



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Utilizando herramientas moleculares para aumentar el conocimiento de quelonios continentales endémicos y amenazados del Chocó Biogeográfico

Sebastián Adolfo Cuadrado Ríos

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Biología
Bogotá DC, Colombia
2025

Utilizando herramientas moleculares para aumentar el conocimiento de quelonios continentales endémicos y amenazados del Chocó Biogeográfico

Sebastián Adolfo Cuadrado Ríos

Tesis o trabajo de investigación presentada(o) como requisito parcial para optar al título
de:

Doctor en Ciencias - Biología

Director (a):

Mario Vargas-Ramírez Biol., M. Sc., Dr. Rer. Nat.

Línea de Investigación:

Sistemática y Evolución

Grupo de Investigación:

Grupo de Biodiversidad y Conservación Genética, Instituto de Genética

Universidad Nacional de Colombia

Facultad de Ciencias, Departamento de Biología

Bogotá DC, Colombia

2025

A mis padres, a ellos dedico toda mi obra

*Al Chocó y su biodiversidad, la más
maravillosa de todas*

Y a su gente, que me acogieron como uno más

Agradecimientos

A mis padres y mi hermana, cuyo apoyo incondicional y amor constante me dieron la fuerza para culminar este proceso.

A mi director, Prof. Dr. Mario Vargas Ramírez, por la confianza que depositó en mí y su constante apoyo académico y personal.

Al Prof. Dr. Uwe Fritz, por su apoyo incondicional en la fase de laboratorio del proyecto, además de las fructíferas charlas.

A Anke Müller, Dr. Christian Kehlmaier y Andrea Criado-Flórez, por su apoyo durante la fase de laboratorio del proyecto.

A mis compañeros del Grupo de Biodiversidad y Conservación Genética, por las fructíferas discusiones. En especial agradezco el apoyo académico y personal de Felipe Hernández y Jessica Rodríguez.

A Juliana Gaviria y Never Cavadia, compañeros de numerosas salidas de campo al Chocó, por su apoyo en los momentos clave de este proyecto de tesis.

A las instituciones Turtle Conservation Fund (TCF) y Turtle Survival Alliance (TSA), por su apoyo financiero para el desarrollo de este proyecto.

Al Ministerio de Ciencia, Tecnología e Innovación y Gobernación del Chocó, que me apoyaron financieramente mediante la “Convocatoria del Fondo de Ciencia, Tecnología e Innovación del Sistema General de Regalías, en el marco del Programa de Becas de Excelencia Doctoral del Bicentenario – Corte II”.

Al Servicio Alemán de Intercambio Académico (Deutscher Akademischer Austauschdienst, DAAD), por apoyarme mediante la “Research Grant (bi-nationally supervised Doctoral Degree/Cotutelle, 2023/24)”.

Finalmente, quiero agradecer a la larguísima lista de personas que me ayudaron durante mi trabajo de campo en el Chocó. El resultado de esta tesis nace de un esfuerzo colectivo que incluye las comunidades, que me apoyaron en todo momento y me hicieron sentir el calor de su amabilidad y gentileza.

Resumen

Utilizando herramientas moleculares para aumentar el conocimiento de quelonios continentales endémicos y amenazados del Chocó Biogeográfico

El Chocó Biogeográfico, en el noroeste de Suramérica, constituye uno de los principales focos de biodiversidad del planeta, pero sus tortugas de agua dulce siguen siendo poco estudiadas. Esta tesis tuvo como objetivo generar recursos genómicos y evaluar la diversidad genética, la conectividad poblacional y las relaciones evolutivas de especies clave de la región, con énfasis en *Trachemys medemi*. En el primer capítulo se secuenciaron y anotaron los tres primeros genomas mitocondriales completos de *T. medemi*, los cuales confirmaron su pertenencia al clado suramericano de *Trachemys* y proporcionaron referencias útiles para estandarizar marcadores de bajo costo como el D-loop. En el segundo capítulo se analizó la filogeografía de *Kinosternon leucostomum* mediante ADN mitocondrial y nuclear, revelando una estructura genética poco profunda dominada por aislamiento por distancia y con discontinuidades genéticas que ocurren por el Escarpe de Hess y el río Prado. En el tercer capítulo se evaluó la genética poblacional de *T. medemi* con secuencias del D-loop y 11 microsatélites en 96 individuos, encontrando poblaciones panmícticas con baja diversidad nuclear ($H_e = 0,479$) y un tamaño efectivo poblacional estimado en 246 (IC95%: 28–1305). Finalmente, el cuarto capítulo abordó la filogenómica del género *Rhinoclemmys*, soportando un escenario de múltiples invasiones durante el Mioceno hacia Sur América. En conjunto, los resultados evidencian tanto la importancia evolutiva del Chocó como la vulnerabilidad de sus tortugas frente a presiones antrópicas, resaltando la necesidad de estrategias de conservación que integren genética, ecología y manejo de hábitats.

