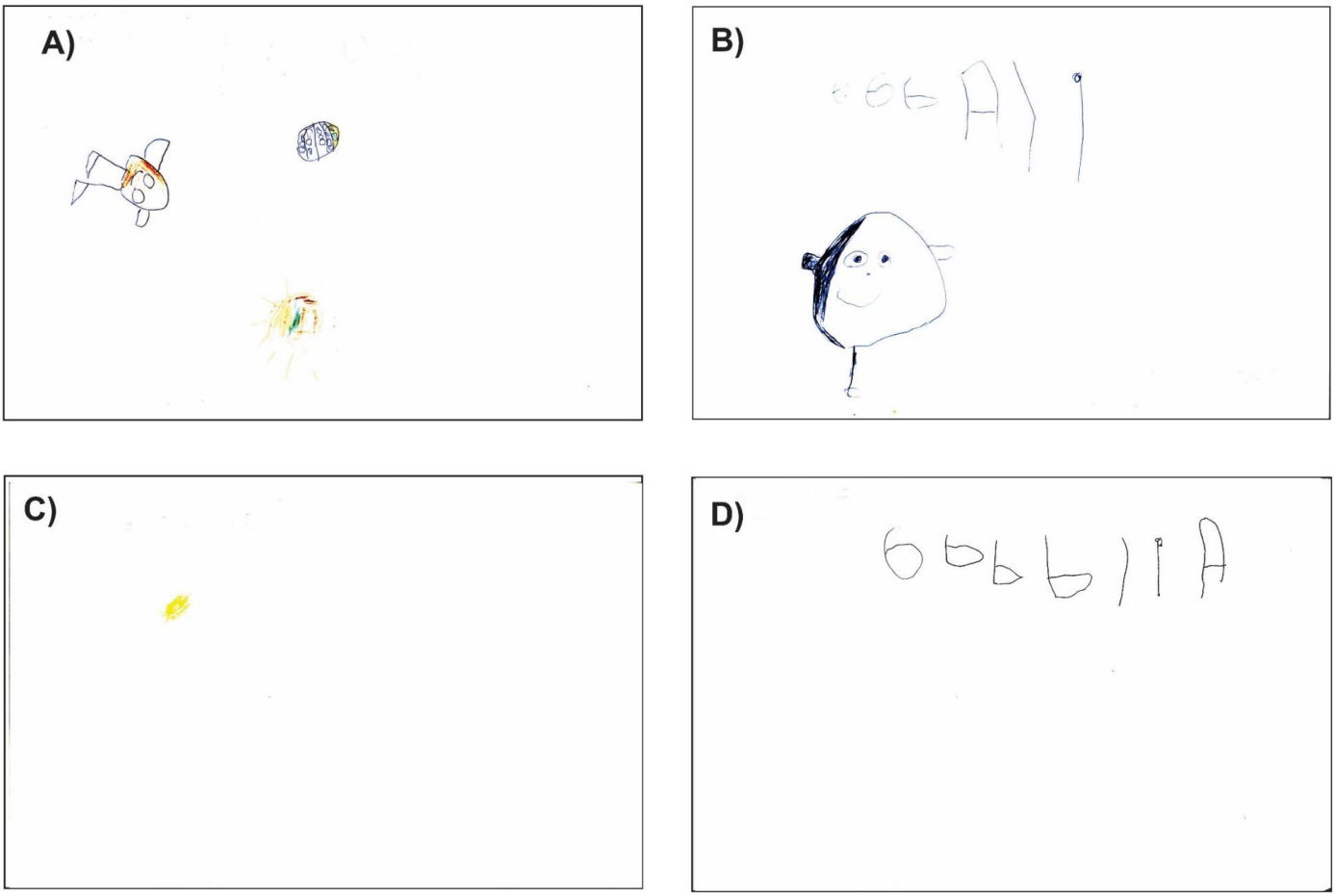

Figura 3. Dibujos espontáneos realizados por la paciente de seis años en hojas A4 (29,7 cm ancho; 21 cm alto) durante la consulta médica. Todos los dibujos fueron hechos en este tamaño y las figuras los muestran en la misma proporción que los originales. Los distintos temas elegidos por la paciente se han identificado como (A), (B), (C), y (D). En (A) la interpretación psicobiométrica señala carencia de sincronización entre lo mental, emocional y el cuerpo. En (B) la interpretación psicobiométrica señala problemas de lateralización funcional, posiblemente por ausencia de desarrollo integral del sistema nervioso central. En (C) la interpretación psicobiométrica señala carencia de sociabilidad y atención con el medio que la rodea. En (D) la interpretación psicobiométrica señala problemas en la sincronización central motora y efector periférica.

Llamó la atención que tres signos que presentó la paciente conciernen a la lateralidad, en este caso vinculada al lenguaje: la dominancia manual izquierda (Francks et al., 2003), la dislexia (Di Scala y Pistoia, 1996) y la glosolalia (McManus, 1985). Se ha observado desde hace bastante tiempo que aumenta la proporción del uso preferencial de la mano izquierda en pacientes con desórdenes neuronales como el autismo (Cornish y McManus, 1996) y la discapacidad intelectual (Grouios et al., 1999). También se han descrito genes cuyas mutaciones se asocian con fenotipos vinculados a la dominancia manual izquierda y la dislexia como el gen *LRRTM1* (Leach et al., 2014) y el gen *PCSK6* (Scerri et al., 2011). La asociación pérdida de la dominancia para el uso de la mano derecha con trastornos cognitivos vinculados a la dislexia y falta de la asimetría cerebral fue investigada por los autores hace algunos años en el fenotipo asociado al gen *HSR* (Ratti et al., 2007, 2010). Se ha propuesto que la pérdida de la asimetría cerebral podría ser a expensas de la alteración de la arquitectura del área del lenguaje de Brocca que es el área que se propone como una adquisición evolutiva en la especie humana en el desarrollo del lenguaje (Ratti et al., 2007). Hasta el momento no se ha descrito al gen *KMT5B* relacionado con la lateralidad.

Es de notar que los genes *HSR* y *LRRTM1* están regulados en su expresión por el mecanismo epigenético de impronta génica. También es posible que ambos genes sean el mismo, ubicados en el locus 2p12. El gen *KMT5B,* como ya se mencionó anteriormente, también está vinculado a la regulación epigenética. Se ha encontrado que la alteración epigenética del gen *HSR* está fuertemente asociada a la pérdida de la metilación en el genoma entero (Ratti et al., 2007), por lo que es razonable proponer que una alteración en un gen como el *KMT5B* podría comprometer la configuración de la

cromatina e involucrar en la expresión, alteraciones en la regulación de otros genes distantes del gen *KMT5B*, incluyendo también otros genes relacionados con la producción de melanina como *MC1R* (Valverde et al., 1995), *OCA2* (Chiang et al., 2008) o *SLC24A5* (Lamason et al., 2005) que completarían aún más las características fenotípicas de nuestra paciente. No se conoce hasta el momento la arquitectura tridimensional del genoma y si esta se modifica a cada instante. Es posible que la descripción fenotípica del gen *KMT5B* aún no esté completa en toda la variabilidad de su expresión, y se espera que esta comunicación pueda contribuir a la misma.

Se desea destacar el aporte valioso que se obtiene al solicitar a los pacientes que realicen durante la consulta un dibujo. En el caso aquí descrito, se ha podido identificar por la interpretación psicobiométrica del mismo, alteración de lateralidad y pérdida de habilidad cognitiva vinculada a la dislexia, aspectos que, sin esta herramienta no se hubieran considerado (Jung, 1977; Békei, 1984; Di Scala y Pistoia, 1996; D'Alfonso y Biedma, 1960; Figura 3); además, se identificó la dominancia manual izquierda al dibujar. Inferir la lateralidad como parte de la expresión del gen *KMT5B*, permite razonar una expresión temprana en el neurodesarrollo de este gen.

Finalmente, sería interesante que los bioinformáticos tengan entrenamiento médico. Habitualmente trabajan en laboratorios sin vinculación con los médicos solicitantes, por lo que sus búsquedas están sesgadas, limitadas y sin compromiso con el diagnóstico de los pacientes.

AGRADECIMIENTOS

Los autores desean agradecer la participación de la Lic. Marta Tambone en la interpretación psicobiométrica de los dibujos.

CONFLICTO DE INTERESES

Los autores declaran ausencia de conflicto de intereses.

BIBLIOGRAFÍA

Ainger S., Jagirdar K., Lee K., Soyer H. (2017) Skin Pigmentation Genetics for the Clinic. Dermatology 233(1):1-15.

Békei M. (1984) Trastornos psicosomáticos en la niñez y la adolescencia. Editorial Nueva Visión, Buenos Aires, Argentina.

Cornish K., McManus I. (1996) Hand preference and hand skill in children with autism. J. Autism Dev. Disord. 26: 597-609.

Chiang P.W., Spector E., Tsai A.C.H. (2008) Evidence suggesting the inheritance mode of the human P gene in skin complexion is not strictly recessive. Am. J. Med. Genet. Part A 146(11): 1493-1496.

D'Alfonso P., Biedma C. (1960) El Lenguaje del Dibujo, Editorial Kapelusz, Buenos Aires, Argentina.

Des Portes V. (2020) Intellectual disability. Handbook of Clinical Neurology 174(9): 113-126.

Di Scala M., Pistoia M.TLI. (1996) Test psicopedagógico de lectoescritura, Departamento de Publicaciones, Facultad de Psicología, UBA, Buenos Aires, Argentina.

Faundes V., Newman W., Bernardini L. (2018) Histone Lysine Methylases and Demethylases in the Landscape of Human Developmental Disorders. Am. J. Hum. Genet. 102(1):175-187.

Francks C., DeLisi L.E., Shaw S.H., Fisher S.E., Richardson A.J., Stein J.F., Monaco A. (2003) Parent-of-origin effects on handedness and schizophrenia susceptibility on chromosome 2p12-q11. Hum. Mol. Genet. 12: 3225-3230.

Grouios G., Sakadami N., Peoderi A., Alevriadou A. (1999) Excess of non-right handedness among individuals with intellectual disability: Experimental evidence and possible explanations. J. Intell. Disabil. Res. 43: 306-313.

Iossifov I., O'Roak B., Sanders S. *et al.* (2014) The contribution of *de novo* coding mutations to autism spectrum disorder. *Nature* 515: 216–221.

Jung C. (1977) Psicología y Simbólica del Arquetipo. 1° Edición. Editorial Paidos, Buenos Aires, Argentina.

Lamason R., Manzoor-Ali P., Mest J., Wong A., Norton H., Aros M., Jurynec M., Mao X., Humphreville V., Humbert J., Sinha S., Moore J., Jadadeeswaran P., Zhao W., Ning G., Akalowska I., McKeigue P., O'Donnell D., Kittles R., Parra E., Mangini N., Grunwald D., Shriver M., Canfield V., Cheng K. (2005) SLC24A5, a Putative Cation Exchanger, Affects Pigmentation in Zebrafish and Humans. Science 310 (5755): 1782-1786.

Leach E., Prefontaine G., Hurd P. (2014) The imprinted gene LRRTM1 mediates schizotypy and handedness in a nonclinical population. J. Hum. Genet. 59: 332-336.

MacKinnon P., Powell B., Rogers G., Baker E., MacKinnon R. (1991) An ultrahigh-sulphur keratin gene of the human hair cuticle is located at 11q13 and cross-hybridizes with sequences at 11p15. Mammalian. Genome 1:53-56.

McManus I.C. (1985) Handedness, language dominance and aphasia: a genetic model. Psychol. Med. 8: 3-40.

Ratti S., Álvarez Toro E., Della Vedova C., Mendoza G., Marsá S. (2024) Hallazgo de variantes genómicas en un paciente con proceso neurodegenerativo. Discusión sobre la relación genotipo-fenotipo. Genética Médica y Genómica 8: 37-43.

Ratti S., Cordoba P., Rearte S., Álvarez E. (2007) Differential expression of Handedness Scalp Hair-whorl Direction and cognitive abilities in primary school children. Int. J. Neuroprot. Neuroregen. 4 (1): 52-60.

Ratti S., Vizioli N., Álvarez E. (2010) Epigenetic modulation expressed as methylation changes in DNA from primary school children of two different geographical environments. II. Am. J. Neuroprot. Neuroregen. 2: 65-70.

Scerri T., Brandler W., Parracchini S., Morris A., Ring S., Richardson A., Talcott J., Stein J., Mónaco A. (2011) PCSK6 is associated with handedness in individual with dyslexia. Human Mol. Gen. 20 (3): 608–614.

Schotta G., Lachner M., Sarma K., Ebert A., Sengupta R., Reuter G., Reinberg D., Jenuwein T. (2004) A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev. 18(11):1251-62.

Valverde P., Healy E., Jackson I., Rees J.L., Thody A.J. (1995) Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. Nat. Genet. 11: 328-330.

von Bertalanaffy L. (1950) The theory of open systems in physics and biology. Science 111:23-9.

Zúñiga A., Haramis A., McMahon A., Zeller R. (1999) Signal relay by BMP antagonism controls the SHH/FGF4 feedback loop in vertebrate limb buds. Nature 401(6753):598-602.

BAG | Journal of Basic & Applied Genetics

Vol XXXV (2); 35-45; July-December 2024

Journal of the Argentine Society of Genetics

HOST IMMUNE RESPONSE TO SARS-CoV-2 INFECTION BY *TLR3*, *TLR4* AND *TLR7* GENE EXPRESSION THROUGH SEMI-QUANTITATIVE RETRO-TRANSCRIPTION PCR

RESPUESTA INMUNE DEL HUÉSPED A LA INFECCIÓN POR SARS-CoV-2 MEDIANTE LA EXPRESIÓN DE LOS GENES *TLR3*, *TLR4* Y *TLR7* A TRAVÉS DE RT-PCR SEMICUANTITATIVA

Martinez Marignac V.L.[1,2], Lirussi D.[3], Oertlin G.[1,2], Favant J.L.[1,2], Fleischman E.[2], Salinas M.[4], Marchetti G.[5], Gassali Z.[5], Richard S.M.[6]

[1] IBIOGEM, Laboratorio Interdisciplinario de Biología y Genética Molecular, Centro de Investigación Científica y de Transferencia Tecnológica a la Producción-CICYTTP, CONICET-UADER, Entre Ríos, Argentina.

[2] Laboratorio Dr. Néstor Bianchi, Hospital San José, Diamante, Entre Ríos, Argentina.

[3] Instituto Nacional de Medicina Tropical ANLIS-Malbrán, Puerto Iguazú, Misiones, Argentina.

[4] Laboratorio Provincial de Epidemiología, Ministerio de Salud, Entre Ríos, Argentina.

[5] Oficina Director – Hospital San José Diamante, Entre Ríos, Argentina.

[6] Instituto Multidisciplinario de Biología Celular (IMBICE), La Plata, Buenos Aires, Argentina.

Corresponding author:
Verónica L. Martinez
vmartinezmarignac@cicyttp.org.ar

ORCID 0000-0002-0486-7280

Cite this article as:
Martinez Marignac V.L., Lirussi D., Oertlin G., Favant J.L., Fleischman E., Salinas M., Marchetti G., Gassali Z., Richard S.M. 2024. HOST IMMUNE RESPONSE TO SARS-CoV-2 INFECTION BY *TLR3, TLR4* AND *TLR7* GENE EXPRESSION THROUGH SEMI-QUANTITATIVE RETRO-TRANSCRIPTION PCR. Journal of Basic and Applied Genetics XXXV (2): 35-45.

Received: 09/24/2024
Revised version received: 11/29/2024
Accepted: 12/11/2024
General Editor: Elsa Camadro
DOI: 10.35407/bag.2024.35.02.03
ISSN online version: 1852-6233

Available online at
www.sag.org.ar/jbag

ABSTRACT

Toll-like receptors (*TLRs*) involved in viral detection or overexpressed during infection, play also a main role in the subsequent development of fatal clinical manifestations in COVID-19 patients. We characterized by semi-quantitative retro-transcription PCR the expression of *TLR* 3, 4, and 7 in nasopharyngeal total RNA samples from 150 and 152 individuals, positive and negative for COVID-19, respectively. All patients were grouped in accordance with the presence of respiratory symptoms, and results on rapid molecular diagnostic test (NeoKit S.A.) during 2021. Four groups were analyzed: a) COVID-19 RNA detected with severe symptomatology; b) COVID-19 RNA detected with low/mild symptomatology; c) COVID-19 RNA not detected with severe symptomatology; and d) COVID-19 RNA not detected with low/mild symptomatology. Other variables studied were age and sex. Our results show (while correcting the sample size bias of previous reports), a non-significant difference in the expression of *TLR4*/7 between COVID-19 positive and negative patients. Noteworthy, the expression of *TLR3* was augmented in patients that resulted negative for COVID-19. When we compared the expression among the four groups, a significant positive correlation between severe symptomatology and *TLR4* expression in patients positive and negative for COVID-19 was found. Furthermore, *TLR4* gene expression was significantly amplified in those COVID-19 patients with severe symptomatology when compared with non-severe COVID-19 cases. Our data suggest that our innate *TLRs* immune system may respond differently to respiratory infections with similar symptomatology. We confirmed that mainly *TLR3* and *TLR4* could be involved in the response to respiratory pathogenesis and particularly *TLR4* in COVID-19 infection with severe symptoms. In concordance with previous studies, we found that in inflammatory respiratory diseases it is important to focus on *TLR3*, and *TLR4* gene expression to understand severe symptoms development. Research on these pathways may help to find modulators or antagonists to these genes, for future treatments.

Key words: gene expression, isothermal amplification, SARS-CoV-2 diagnoses, semi-quantitative retro-transcription PCR, toll-like receptor genes.