Palabras clave: Chocó Biogeográfico, genética poblacional, filogeografía, conservación, tortugas dulceacuícolas.

Abstract

Using molecular tools to increase knowledge of continental endemic and threatened chelonians from the Biogeographic Chocó

The Biogeographic Chocó, located in northwestern South America, is one of the planet's major biodiversity hotspots, yet its freshwater turtles remain poorly studied. This dissertation aimed to generate genomic resources and to evaluate genetic diversity, population connectivity, and evolutionary relationships of key species in the region, with emphasis on *Trachemys medemi*. In the first chapter, the three first complete mitochondrial genomes of *T. medemi* were sequenced and annotated, confirming its placement within the South American *Trachemys* clade and providing useful references to standardize cost-effective markers such as the D-loop. The second chapter analyzed the phylogeography of *Kinosternon leucostomum* using mitochondrial and nuclear DNA, revealing shallow genetic structure dominated by isolation by distance, with genetic breaks at the Hess Escarpment and the Río Prado. The third chapter assessed the population genetics of *T. medemi* using D-loop sequences and 11 microsatellites from 96 individuals, finding panmictic populations with low nuclear diversity ($H_e = 0.479$) and an estimated effective population size of 246 (95% CI: 28–1305). Finally, the fourth chapter addressed the phylogenomics of the genus *Rhinoclemmys*, supporting a scenario of multiple Miocene invasions into South America. Taken together, these results highlight both the evolutionary importance of the Chocó and the vulnerability of its turtles to anthropogenic pressures, emphasizing the need for conservation strategies that integrate genetics, ecology, and habitat management.

Keywords: Biogeographic Chocó, population genetics, phylogeography, conservation, freshwater turtles.

Contenido

	Pág.
Resumen	V
Introduction	1
1. Capítulo 1	9
1.1 Introduction.....	9
1.2 Materials and Methods.....	10
1.3 Results.....	12
1.4 Discussion	16
1.5 References	17
2. Capítulo 2	19
2.1 Introduction.....	19
2.2 Materials and Methods.....	21
2.3 Results.....	24
2.4 Discussion	26
2.5 References	30
3. Capítulo 3	37
3.1 Introduction.....	38
3.2 Materials and Methods.....	39
3.3 Results.....	46
3.4 Discussion	53
3.5 References	58
4. Capítulo.....	69
4.1 Introduction.....	69
4.2 Materials and Methods.....	70
4.3 Results.....	73
4.4 Discussion	82
4.5 References	87
5. Conclusions	99
5.1 Conclusions.....	99
5.2 Recommendations	101
5.3 References	102

x

Introduction

The Biogeographic Chocó (also referred to as the Tumbes–Chocó–Magdalena region; Forero-Medina and Joppa 2010) spans the Pacific lowlands of Ecuador, Colombia, and Panama, and represents one of the planet’s most biologically rich and unique areas. As one of the wettest regions on Earth, it experiences record-breaking rainfall levels (Poveda et al. 2004, Yepes et al. 2019). This combination of climatic stability and complex topography has fostered extraordinary biodiversity and high endemism (Aguirre-C and Rangel-Ch 2008, Mittermeier et al. 2011, Christenhusz et al. 2017). Like the Atlantic Forest in Brazil, the Chocó has functioned as both a Pleistocene refuge and a modern cradle of speciation (Jaramillo-Vivanco et al. 2010, Arteaga et al. 2016, Pérez-Escobar et al. 2019). Molecular studies have revealed intricate diversification patterns across elevational gradients (Arteaga et al. 2016), heterogeneous landscapes (Cucalón et al. 2022), and historical connections with adjacent biomes (Elías et al. 2023, Sedano-Cruz et al. 2024, van der Kamp and Ortega-Andrade 2024). This dual role as a refuge and an engine of diversification underscores the region’s global evolutionary importance, while escalating threats from habitat conversion and climate change highlight its urgent conservation status (Myers et al. 2000, FEOW 2021).