RESUMEN

Los receptores tipo Toll (*TLRs*) implicados en la detección de virus o sobreexpresados durante la infección, desempeñan un papel principal en el desarrollo de manifestaciones clínicas fatales en pacientes con COVID-19. Caracterizamos, mediante retro transcripción y PCR semi-cuantitativa, la expresión de los genes *TLR* 3, 4 y 7 en muestras de ARN nasofaríngeo de 150 individuos positivos y 152 negativos para COVID-19. Todos los pacientes fueron agrupados de acuerdo con la presencia de síntomas respiratorios y los resultados de diagnóstico molecular rápido (NeoKit S.A.) durante 2021. Se analizaron cuatro grupos: a) COVID-19 detectado con sintomatología grave; b) COVID-19 detectado con sintomatología leve/mínima; c) COVID-19 no detectado con sintomatología grave; y d) COVID-19 no detectado con sintomatología leve/mínima. Otras variables estudiadas fueron la edad y el sexo. Los resultados muestran, corrigiendo el sesgo de tamaño de muestra de informes previos, una diferencia no significativa en la expresión de *TLR4/7* entre pacientes positivos y negativos para COVID-19. Se destaca la expresión de *TLR3* incrementada en pacientes que resultan negativos para COVID-19. Al comparar la expresión entre los cuatro grupos, se encontró una correlación positiva significativa entre la sintomatología grave y la expresión de *TLR4* en el total de pacientes. La expresión del gen *TLR4* se amplificó significativamente en los pacientes con COVID-19 con sintomatología grave en comparación con los casos de COVID-19 no graves. Nuestros datos sugieren que el sistema inmune innato mediado por *TLRs* podría responder de manera diferente a infecciones respiratorias aun con sintomatología similar. Confirmamos que *TLR3* y *TLR4* podrían estar involucrados en la respuesta a la patogénesis respiratoria, y particularmente *TLR4* en infecciones por COVID-19 con síntomas graves. Concluimos que es importante evaluar la expresión de los genes *TLR3* y *TLR4* para comprender el desarrollo de síntomas graves en enfermedades respiratorias; esto ayudará a identificar moduladores o antagonistas de estos genes para futuros tratamientos.

Palabras clave: expresión génica, amplificación isotérmica, diagnóstico de SARS-CoV-2, PCR semi-cuantitativa de retro-transcripción, genes de receptores tipo Toll.

INTRODUCTION

Toll-like receptors (*TLR*s) are a family of receptors comprising 10 members (*TLR1-TLR10*), which are expressed in innate immune cells such as macrophages, as well as in epithelial and fibroblast cells. These have been reported as a cause in the initial failure of viral clearance (O'Neill et al., 2013) and it was suggested that they are involved in the subsequent development of fatal clinical manifestations in severe COVID-19 patients, related to fatal respiratory failure in acute respiratory distress syndrome (ARDS) (Onofrio et al., 2020). *TLR*s activation can be induced by a multitude of pathogen associated molecular patterns (PAMPs) present in bacteria, viruses, and other microorganisms (Duran et al., 2014). Hence, it was suggested that *TLR*s play an important role in priming innate immune responses, with production of inflammatory cytokines, like interleukins 4/6, (IL-4, IL6), IL-1β and other mediators (Bortolotti et al., 2021).

*TLR*s are expressed on the cell surface (*TLR* 1, 2, 4, 5, 6 and 10) or in the endosomal compartment (*TLR* 3, 7, 8 and 9) (O'Neill et al., 2013; Onofrio et al., 2020). While *TLR*3 in the endosome recognizes viral double-stranded RNA (dsRNA), *TLR*7 recognizes viral single-stranded RNA and is therefore suggested to be involved in the clearance of SARS-CoV-2 and similar viruses (Imai et al., 2008; Onofrio et al., 2020). Although COVID-19 viral entry is mainly via the interaction of spike glycoprotein with the angiotensin-converting enzyme 2 (ACE2) of the host cell, other interactions have been described. It is now well known that the spike protein can interact with immune ligands like TLRs and C-lectin-like receptors (CLR) as well as with non-immune receptors as neuropilin-1 (NRP1) and glucose regulated protein 78 (GRP78) for the viral entry (Gadanec et al., 2021).

An intact SARS-CoV-2 virus is surrounded by a lipid enclosure that contains the envelope protein (E), the membrane protein (M), and the spike glycoprotein (S). The genome of SARS-CoV-2 consists of large, single-stranded positive RNA (from 29.8 to 29.9 kb) that contains fourteen (14) open-reading frames (ORFs) encoding twenty-seven (27) proteins; it is possible to suggest that before packaging, COVID-19 RNA virions could be detected by TLR7 that is localized in the intracellular membranes of endosomes (Nishiya et al., 2005). The genome sequence of SARS-CoV-2 displays 79.0% homology with SARS-CoV and 51.8% with MERS-CoV. Nucleocapsid (N) proteins form complexes with genomic RNA for genome packaging. In consequence, it is plausible that *TLR7* and *TLR3* are molecules that could recognize single- and double-stranded viral RNA structures before packaging (Nishiya et al., 2005).

During viral entry into host cells, the surface trimeric S glycoprotein mediates receptor recognition and virus-host cell membrane fusion. The host protease furin cleaves the S protein into S1 and S2 subunits for pre-activation, and the receptor-binding domain (RBD) of S1 binds to ACE2 expressed on the surfaces of host cells. Then, SARS-CoV-2 enters host cells by either direct fusion or endocytosis (Jung and Lee, 2021).

In case of TLR4 expressed on the cell surface, it has been shown that it binds SARS-CoV-2 spike glycoprotein and this activates TLR4 signaling to increase cell surface expression of ACE2, then facilitating SARS-CoV-2 entry. This is in part the mechanism that destroys the type II alveolar cells, the ones that secrete pulmonary surfactants. As a result, all these factors that otherwise contribute to a physiological lung equilibrium, during COVID-19 infection promote ARDS and inflammation (Manik and Singh, 2022). Thus, we suggest that an available antagonist against TLR4 could prevent the onset of severe COVID-19 or other viral induced ARDS in symptomatic patients, and synergize with active antiviral therapy.

While TLR3 recognizes viral double-stranded RNA (dsRNA), TLR7 recognizes viral single stranded RNA and is therefore involved in SARS-CoV-2 clearance, as evidenced by its loss of function in some cases of severe COVID-19 (van der Made et al., 2020). On the other hand, TLR4 at the surface of cells plays a role in the induction of damaging inflammatory responses during acute viral infections as it functions as a sensor for damage-associated molecular patterns (DAMPs). These patterns include a wide variety of molecules released from injured or dying tissues as well as molecules actively released in response to cellular stress from intact cells. Besides these functions in viral infections, the paramount role of TLR4 is the recognition of lipopolysaccharides (LPS) in Gram-negative bacteria (Poltorak et al., 1998) which could explain why TLR4 is not only associated to COVID-19 infection or ARDS.

It is well known that *TLR*s family members play also an important role in cancer progression. A comparison with cancer treatments is appropriate herein, as TLR stimulation of cancer cells can lead to tumor progression or inhibition. Stimulation of *TLR* 2, 4, and 7/8 can lead to tumor progression through production of immunosuppressive cytokines, increased cell proliferation, and resistance to apoptosis. On the other hand, stimulation of *TLR* 2, 3, 4, 5, 7/8 and 9 gene expression, often combined with chemo or immunotherapy, can lead to tumor inhibition through different pathways (Urban-Wojciuk et al., 2019). In addition, TLR stimulation in natural killer (NK) cells and APCs (antigen-presenting DC-dendritic cells and macrophages) can induce CTL (cytotoxic T lymphocytes) to further inhibit tumor growth (Urban-Wojciuk et al., 2019). In consequence, using cancer *TLR* insight could be an alternative approach to selectively activate some of these pathways in patients suffering a severe COVID-19 infection, for whom it could be a disadvantage to develop

the wrong cytokine storm (Chahal et al., 2013; Bezemer and Garssen, 2021). Thus, it is plausible to postulate that to decrease hyper-inflammation and thrombotic complications in vulnerable population (severe stages of COVID-19), a control of the expression of *TLR3*, *TLR4* and *TLR7* genes could be instrumental. However, there is a need of more evidence concerning the importance of the relative expression level of *TLR3*, *TLR4* and *TLR7* (Bortolotti et al., 2021; Manik and Singh, 2022) in ARDS and COVID-19 patients.

As the *TLR*s have been studied in the field of oncology, different anti-tumor treatments have been developed as activators, inhibitors or antagonists of these receptors (Urban-Wojciuk et al., 2019). However, there are few studies recently published showing the gene expression profile of some of these genes in COVID-19 infected patients (Chahal et al., 2013; Bezemer and Garssen, 2021) which were carried out on small cohorts of patients (with different symptomatology). Furthermore, there are also few studies relating treatment of COVID-19 by agonists of TLR7 and TLR8. For instance, one of these studies showed that *TLR7* stimulation may not only help viral clearance, but also exert an anti-inflammatory effect that decreases severe symptoms while synergizing with treatments (Khalifa and Ghoneim, 2021).

Here, our goal was to evidence *TLR* 3, 4, and 7 gene expression by semi-quantitative retro-transcription PCR in nasopharyngeal total RNA samples from patients admitted to the service of public hospitals of five departmental jurisdictions of the province of Entre Ríos, Argentina, during 2021.

MATERIALS AND METHODS

Patient groups (cohorts)

This study included a group of patients admitted during 2021 to the service of public hospitals in Diamante departmental jurisdiction and from other four departmental jurisdictions of the Province of Entre Ríos, Argentina. All patients showed symptomatology related to respiratory infections and at least two of the twelve symptoms were attributed to SARS-CoV-2 infection, according to the Health Ministry of Argentina and the World Health Organization (WHO) interim guidance.

Patients were divided into two main groups: 1- patients with active SARS-CoV-2 infection confirmed by a new isothermal amplification kit (developed by Cassara and Neokit S.A., Argentina) from nasopharyngeal swabs (DET, n=150) and, 2- SARS-CoV-2 negative patients and/or with SARS-CoV-2 IgG antibody-negative serology (ND, n=152). Disease severity was classified into three clinical types (low, mild/moderate and severe), according to the Health Ministry of Argentina and the WHO interim guidance. We have considered the following symptoms: 1=headache, 2=myalgia, 3=sore throat, 4=fever (no higher than 38 ºC), 5=fever (38 ºC or higher), 6=cough, 7=rhinitis/nasal congestion, 8=vomiting, 9=diarrhea, 10=respiratory failure, 11=anosmia, and 12=dysgeusia and death.

Patients with low symptomatology showed one or two of the symptoms 2, 8 or 9 (n=182), patients with mild/moderate symptomatology showed symptoms 3, 4 or 5 (n=76) and patients with severe symptomatology had symptoms 5, 6, 7, 10, 11 and 12 (n=44) (Table 1).

In brief, patients with low symptomatology did not present evidence of viral pneumonia or hypoxia. Symptoms were non-specific and could include separately fatigue, headache, myalgia, abdominal pain, vomiting and diarrhea. Mild/moderate symptomatology patients presented symptoms and signs of non-severe pneumonia, fever, sore throat (cough or fast breathing) and included patients with comorbidities such as cardiac or respiratory disease and/or diabetes. The cohort with severe symptomatology contained patients under severe disease, the majority with fever over 38 ºC, nasal congestion, respiratory failure, anosmia and dysgeusia. The characteristics of participants with or without COVID-19 are shown in Table 1.

The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board CEYSTE - Committee of Ethical and Experimental Work Safety (CEYSTE, CONICET), whose members reviewed and approved the informed consent, the sample collection and the overall study protocol (Expedient reference: CEYSTE-CE0112/2020).

COVID-19 determination

We used isothermal amplification by the COVID-19 NEOKIT-PLUS TECNOAMI, developed by CONICET and

Table 1. Characteristics of the cohorts of patients admitted to public hospitals in five jurisdictions of the Province of Entre Ríos with COVID-19 and miscellaneous respiratory infections. n=sample size; DET=COVID-19 detected infection; ND= COVID-19 non detected infection.

Patients Characteristics		DET (n=150)	ND (n=152)
Age/years	<2 baby	9	17
	<7 child	21	21
	< 18 teen	17	21
	< 36 young adult	49	40
	< 65 adult	47	46
	>65 old adult	8	7
Sex	female	70	77
	male	80	75
Symptomatology	low	91	91
	mild/moderate	40	36
	severe	19	25

Cassara S.A., Argentina. This diagnostic test is used for the simplified molecular detection of the SARS-CoV-2 virus (the etiological agent of atypical pneumonia or severe acute respiratory syndrome). The NEOKIT-PLUS system allowed the purification of total mRNA from patients swabs according to the manufacturer protocol. We obtained 50 µl of total RNA and used 10 µl to determine the presence of viral RNA.

Human cDNA

We used 5 µl of each total RNA sample purified using the NEOKIT PLUS system, which were stored at -80 °C. A random hexamer retro-transcription PCR method was performed by 1 µM of hexamer primers (PB-L SRL, Argentina), dNTPs 500 µM and MilliQ water in a final volume of 15 µl. This was exposed to thermo-cycling (5 min at 65 °C and 2 min on ice), and finally, the reaction was brought to a final volume of 20 µl by adding 50 mM of $MgCl_2$, 100 U of Moloney Murine Leukemia Virus (MMLV) reverse transcriptase, and 4 µl of MMLV buffer (PB-L, Argentina); thermo-cycling was performed at 25 °C for 10 min and 37 °C for 50 min.

TLR3, TLR4 and TLR7 gene expression determination

We characterized the expression of *B-actin*, *TLR* 3, 4, and 7 genes. The determination of gene expression was performed following the protocols of Allborn et al. (2008) and Ghasemi et al. (2014). We must underline that the selection of semi-quantitative retro-transcription PCR rather than quantitative PCR (qPCR), was because we focused on very high deltas in order to found more biological relevant or significant changes between cohorts.

We used 5 µl of cDNA in a semi-quantitative gene expression protocol. Human *B-actin* gene was used as a housekeeper gene and the amplification reaction was performed with 200nM of forward (GAGCACAGAGCCTCGCCTTT) and reverse (ACATGCCGGAGCCGTTGTC) primers, 1U of Taq polymerase (Pegasus-PB-L SA Argentina), 3 mM of MgCl2 and 200 uM of dNTPs; the thermocycling conditions were: 2 min at 94 °C; 40 cycles of 45 s at 92 °C, 45 s at 60 °C, and 45 s at 72 °C; and an end step of 5 min at 72 °C. In Table 2 we show the *TLRs* primers. The conditions of end point PCR used for *TLR* 3 and 4 were the following: 2 mM of MgCl2, 200 uM of dNTPs, 200 nM of primers and 1 U of Taq polymerase (Pegasus –PB-l SA, Argentina). The PCR conditions for *TLR7* were the following: 2.6 mM of MgCl2, 200 uM of dNTPs, 200 nM of primers and 1 U Taq polymerase. The thermocycling conditions for all the genes were the following: 2 min at 94 °C; 40 cycles of 30 s at 95 °C, 1 min at 62 °C, and 1 min at 72 °C; and an end step of 5 min at 72 °C. PCR products (10 µl of each sample) were separated by electrophoresis in a 2% agarose gel (Sigma) with 1x TAE buffer (Invitrogen) and a voltage of 95 V for 30-40 min. The bands were visualized by using an ultraviolet trans-illumination device and digital images were captured by Gel documentary machine (Care stream, Berlin, Germany).

Statistical analysis

For groups (DET vs. ND) and subgroups (sex, age, severity, gene expression) comparisons we used a nested analysis, non-paired t-Test and Spearman's Correlation Coefficient to evidence association between level of symptomatology and the expression of the three *TLRs* genes. We also used a Point-Biserial Correlation Calculator to evidence any correlation between two variables in the special circumstance of dichotomous variables (DET/ND; female/masculine); for this calculator there are only two possible values, 1 or 0. We used Pearson's correlation to measure the strength and direction of the relationship between two paired variables (age or severity vs. gene expression)

Table 2. List of primers used in the experiment for the amplification of human *B-actin, TLR* 3, 4 and 7 genes by semi-quantitative RT-PCR. Forward and reverse primer sequences, annealing temperature and product size (bp= base pairs).

Gene	Forward primer	Reverse primer	Annealing temperature (°C)	Product size (bp)
TLR3	GTATTGCCTGGTTTGTTAATTGG	AAGAGTTCAAAGGGGGCACT	62	112
TLR4	AAGCCGAAAGGTGATTGTTG	CTGAGCAGGGTCTTCTCCAC	62	114
TLR7	CCTTGAGGCCAACAACATCT	GTAGGGACGGCTGTGACATT	62	285
B-actin	GAGCACAGAGCCTCGCCTTT	ACATGCCGGAGCCGTTGTC	60	108

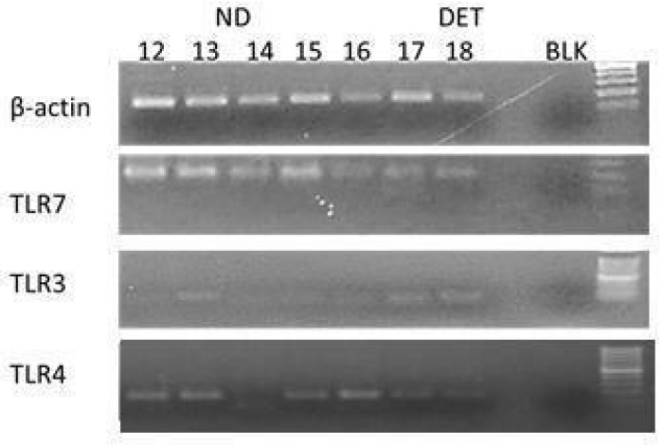

Figure 1. Characteristic results of the semi-quantitative retro-transcription PCR for *B-actin, TLR3, TLR4* and *TLR7* human RNAs. Agarose 2% gels showing characteristic results, BLK is retro-transcription and semi-quantitative PCR blank, the samples correspond to typical results from ND which are COVID-19 non detected samples and DET are COVID-19 detected samples.

and Spearman's Rho (RS) coefficient of correlation to compare the range of symptoms and other ordinal as well as continuous variables data, since ranks were used instead of assumptions of normality, especially for severity, sex and age. We used also a McNemar's test to evidence the strength of representation of each cohort for further analysis. We used a Welch's ANOVA for age and sex differences and gene expression. We further tested differences on sex and age by the Games-Howell post-hoc test from ANOVA as a nonparametric approach to compare combinations of cohorts without assuming equal variances and sample sizes.