The Atrato River basin in northwestern Colombia exemplifies the ecological uniqueness of the Chocó. It comprises a vast, intricate hydrological network of rivers, streams, and floodplain wetlands embedded within tropical rainforests (Anaya-Acevedo et al. 2017, Palomino-Ángel et al. 2019). This system provides critical habitat for aquatic fauna but has been severely impacted by gold mining, deforestation, agricultural expansion, and contamination (Hurtado-Gómez et al. 2015, Calzadilla 2019, Caicedo-Rivas et al. 2022, Bernal-Alviz et al. 2025). These anthropogenic pressures have degraded ecosystems and threaten species with narrow distributional ranges. Despite its importance, the fauna of the Biogeographic Chocó remains poorly studied. Logistical challenges—including difficult access, transportation barriers, and complex sociopolitical conditions—continue to hinder research efforts (Ortiz Lancheros 2022, Otálvaro-Marín et al. 2023). Consequently, critical

knowledge gaps persist regarding the population status, connectivity, and conservation needs of many species, undermining the development of effective management strategies.

The Biogeographic Chocó harbors one of the richest reptile assemblages in Colombia, with more than 180 species reported across its ecosystems, including a remarkable proportion of endemics (Tailor-Rengifo and Rentería-Moreno 2011). This exceptional diversity reflects both the region's role as a transition zone between Central and South America and as a refuge fostering unique evolutionary trajectories. The Chocó department also stands out as one of Colombia's turtle-richest regions, harboring 12 species, a number only surpassed by Amazonas department with 13 (Turtle Taxonomy Working Group 2017). The regional turtle fauna includes both widespread and localized taxa. Its greatest significance lies in the presence of narrow endemics such as the Atrato slider (*Trachemys medemi*), restricted to the Atrato River basin, and Dunn's mud turtle (*Kinosternon dunni*), confined to the San Juan, Baudó, and Atrato drainages. Other species in the region include the brown wood turtle (*Rhinoclemmys nasuta*), the Colombian wood turtle (*Rhinoclemmys melanosterna*), the white-lipped mud turtle (*Kinosternon leucostomum*), and the red-footed tortoise (*Chelonoidis carbonarius*). Freshwater turtles in the Chocó are highly vulnerable to hunting, egg collection, and illegal trade (Epperson and Heise 2003, Moll and Moll 2004, Rhodin et al. 2011, TCC 2011, Morales-Betancourt et al. 2015, Páez et al. 2012, Lovich et al. 2018, Stanford et al. 2020). *Trachemys medemi*, in particular, has been identified as one of the most trafficked turtles in the region (Asprilla-Perea et al. 2013, Arroyave Bermúdez et al. 2014), exacerbating pressure to its already limited populations (Páez et al. 2022). Although some species are relatively widespread, the endemic turtles of the Biogeographic Chocó are especially imperiled due to their restricted ranges, small effective population sizes, and ongoing habitat degradation. The coexistence of widespread and narrowly distributed turtles in this hotspot illustrates the extraordinary ecological niche space available in the Biogeographic Chocó, which facilitates the establishment of multiple species in sympatry while simultaneously exposing them to common conservation threats.