We considered significant a p-value of 0.05 or lower. All statistical analyses were performed by the free access websites https://www.socscistatistics.com/tests/, https://www.graphpad.com/quickcalcs/ and https://www.biostathandbook.com/spearman.html.

RESULTS

By a low cost and efficient semi-quantitative retro-transcription and end point PCR, we analyzed on 2% agarose gels the gene expression of *B-actin*, *TLR3*, *TLR4* and *TLR7* (Figure 1).

From the collected nasopharyngeal swabs samples, we extracted RNA, and purified it using NeoKit Plus (SAS Argentina, SA). Only samples for which it was feasible to obtain very good retro-transcription and PCR product of *B-actin* gene were used for further analysis. By analyzing this important RNA collection of patients with respiratory disease symptoms sampled during 2021 and 2022, we evidenced that the expression of *TLR3*, *TLR4* and *TLR7* genes have no correlation with the presence of COVID-19 viral genome. This applies to the total cohort of patients by using the Point-Biserial Correlation calculator (Table 3, Figure 2). By applying the McNemar's test we did not evidence significant association between severe cases and COVID-19 presence in the patients, neither a bias in the representation of each used cohort for diverse comparisons (p>0.5). However, by applying a

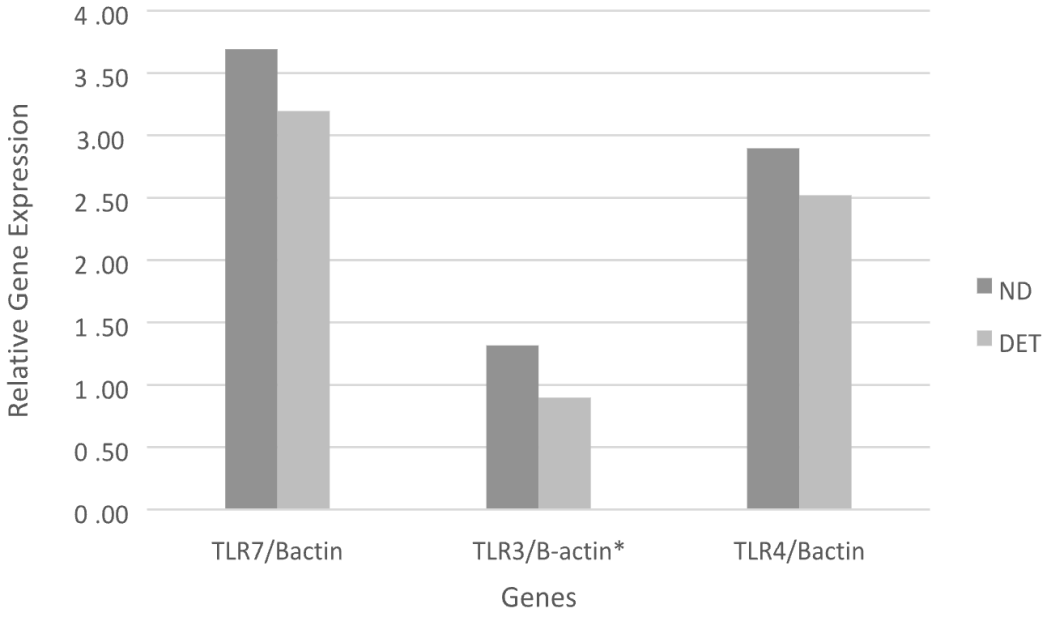

Figure 2. Gene expression by comparing COVID-19 Non-detected (ND) and Detected (DET) patients admitted to public hospitals in five jurisdictions of the Province of Entre Ríos with COVID-19 and miscellaneous respiratory infections.

Table 3. Semi-quantitative expression of *TLR3, TLR4* and *TLR7* normalized by housekeeping gene *B-actin* in two cohorts of patients: infected and not infected with COVID-19. SE = standard error, taking in account the sample size.

		TLR7/B-actin	*TLR3/B-actin*	*TLR4/B-actin*
COVID-19 non-detected	Average	3.69	1.32	2.90
	SE (±)	0.22	0.15	0.21
COVID-19 detected	Average	3.20	0.90	2.52
	SE (±)	0.19	0.10	0.17

non-paired t-Test, we found significant differences of expression for *TLR3*, in the 302 studied patients. Noteworthy, those who were infected with COVID-19 (n=150) had a significantly lower expression of *TLR3* (t=2.2, p=0.025) than the non infected ones (Figure 2).

When we took into account the severity of the disease, we observed non-significant differences between the cohorts and its *TLR3* or *TLR7* expression by ANOVA. For *TLR4*, we observed by ANOVA (followed by a Tukey's test) a low but significant difference between the gene expression of the cohorts (p=0.046).

When we analyzed the relationship between the severity of the symptoms and *TLR3* expression in the total sample (Table 4, Figure 3), we did not evidence a correlation between expression and level of symptomatology. On the other hand, *TLR4* expression showed a moderate positive and significant correlation in all patients with severe symptoms when we did not discriminate by other variables (Sperman's r_s=0.15054, *p* 2-tailed =0.00925). However, the differences in gene expression between low/mild and severe symptoms were non-significant (t-Test, p=0.054). The comparison of the four cohorts, DET low/mild/moderate, DET severe, NT low/mild/moderate and NT severe, revealed that *TLR4* expression was 1.5-fold higher in patients with severe symptoms and COVID-19 infection, although the difference was not significant (t-Test, p=0.06). Nevertheless, our results show a significant positive correlation when we took into account only infected patients (DET) and *TLR4* expression, which resulted to be significantly increased in patients with severe confirmed COVID-19 disease (t-Test, p<0.05). We also analyzed the gene expression according to sex and age of patients by a nested ANOVA. Only COVID-19 DET samples, by Game-Howell's test showed significant differences of *TLR7*, *TLR3* and *TLR4* expression when discriminated by sex. The chi-square test of independence showed that there was no significant association between sex and been infected or not, X^2 (2, n=302, p>0.5) as well as between severe symptoms and be female or male, X^2 (2, n=302, p>0.07). Briefly, the differences in expression were not significant associated with symptomatology or been in DET or ND groups.

When analyzing by age of patient, DET and ND displayed the same patient's age distribution. The *TLR3*, *TLR4* and *TLR7* gene expression displayed a non-significant and low negative correlation with age, as determined by a nested and correlation analysis. When analyzed separately by Spearman's correlation, *TLR4* expression level showed a significant negative correlation with age of ND or DET cohorts (r_s = -0.14216, *p* 2-tailed = 0.01372). In conclusion, of all genes analyzed, only *TLR4* was significantly correlated in its expression with age. On the other hand, when analyzed by the Spearman's correlation, severity and age resulted in a significant positive correlation (r(302)=0.2676, p<0.01). These results were confirmed by Point-Biserial Correlation taking into account ranks of ages (6) and scoring symptoms low/mild/moderate as 0 and severe as 1. A significant correlation was also found (r=-0.143, p=0.012), in both non-detected and detected COVID-19 patients, meaning that older patients of these cohorts would develop high severity of symptoms (of a respiratory disease) disregarding if infected or not with COVID-19 and their *TLRs* gene expression.

DISCUSSION

Our data partially contradict previous findings, showing here now in a larger cohort, a non-significant difference between *TLR4* and *TLR7* expression in patients with non-detected and detected COVID-19. We also show here that *TLR3* gene expression was increased in non-detected COVID-19 patients independently of the form of the clinical symptoms of the disease (low, mild and severe). Finally, when the severity of symptoms and gene expression were together taken into account,

TLR4 expression was positively correlated with severe symptomatology/deaths in COVID-19 positive patients.

A previous work of Menezes et al. (2021) was carried out in 79 patients with severe COVID-19 detected at admission, that according to the WHO classification, were divided into two groups: patients who needed mechanical ventilation and/or deceased (severe, n=50) and patients who used supplementary oxygen but not mechanical ventilation and survived (mild/moderate, n=29). A control group (control, n=17) of healthy people was enrolled too. Instead of nasopharyngeal tissue, they use the peripheral blood (which will account for systemic immunity, not mucosal) to characterize gene expression (mRNA) of Toll-like receptors (*TLRs*) 3, 4, 7, 8, and 9, as well as other immune response genes (*RIGI*, *NLRP3*, *IFN-α*, *IFN-β*, *IFN-γ*, *IFN-λ*, *pro-interleukin(IL)-1β (pro-IL-1β)*, and *IL-18*), all determination according to the authors were done on admission, between 5–9 days, and between 10–15 days (Menezes et al., 2021). Circulating cytokines in plasma were also measured. When they compared the COVID-19 mild/moderate group with the COVID-19 severe group, the former had lower expression of *TLR3* and overexpression of *TLR4*. In the present work, we informed the expression level of three *TLRs* at the moment of infection determination, between 5-9 days of symptoms, using nasopharyngeal swaps. Importantly, our samples not only corresponded to a larger cohort but also focused on mucosal immunity rather than systemic immunity. Furthermore, gene expression was analyzed over a shorter timeframe following symptom onset compared to the study by Menezes et al. (2021) and different severity levels, age and sex of subjects presenting some of the twelve WHO recognized COVID-19 symptoms and with SARS-CoV-2 viral genome determined were taken into account. In concordance with the findings of Menezes et al. (2021), we confirmed, in a much larger cohort of patients, that *TLR4* expression was increased in patients with severe symptomatology; on the contrary, in COVID-19 positive infected patients, we observed a non-significant tendency to higher expression.