Conservation biology increasingly recognizes that effective strategies must integrate molecular evidence, as ecological data alone rarely captures the evolutionary and demographic processes underlying species persistence. Molecular approaches including phylogeography, population genetics, and phylogenetics, provide unique insights into

historical connectivity, lineage diversification, and levels of genetic variation within and among populations (Moritz 1994, Avise 2000, Funk et al. 2012). Understanding the evolutionary history of organisms is particularly important in regions like the Biogeographic Chocó, where complex geological and climatic dynamics have shaped both ancient radiations and recent diversification. In turtles, molecular tools are especially relevant given their unusual life-history traits—extreme longevity, delayed maturity, low fecundity, temperature-dependent sex determination, and high juvenile mortality (Gibbons 1987)—which complicate demographic inference. By integrating molecular and ecological data, conservation practitioners can identify Evolutionarily Significant Units (ESUs) and Management Units (MUs), estimate effective population sizes, infer demographic history, and detect adaptive variation critical for resilience under environmental change (Fraser and Bernatchez 2001, Ouborg et al. 2010, Hohenlohe et al. 2021). Moreover, genetic studies elucidate how processes such as drift, inbreeding, and the accumulation of deleterious mutations can interact in a self-reinforcing extinction vortex (Amos and Balmford 2001), underscoring the urgency of monitoring genetic diversity as a cornerstone of long-term conservation planning.

The advent of next-generation sequencing (NGS) has driven a transition to conservation genomics. Approaches like reduced-representation sequencing (e.g., RAD-seq) and whole-genome sequencing enable the generation of thousands of single nucleotide polymorphism (SNPs), providing high resolution for disentangling processes like migration, genetic drift, adaptation, refining estimates of genetic diversity, effective population sizes, and adaptive differentiation (Kohn et al. 2006, Hohenlohe et al. 2021, Chattopadhyay et al. 2019, Gallego-García et al. 2019). In parallel, the use of complete mitochondrial genomes has become widespread, offering more robust phylogeographic and phylogenetic frameworks and higher-resolution insights into female population history than single-locus approaches. More recently, whole-genome sequencing has also opened unprecedented opportunities to investigate adaptive variation, demographic history, and genome-wide patterns of connectivity, marking a transformative advancement for conservation genetics and evolutionary biology.

This dissertation employs molecular tools to address critical knowledge gaps for endemic and threatened chelonians of the Biogeographic Chocó, integrating both classical and next-generation approaches. **Chapter 1** presents the first complete mitochondrial genomes of *Trachemys medemi*, describing their genomic organization, base

composition, and skew patterns, and evaluating their phylogenetic placement among other Neotropical *Trachemys*. These results provide an essential reference for future evolutionary and conservation studies of the species and genus. **Chapter 2** examines the phylogeography of *Kinosternon leucostomum*, integrating new mitochondrial and nuclear data with published datasets to test whether Chocoan populations represent distinct genetic lineages. Analyses reveal that isolation by distance predominates, with significant genetic breaks at the Hess Escarpment and Río Prado, Tolima (upper Magdalena valley); Chocoan populations belong to a widely distributed clade spanning the Magdalena Valley and Central America. These findings underscore the importance of the Chocó as a transition zone between Central and South America and highlight the need for targeted sampling in this region. **Chapter 3** provides the first population genetic assessment of *T. medemi* using mitochondrial control region sequences and 11 microsatellites from 96 individuals across seven localities. Results indicate shallow genetic structuring, panmixia and low nuclear diversity. Effective population sizes, while above short-term thresholds to avoid inbreeding, fall below benchmarks for maintaining long-term adaptive potential. These findings emphasize the vulnerability of *T. medemi* to demographic stochasticity and genetic drift, underscoring the urgent need for conservation strategies that ensure habitat quality, protect aquatic corridors, and implement genetic monitoring. **Chapter 4** evaluates the phylogenomic relationships within the genus *Rhinoclemmys*, with a particular focus on species distributed in the Biogeographic Chocó. This is of special relevance because the region harbors one of the highest concentrations of *Rhinoclemmys* species on the continent, including *R. melanosterna*, *R. annulata*, and *R. nasuta*. By clarifying their evolutionary relationships, this chapter contributes to understanding the macroevolutionary drivers of diversity, providing a broader framework for interpreting both regional and lineage-specific diversification.

Collectively, this dissertation provides essential genomic resources and address major data deficiencies for turtles of the Chocó. By filling critical knowledge gaps on the conservation genetics of endemic and threatened chelonians, this work supports the re-evaluation of threat categories, the delineation of conservation units, and the design of future programs for genetic rescue and demographic monitoring. Ultimately, it aims to strengthen the scientific foundation for the persistence of these unique lineages in a region where evolutionary uniqueness and anthropogenic pressures are profoundly intertwined.