In the case of *TLR3*, there were no difference in expression within the positive COVID-19 patients when contrasted according the severity of symptoms. *TLR3* was found to be increased in non-positive COVID-19 patients disregarding their disease severity. Since *TLR3* recognizes viral double stranded RNA (dsRNA), this increased expression is therefore likely to be related to other viral infections present in the ND cohort with low to severe symptomatology.

Other authors, such as Bagheri-Hosseinabadi et al. (2022) used the same type of biological samples though in a smaller set of individuals (90 COVID-19 patients and 50 controls) with symptomatology of COVID-19 infection. However, they were compared with controls with low or no symptoms, healthy and not infected or suffering from a respiratory illness. Samples were classified into those without symptoms, with symptoms but not hospitalized, and with symptoms and hospitalized. They showed that *TLR3*, *TLR7*, *TLR8*, and *TLR9* were over expressed in the COVID-19 patients with clinical symptoms and in need of hospitalization. By the contrary, we did not find differences between the expressions of *TLR3*, *TLR4* and *TLR7* in those hospitalized non COVID-19 (n=8) or hospitalized COVID-19 patients (n=5) (data not shown). Furthermore, we found no differences in the

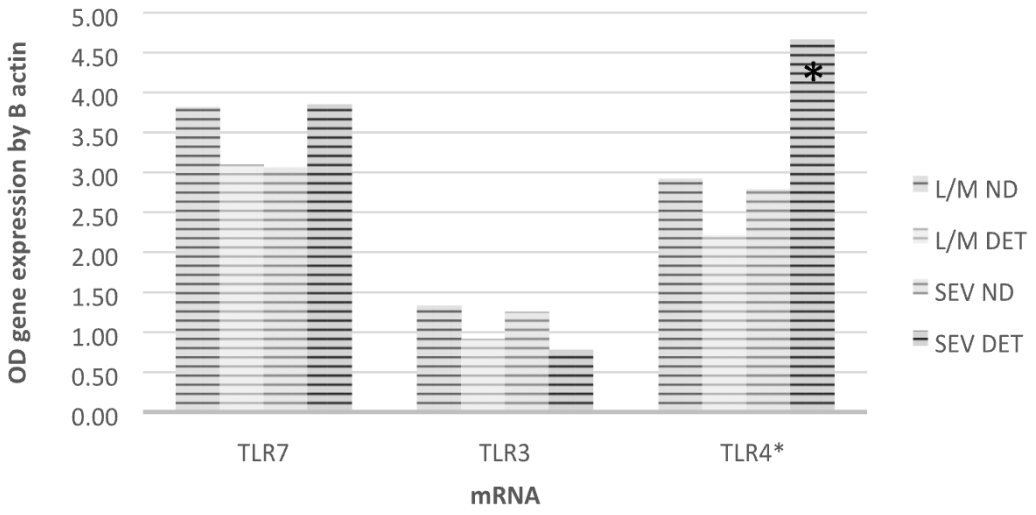

Figure 3. *TLRs* expression in the four studied cohorts: L/M ND: low or mild symptomatology, COVID-19 non-detected; SEV ND: severe symptomatology, COVID-19 non-detected; L/M DET: low or mild symptomatology, COVID-19 detected; SEV DET: severe symptomatology, COVID-19 detected. *Significant statistical differences ($p<0.05$).

Table 4. Semi-quantitative expression of *TLR3, TLR4* and *TLR7* normalized by housekeeping gene *B-actin* in four cohorts of patients: non-detected COVID-19 with low/mild symptoms, non-detected COVID-19 with severe symptoms, detected COVID-19 with low/mild and detected COVID-19 with severe symptoms. n= sample size; SE= standard error taking into account sample size. *Significant statistical differences ($p<0.05$).

	COVID-19 detected		**COVID-19 non-detected**	
	Low/Mild (n=131)	Severe (n=19)	Low/Mild (n=127)	Severe (n=25)
TLR7/B-actin	3.1	3.85	3.82	3.06
SE	0.2	0.53	0.24	0.37
TLR3/B-actin	0.91*	0.78*	1.33	1.25
SE	0.11	0.20	0.17	0.31
TLR4/B-actin	2.21	4.66*	2.92	2.79
SE	0.16	0.61	0.23	0.33

expression for *TLR3* and *TLR7* on COVID-19 patients vs. those without this infection (n=302). As expected, they concluded that the upregulation of *TLRs* was associated with clinical presentations of severity of symptoms on COVID-19 positive patients. They also concluded that the modulation of *TLR3*, *TLR7*, *TLR8* and *TLR9* in the epithelial cells of COVID-19 cases may estimate the disease severity and requirement for hospitalization. On the contrary, the analysis of our data demonstrated that *TLR3* is overexpressed in ND cohorts, while *TLR7* showed no significant differences between the samples. Our results showed that increased *TLR4* expression was correlated with the severity of infections.

When the severity is taken into account, the *TLR4* expression corroborates previous results of the higher level of expression in COVID-19 patients with high or severe symptomatology and hospitalization. Accordingly, *TLR4* expression is suggested as the only factor correlated to unfavorable outcome for COVID-19 infection (Bagheri-Hosseinabadi et al., 2022). On the other hand, the higher expression of *TLR3* correlates positively with severe symptoms in those patients without SARS-Cov-2 infection, a discrepancy with previous results, as those from Menezes et al. (2021). Noteworthy, the results are not comparable, since these authors set as controls patients without respiratory syndrome and analyzed a smaller sample size.

To the best of our knowledge, none of the previous studies (as well as this work) took into account the mRNA expressions in vaccinated patients. More likely because vaccinated patients will rarely develop severity symptoms like those common in non-vaccinated patients. As shown in mice (Do et al., 2023), the effects of co-administration of *TLR* agonists (*TLR* 3, 4, 9) as SARS-CoV vaccine adjuvants, fail to boost mucosal humoral immunity (i.e.: anti-spike neutralizing antibodies) underlining the effect of vaccination on *TLR*s expression and other immunity respondent. These effects are similarly reproduced in humans where the primary innate response to *TLR*s can be used to predict SARS-CoV vaccine protective humoral responses, as it was shown in a small clinical trial when using AstraZeneca ChAdOx1 nCoV-19(AZD1222) (Shen et al., 2022). Consequently, vaccination would introduce a plausible bias by modulating the immune response of patients towards more effective (less exaggerated) responses, in most cases, which can account for the differences observed here for *TLR3* and *TLR4*. Although we cannot completely discard vaccination effect for our study, we can assume the assayed population as naive for COVID-19 immunizations, since the first COVID-19 vaccination campaign in Argentina started only for vulnerable population (such as older adults, adults, younger adults) by late May 2021, meanwhile massive vaccination for all ages started at the end of 2021 in province of Entre Ríos. Another indication of the naive immune status of the analyzed cohorts in this study, is that we found a significant positive correlation between severity and age, in both non detected and detected COVID-19 patients, meaning that older patients of this cohort tend to develop severe symptoms (an almost opposite scenario than the attained upon immunization).

If we take into account the conclusion of a previous work carried out in our laboratory (Eberhardt et al., 2022) we can support that the vast majority of our samples are from non-vaccinated patients. In this sense, according with our results for *TLR3* and *TLR7* expressions, there were a non-significant decrease in expression by age, while, *TLR4* was found to be significantly increased in children, young adults and adults. By three-nested ANOVA analyses, we found that age and severity in DET cohorts only have a significant impact on *TLR4* expression. Repeating, *TLR4* is an innate immune response effector and it has the role of

pattern recognition receptors (PRRs) in the immune system. PRRs are crucial for detecting signs of infection or damage, helping to initiate appropriate immune responses. Specially it recognizes pathogen-associated molecular patterns (PAMPs) and damage-associated molecular patterns (DAMPs). *TRL4* intrinsic relationship with aging, could be part of its normal response in young and adult cohorts, overexpressing *TLR4* during infection in a significant correlation with severe disease (Wong et al., 2020; Khanmohammadi and Rezaei, 2021; Seery et al., 2021; Dai et al., 2022; Kim et al., 2023).

Overall, we found that these genes are expressed in a different manner between women and men. This can be supported specially from reports showing that *TLR7* expression depends on the methylation pattern of the X chromosome (Gomez-Carballa et al., 2022), however, in contrast to other reports there were no association between been infected or not by COVID-19, nor the symptomatology by sex, on the expression levels of *TLR7*. Importantly, regarding *TLR7*, it is known that its expression may be different in females and males due to the altered methylation patterns in the X-linked *TLR7* gene and consequential down-regulation of its expression in males (a difference that is not disease or infection related) (Gomez-Carballa et al., 2022).