References

- Aguirre-C, J., and Rangel-Ch, O. (2008). *El Chocó biogeográfico*. In: *Colombia diversidad biótica VI: Riqueza y diversidad de los musgos y líquenes en Colombia*, 77–84.
- Amos, W., and Balmford, A. (2001). When does conservation genetics matter?. *Heredity*, 87(3), 257-265.
- Anaya-Acevedo, J. A., Escobar-Martínez, J. F., Massone, H., Booman, G., Quiroz-Londoño, O. M., Cañón-Barriga, C. C., ... and Palomino-Ángel, S. (2017). Identification of wetland areas in the context of agricultural development using Remote Sensing and GIS. *Dyna*, 84(201), 186–194.
- Antonelli, A., and Sanmartín, I. (2011). Why are there so many plant species in the Neotropics? *Taxon*, 60(2), 403–414.
- Arteaga, A., Pyron, R. A., Peñafiel, N., Romero-Barreto, P., Culebras, J., Bustamante, L., Yáñez-Muñoz, M. H., and Guayasamin, J. M. (2016). Comparative phylogeography reveals cryptic diversity and repeated patterns of cladogenesis for amphibians and reptiles in northwestern Ecuador. *PLoS ONE*, 11(4), e0151746.
- Asprilla-Perea, J., Serna-Agudelo, J. E., and Palacios-Asprilla, Y. (2013). Diagnóstico sobre el decomiso de fauna silvestre en el departamento del Chocó (Pacífico Norte colombiano). *Revista UDCA Actualidad and Divulgación Científica*, 16(1), 175–184.
- Avice, J. C. (2000). *Phylogeography: the history and formation of species*. Harvard University Press.
- Caicedo-Rivas, G., Salas-Moreno, M., and Marrugo-Negrete, J. (2022). Health risk assessment for human exposure to heavy metals via food consumption in inhabitants of middle basin of the Atrato River in the Colombian Pacific. *International Journal of Environmental Research and Public Health*, 20(1), 435.
- Calzadilla, P. V. (2019). A Paradigm Shift in Courts' View on Nature: The Atrato River and Amazon Basin Cases in Colombia. *Law, Environment and Development Journal*, 15, 12.
- Cano, A., Manrique, H. F., Hoyos-Gómez, S. E., Echavarría, N., Upedui, A., Gonzalez, M. F., et al. (2017). Palms of the Darién Gap (Colombia-Panama). *Palms*, 61, 5–20.

- Christenhusz, M. J., Fay, M. F., and Chase, M. W. (2017). *Plants of the world: an illustrated encyclopedia of vascular plants*. London: Kew Publishing.
- Epperson, D. M., and Heise, C. D. (2003). Nesting and hatchling ecology of gopher tortoises (*Gopherus polyphemus*) in southern Mississippi. *Journal of Herpetology*, 37(2), 315–324.
- Fraser, D. J., and Bernatchez, L. (2001). Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular Ecology*, 10(12), 2741-2752.
- Frankham, R., Bradshaw, C. J., and Brook, B. W. (2014). Genetics in conservation management: revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation*, 170, 56–63.
- Forero-Medina, G., and Joppa, L. (2010). Representation of global and national conservation priorities by Colombia's protected area network. *PLoS One*, 5(10), e13210.
- Funk, W. C., McKay, J. K., Hohenlohe, P. A., and Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in ecology and evolution*, 27(9), 489-496.
- Gallego-García, N., Vargas-Ramírez, M., Forero-Medina, G., and Caballero, S. (2018). Genetic evidence of fragmented populations and inbreeding in the Colombian endemic Dahl's toad-headed turtle (*Mesoclemmys dahli*). *Conservation Genetics*, 19(1), 221–233.
- Gibbons, J. W. (1987). Why do turtles live so long?. *BioScience*, 37(4), 262-269.
- Hohenlohe, P. A., Funk, W. C., and Rajora, O. P. (2021). Population genomics for wildlife conservation and management. *Molecular Ecology*, 30(1), 62-82.
- Hurtado-Gómez, J. P., Vargas-Ramírez, M., Iverson, J. B., Joyce, W. G., McCranie, J. R., Paetzold, C., and Fritz, U. (2024). Diversity and biogeography of South American mud turtles elucidated by multilocus DNA sequencing (Testudines: Kinosternidae). *Molecular Phylogenetics and Evolution*, 197, 108083.
- Lovich, J. E., Ennen, J. R., Agha, M., and Gibbons, J. W. (2018). Where have all the turtles gone, and why does it matter? *BioScience*, 68(10), 771–781.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., and Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403(6772), 853-858.

Mittermeier, R. A., Turner, W. R., Larsen, F. W., Brooks, T. M., and Gascon, C. (2011). Biodiversity hotspots. In: Zachos, F. E., and Habel, J. C. (eds) *Global biodiversity conservation: the critical role of hotspots*. Springer, London, pp. 3–22.

Morales-Betancourt, M. A., Lasso, C. A., Páez, V. P., and Bock, B. C. (2015). *Libro rojo de los reptiles de Colombia*. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Universidad de Antioquia. Bogotá.

Ortiz Lancheros, C. A. (2022). Entre la fragilidad de la paz y la persistencia de la guerra: El caso de la subregión del Bajo Atrato, Chocó, Colombia. *Revista Ratio Juris*, 17(34), 319–342.

Otálvaro-Marín, B., Parra-López, M. Y., and Klinger-Cundumí, E. (2023). Análisis de las injusticias sociales, ambientales y territoriales del departamento del Chocó, Colombia. *Prospectiva*, 36, e20212476.

Páez, V. P., Bock, B. C., Alzate-Estrada, D. A., Barrientos-Muñoz, K. G., Cartagena-Otalvaro, V. M., Echeverry-Alcendra, A., ... and Vallejo-Betancur, M. M. (2022). Turtles of Colombia: an annotated analysis of their diversity, distribution, and conservation status. *Amphibian and Reptile Conservation*, 16(1), 106–135.

Palomino-Ángel, S., Anaya-Acevedo, J. A., Simard, M., Liao, T. H., and Jaramillo, F. (2019). Analysis of floodplain dynamics in the Atrato River Colombia using SAR interferometry. *Water*, 11(5), 875.

Poveda, G., Rojas, C. A., Rudas, A., and Rangel, O. (2004). El Chocó biogeográfico: ambiente físico. In: *Colombia diversidad biótica IV, El Chocó biogeográfico*/Costa Pacífica, 1–22.

Rhodin, A. G. J., Stanford, C. B., van Dijk, P. P., Eisemberg, C., Luiselli, L., Mittermeier, R. A., ... and Vogt, R. C. (2018). Global conservation status of turtles and tortoises (order Testudines). *Chelonian Conservation and Biology*, 17(2), 135–161.

- Stanford, C. B., Iverson, J. B., Rhodin, A. G. J., van Dijk, P. P., Mittermeier, R. A., Kuchling, G., ... and Walde, A. D. (2020). Turtles and tortoises are in trouble. *Current Biology*, 30(12), R721–R735.
- Tailor-Rengifo, M., and Rentería-Moreno, A. (2011). Diversidad de reptiles en el Chocó. In: *Colombia diversidad biótica XI*. Instituto de Ciencias Naturales, Bogotá.
- TCC [Turtle Conservation Coalition], Rhodin, A. G. J., Walde, A. D., Horne, B. D., van Dijk, P. P., Blanck, T., and Hudson, R. (2011). *Turtles in Trouble: The World's 25+ Most Endangered Tortoises and Freshwater Turtles—2011*. IUCN/SSC Tortoise and Freshwater Turtle Specialist Group, Turtle Conservation Fund, Turtle Survival Alliance, Turtle Conservancy, Chelonian Research Foundation, Conservation International, Wildlife Conservation Society, and San Diego Zoo Global.
- TTWG [Turtle Taxonomy Working Group: Rhodin, A. G. J., Iverson, J. B., Fritz, U., Gallego-García, N., Georges, A., Shaffer, H. B., and van Dijk, P. P.] (2025). *Turtles of the World: Annotated Checklist and Atlas of Taxonomy, Synonymy, Distribution, and Conservation Status (10th Ed.)*. Chelonian Research Monographs, 10, 1–575.
- Vargas-Ramírez, M., del Valle, C., Ceballos, C. P., and Fritz, U. (2017). *Trachemys medemi* n. sp. from northwestern Colombia turns the biogeography of South American slider turtles upside down. *Journal of Zoological Systematics and Evolutionary Research*, 55(4), 326–339.
- Yepes, J., Poveda, G., Mejía, J. F., Moreno, L., and Rueda, C. (2019). Choco-JEX: A research experiment focused on the Chocó low-level jet over the far eastern Pacific and western Colombia. *Bulletin of the American Meteorological Society*, 100(5), 779–796.