Finally, the expression of *TLR4* in detected patients with severe symptoms was significantly higher than in detected patients with low or moderate symptoms, while the expression of *TLR3* in those patients without the infection was significantly higher than in detected patients. Therefore, we confirmed that the immune response of *TLR4* and *TLR3*, as measured by its gene expression, can be associated with harm (damage) rather than host protection, due to persistent inflammation and tissue destruction in the pathogenesis of COVID-19. This applies also to the other severe respiratory diseases. We suggest that research efforts should focus on therapies based on the regulation of *TLR4* and *TLR3* expression, to diminish inflammation in severe respiratory diseases, with or without COVID-19 infection. At the same time, the role of *TLR3* in different respiratory pathologies needs to be evaluated as it probably mitigates symptoms in other respiratory viral infections. Nonetheless, *TLR4* high expression could be monitored as a marker or cause of severe symptoms in COVID-19 infections.

In agreement with Onofrio et al. (2020), we suggest that *TLRs* are involved in the development of fatal clinical manifestations in severe COVID-19 patients, essentially in ARDS, and therefore related to the progression towards fatal respiratory failure. Our data also suggest that even though our innate immune system contributes to the elimination of viruses by *TLRs* sensing and signaling, the maintenance of inflammation and tissue destruction in the host could be linked with *TLR4* expression in COVID-19 infected patients. We confirmed that *TLR3*, *TLR4* and *TLR7* could be involved in the pathogenesis of other respiratory diseases and the severity in aged populations. We suggest that treatments focused on the expression of these genes during the progression of inflammatory respiratory diseases, could be an interesting starting point for the development of therapies that avoid unfavorable outcomes and severe symptoms in younger patients. We do not need to underline the importance of sex in the infection outcome, because both genders showed similar symptomatology (although with differences in gene expression).

We suggest to analyze the repositioning of TLR antagonist and agonist drugs, developed as consequence of anti-tumoral and cancer studies (Chahal et al., 2013; Urban-Wojciuk et al., 2019; Khalifa and Ghoneim, 2021; Dai et al., 2022), for its use in ARDS and severe COVID-19 infected patients.

We found a low positive correlation between severity and age. Taking into account these findings and those from previous studies showing cytokine storms and age-related disease development (Khanmohammadi and Rezaei, 2021; Dai et al., 2022), we can conclude that acute lung injury is a feature of severe respiratory disease in adults, disregarding the gene expression here studied.

As a consequence of the differences found in *TLR3* and *TLR4*, we suggest conducting a much larger study, with patients with symptoms of respiratory disease during the Delta and Omicron waves in Argentina (2020-2021), which would decrease biases by number and allow analysis from different detection systems, including the vaccination status. A thorough knowledge of the status of *TLR3*, *TLR4* and *TLR7* expression in infected patients with ARDS or SARS- Cov-2, can help in repositioning cancer drugs for ARDS treatment for those patients with increased *TLR4* expression.

Advances in the development of TLR agonists and antagonists as immunomodulators, including their roles as vaccine adjuvants and treatments for hyperinflammatory responses, hold promise for future interventions against SARS-CoV-2 or similar infections (Ghasemi et al., 2014; Eberhardt et al., 2022).

BIBLIOGRAPHY

Allhorn S., Carsten Böing A., Koch A., Kimmig R., Gashaw I. (2008) TLR3 and TLR4 expression in healthy and diseased human endometrium. Reprod. Biol. Endocrinol. 6: 1-11. doi: 10.1186/1477-7827-6-40

Bagheri-Hosseinabadi Z., Zarandi E.R., Mirabzadeh M., Amiri A., Abbasifard M. (2022) mRNA expression of toll-like receptors 3, 7, 8, and 9 in the nasopharyngeal epithelial cells of coronavirus disease 2019 patients. BMC Infect. Dis. 22 (1): 448. doi: 10.1186/s12879-022-07437-9

Bezemer G.F.G., Garssen J. (2021) TLR9 and COVID-19: a multidisciplinary theory of a multifaceted therapeutic target. Front. Pharmacol. 11: 601685. doi:10.3389/fphar.2020.601685

Bortolotti D., Gentili V., Rizzo S., Schiuma G., Beltrami S., Strazzabosco G., Fernandez M., Caccuri F., Caruso A., Rizzo R. (2021) TLR3 and TLR7 RNA sensor activation during SARS-CoV-2 infection. Microorganisms 9 (9): 1820. doi: 10.3390/microorganisms9091820

Chahal D.S., Raja K., Sivamani R., Isseroff R., Dasu M. (2012) Plant-based modulation of toll-like receptors: an emerging therapeutic model. Phytother. Res. 27 (10): 1423-1438. doi: 10.1002/ptr.4886

Dai J., Wang Y., Wang H., Gao Z., Wang Y., Fang M., Shi S., Zhang P., Wang H., Su Y., Yang M. (2022) Toll-like receptor signaling in severe acute respiratory syndrome coronavirus 2-induced innate immune responses and the potential application value of toll-like receptor immunomodulators in patients with coronavirus disease 2019. Front. Microbiol. 13: 948770. doi: 10.3389/fmicb.2022.948770

Do K.T.H., Willenzon S., Ristenpar J., Janssen A., Volz A., Sutter G., Förster R., Bošnjak B. (2023) The effect of Toll-like receptor agonists on the immunogenicity of MVA-SARS-2-S vaccine after intranasal administration in mice. Front. Cell. Infect. Microbiol. 13: 1259822. doi: 10.3389/fcimb.2023.1259822

Durán A., Álvarez-Mon M., Valero N. (2014) Papel de los receptores tipo toll (TLRs) y receptores para dominios de oligomerización para la unión a nucleótidos (NLRs) en las infecciones virales. Invest. Clin. 55 (1): 61–81.

Eberhardt A.T., Simoncini M., Piña C., Galoppo G., Parachú-Marco V., Racca A., Arce S., Viotto E., Facelli F., Valli F., Botto C., Scarpa L., Junges C., Palavecino C., Beccaria C., Sklar D., Mingo G., Genolet A., Muñoz de Toro M., Aimar H., Martinez Marignac V., Beldomenico P. M. (2022) Preceding anti-spike IgG levels predicted risk and severity of COVID-19 during the Omicron-dominant wave in Santa Fe city, Argentina. Epidemiol. Infect. 150: e187. doi: 10.1017/S0950268822001716

Gadanec L.K., McSweeney K.R., Qaradakhi T., Ali B., Zulli A., Apostolopoulos V. (2021) Can SARS-CoV-2 virus use multiple receptors to enter host cells? Int. J. Mol. Sci. 22 (3): 992. doi: 10.3390/ijms22030992

Ghasemi N., Amjadi F., Salehi E., Shakeri M., Aflatoonian A., Aflatoonian R. (2014) Expression of Toll-like receptors 7-10 in human fallopian tubes. Iran. J. Reprod. Med. 12(6): 389–394.

Gomez-Carballa A.G., Pardo-Seco J., Pischedda S., Rivero-Calle I., Butler-Laporte G., Richards J.B., Viz-Lasheras S., Martinón-Torres F., Salas A. (2022) Sex-biased expression of the *TLR7* gene in severe COVID-19 patients: Insights from transcriptomics and epigenomics. Environ. Res. 215: 114288. doi: 10.1016/j.envres.2022.114288

Imai Y., Kuba K., Neely G.G., Yaghubian-Malhami R., Perkmann T., van Loo G., Ermolaeva M., Veldhuizen R., Leung Y.H., Wang H., Liu H., Sun Y., Pasparakis M., Kopf M., Mech C., Bavari S., Peiris J.S., Slutsky A.S., Akira S., Hultqvist M., Penninger J.M. (2008) Identification of oxidative stress and Toll-like receptor 4 signaling as a key pathway of acute lung injury. Cell, 133(2): 235–249. doi: 10.1016/j.cell.2008.02.043

Jung H.E., Lee H.K. (2021) Current understanding of the innate control of toll-like receptors in response to SARS-CoV-2 infection. Viruses 13 (11): 2132. doi: 10.3390/v13112132

Khalifa A.E., Ghoneim A.I. (2021) Potential value of pharmacological agents acting on toll-like receptor (TLR) 7 and/or TLR8 in COVID-19. CRPHAR 2: 100068. doi: 10.1016/j.crphar.2021.100068

Khanmohammadi S., Rezaei N. (2021) Role of toll-like receptors in the pathogenesis of COVID-19. J. Med. Virol. 93(5): 2735–39. doi: 10.1002/jmv.26826

Kim H.J., Kim H., Lee J.H., Hwangbo C. (2023) Toll-like receptor 4 (TLR4): new insight immune and aging. I & A 20(1): 67. doi: 10.1186/s12979-023-00383-3

Manik M., Singh R.K. (2021) Role of toll-like receptors in modulation of cytokine storm signaling in SARS-CoV-2-induced COVID-19. J. Med. Virol. 94: 869-877. doi: org/10.1002/jmv.27405

Menezes M.C.S., Veiga A.D.M., Martins de Lima T., Kunimi Kubo Ariga S., Vieira Barbeiro H., de Lucena Moreira C., Souza H.P. (2021) Lower peripheral blood Toll-like receptor 3 expression is associated with an unfavorable outcome in severe COVID-19 patients. Sci. Rep. 11(1): 15223. doi: 10.1038/s41598-021-94624-4

Nishiya T., Kajita E., Miwa S., DeFranco A.L. (2005) TLR3 and TLR7 are targeted to the same intracellular compartments by distinct regulatory elements. JBC 280(44): 37107-37117. doi: 10.1074/jbc.M504951200

O'Neill L.A.J., Golenbock D.T., Bowie A.G. (2013) The history of Toll-like receptors — redefining innate immunity. Nat. Rev. Immunol. 13(6): 453-460. doi: 10.1038/nri3446

Onofrio L., Caraglia M., Facchini G., Vincenzo M., De Placido S., Buonerba C. (2020) Toll-like receptors and COVID-19: a two-faced story with an exciting ending. Future Science OA 6(8): FSO605. doi: 10.2144/fsoa-2020-0091

Poltorak A., He X., Smirnova I.I., Liu M., Van Huffel C., Du X., Birdwell D., Alejos E., Silva M., Galanos C., Freudenberg M.A., Ricciardi-Castagnoli P., Layton B., Beutler B. (1998) Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. Science 282 (5396): 2085-2088. doi: 10.1126/science.282.5396.2085

Seery V., Raiden S., Algieri S.C., Grisolía N.A., Filippo D., De Carli N., Di Lalla S., Cairoli H., Chiolo M.J., Meregalli C.N., Giménez L.N., Gregorio G., Sarli M., Alcalde A.L., Davenport C., Bruera M.J., Simaz N., Pérez M., Nivela V., Zaragoza Bayle V., Tuccillo P., Agosta M.T., Eisenhardt Perez H.E., Villa Nova S., Suárez P., Takata E.M., García M., Lattner J., Rolón M.J., Coll P., Sananez I., Holgado M.P., Ferrero F.C., Geffner J., Arruvito L. (2021) Blood neutrophils from children with COVID-19 exhibit both inflammatory and anti-inflammatory markers. EBioMedicine 67: 103357. doi: 10.1016/j.ebiom.2021.103357

Shen C.F., Yen C.L., Fu Y.C., Cheng C.M., Shen T.C., Chang P.D., Cheng K.H., Liu C.C., Chang Y.T., Chen P.L., Ko W.C., Shieh C.C. (2022) Innate immune responses of vaccinees determine early neutralizing antibody production after ChAdOx1nCoV-19 vaccination. Front. Immunol. 13: 807454. doi: 10.3389/fimmu.2022.807454

Urban-Wojciuk Z., Khan M.M., Oyler B.L., Fåhraeus R., Marek-Trzonkowska N., Nita-Lazar A., Hupp T.R., Goodlett D.R. (2019) The Role of TLRs in anti-cancer immunity and tumor rejection. Front. Immunol. 10: 2388. doi: 10.3389/fimmu.2019.02388

van der Made C.I., Simons A., Schuurs-Hoeijmakers J., van den Heuvel G., Mantere T., Kersten S., van Deuren R.C., Steehouwer M., van Reijmersdal S.V., Jaeger M., Hofste T., Astuti G., Corominas Galbany J., van der Schoot V., van der Hoeven H., Hagmolen Of Ten Have W., Klijn E., van den Meer C., Fiddelaers J., de Mast Q., Hoischen A. (2020) Presence of genetic variants among young men with severe COVID-19. JAMA 324 (7): 663. doi: org/10.1001/jama.2020.13719

Wong L.S.Y., Loo E.X.L., Kang A.Y.H., Lau H.X., Tambyah P.A., Tham E.H. (2020) Age-related differences in immunological responses to SARS-CoV-2. JACI: In practice 8(10): 3251-3258. doi: 10.1016/j.jaip.2020.08.026

ACKNOWLEDGEMENTS

The authors want to thank the NEOKIT Cassara team for technical access and guidance, Tec. Fiorella De La Lama for technical support, and to all the employees of Hospital San Jose who served during the COVID-19

pandemic health emergency. We acknowledge VLMM for the financing support through ST2719 CICYTTP-CONICET funds, and to the Ministry of Health of the Province of Entre Ríos, Hospital San Jose and CONICET for financial and salary support.

AUTHOR CONTRIBUTION STATEMENT

Conceptualization: Martinez Marignac V.L. and Lirussi D. Data collection and curation: Martinez Marignac V.L., Gazali Z., Oertlin G., Salinas M. and Fleischmann E. Formal analysis: Martinez Marignac V.L. and Lirussi D. Funding acquisition and Resources: Marchetti G. and Martinez Marignac V.L. Methodology: Richard S., Martinez Marignac V.L., Oertlin G., Fleishmann E. and Lirussi D. Writing and revision: Richard S., Martinez Marignac V.L. and Lirussi D. Final editing: Martinez Marignac V.L. and Lirussi D.

CONFLICTS OF INTEREST

The authors inform that they are all public employees belonging to CONICET or to the Ministry of Health of the Province of Entre Ríos. The authors declare that they have no conflict of interest.

BAG | Journal of Basic & Applied Genetics
Vol XXXV (2); 47-48; July-December 2024
Journal of the Argentine Society of Genetics

LETTER TO JBAG

Recently, Dr. Frank W. Nicholas, from the School of Veterinary Science at the University of Sydney (Sydney, New South Wales, Australia), published an article titled "Animal genetics 100 years ago" (DOI: 10.1111/age.70017) in the journal Animal Genetics. This publication commemorates the centennial anniversary of the first book to include the term Animal Genetics in its title. This text, authored by Francis A. E. Crew, Professor of Genetics and Director of the Department of Animal Breeding at the University of Edinburgh, presents a historical perspective on the state of animal genetics a century ago. Dr. Nicholas highlights the substantial advances in understanding inheritance mechanisms during the two decades following the rediscovery of Mendel's laws. In this context, Crew's book, "Animal Genetics: An Introduction to the Science of Animal Breeding", encompasses twelve chapters covering foundational concepts, including: the recognition that genes are located on chromosomes (Chapter IV), evidence supporting chromosomal sex determination (Chapters VI–VIII), the understanding that quantitative trait variation is determined by multiple genes and non-genetic factors (Chapter II), the detrimental effects of inbreeding on fertility, attributed to increased homozygosity of deleterious alleles (Chapter IX), the observation that crossbreeding enhances hybrid vigor (Chapter IX), and the identification of Mendelian inheritance patterns in various diseases, enabling genetic disease control through pedigree recording (Chapter XI). Additionally, the book discusses key genetic concepts such as gene, chromosome, locus, linkage, recombination, linkage map, pleiotropy, and multiple alleles. Despite the exponential growth of animal genetics over the past century, many principles and examples outlined in Crew's work remain fundamental and are frequently cited in contemporary undergraduate genetics textbooks.

Dr. Guillermo Giovambattista

CARTA A JBAG

Recientemente, el Dr. Frank W. Nicholas de la Escuela de Ciencias Veterinarias de la Universidad de Sydney (Sydney, New South Wales, Australia) publicó en la revista Animal Genetics un artículo denominado "Animal genetics 100 years ago" (https://doi.org/10.1111/age.70017), en conmemoración a los cien años de la edición del primer libro que incluyó un su título las palabras "Animal Genetics". Este texto escrito por Francis A. E. Crew, profesor de Genética y director del Departamento de Animal Breeding en la Universidad de Edimburgo, refleja un interesante visón del estado del arte sobre genética animal hace cien años. El artículo del Dr. Frank W. Nicholas resalta cómo durante las dos décadas posteriores al redescubrimiento de las leyes de Mendel se avanzó significativamente en la comprensión de los mecanismos de la herencia. En este sentido, el libro de Francis A. E. Crew denominado "Animal Genetics: an Introduction to the Science of Animal Breeding" incluye en sus doce capítulos temáticas y conceptos tales como, el reconocimiento que los genes se localizan en los cromosomas (capítulo IV), evidencias de la determinación cromosómica del sexo (capítulos VI – VIII), que la variación de los caracteres cuantitativos está determinada por múltiples genes y factores no genéticos (capítulo II), que la consanguinidad resulta en la disminución de la fertilidad debido al aumento de la homocigosidad de alelos indeseables (capítulo IX), que el cruzamiento entre líneas o razas produce vigor híbrido (capítulo IX), y que muchas enfermedades tienen una herencia Mendeliana y por lo tanto pueden ser controladas registrando el pedigrí (capítulo XI). Además, se discuten conceptos como gen, cromosoma, locus, ligamiento, recombinación, mapa de ligamiento, pleiotropía y alelos múltiples. En los cien años transcurridos desde la publicación del libro de Francis A. E. Crew, la genética animal ha experimentado un crecimiento exponencial. Sin embargo, muchos de los temas de los temas y ejemplos incluidos en este texto aún los podemos encontrar como ejemplos clásicos en los libros de genética general utilizados en los cursos de grado universitarios.

Dr. Guillermo Giovambattista

BAG

Journal of Basic & Applied Genetics