1. Capítulo 1

Complete mitochondrial genome and phylogenetic analyses of the Atrato Slider, *Trachemys medemi* (Testudines: Emydidae)

Capítulo publicado en “Cuadrado-Ríos, S., Vargas-Ramírez, M., Kehlmaier, C., and Fritz, U. (2025). Complete mitochondrial genome and phylogenetic analysis of the Atrato slider, *Trachemys medemi* (Testudines, Emydidae). *ZooKeys*, 1224, 253”.

Abstract

The complete mitochondrial genome of three *Trachemys medemi* specimens were sequenced and annotated. These represent the first complete mitogenomes reported for the species. The mitochondrial genome is a circular DNA molecule of 16,711-16,810 bp in size, with 61.0% AT content. It includes 13 protein-coding genes, 22 tRNAs, two ribosomal RNA genes, and a non-coding control region. The genome composition presents a slight positive AT skew (0.12274) and a negative GC skew (-0.34190). Phylogenetic analyses based on complete mitogenomes, which lack some *Trachemys* species, placed *T. medemi* as sister to *T. venusta*. Phylogenies from the same dataset but including available mtDNA information of most of the Neotropical *Trachemys* species, recovered *T. medemi* as sister to *T. dorbigni*, and this clade was sister to *T. venusta*, *T. yaquia*, and *T. ornata*. The newly obtained mitogenomes provide valuable data for further studies on population genetics, functional genomics, and evolutionary genomics of *Trachemys*.

Keywords: Chocó; mitogenome; primer walking; turtle; phylogeny.

1.1 Introduction

The Atrato Slider (*Trachemys medemi*) is a freshwater turtle species with a narrow distribution. It is restricted to the Atrato Basin of northwestern Colombia (Vargas-Ramírez et al. 2017; TTWG 2021). This makes it one of the endemic turtle species in the country

(Páez et al. 2022). Described in 2017 (Vargas-Ramírez et al. 2017), *Trachemys* individuals from the Atrato Basin were previously identified, among others, as *T. venusta uhrigi*, a Central American taxon (TTWG 2017, 2021; for a more detailed historical account of the species' taxonomy, see Vargas-Ramírez et al. 2017). However, phylogenetic analyses (Vargas-Ramírez et al. 2017; Fritz et al. 2023, 2024) recover *T. medemi* as sister to *T. dorbigni*, which occurs in Brazil, Uruguay and Argentina. This finding prompted a reevaluation of the evolutionary history of the genus and its colonization process into South America. To further our understanding of the phylogenetic position of *T. medemi*, and to generate new information useful for further studies, we aimed to sequence and obtain the entire mitogenome of *T. medemi* using Long-Range PCR and primer walking procedures. Since these are the first complete mitogenomes reported for the species, they significantly augment the genetic information available for *T. medemi* and provide an opportunity to test its phylogenetic position at the mitogenomic scale. In the context of taxonomic uncertainties and identified introgression in the genus *Trachemys* (Fritz et al. 2023, 2024), this data can be critical for further taxonomic and phylogenetic studies.

1.2 Materials and Methods

We used ethanol-preserved blood samples from three *Trachemys medemi* deposited at the Banco de Tejidos de la Biodiversidad Colombiana (BTBC), Instituto de Genética, Universidad Nacional de Colombia. The museum codes for these samples are BTBC12643, BTBC13199, and BTBC13207. Samples were collected from wild turtles captured during a population genetics assessment on the Biogeographic Chocó of Colombia: the Ciénaga de Napipí, Chocó (BTBC12643), Ciénaga de Necoclí, Antioquia (BTBC13199) and Las Brisas, Belén de Bajirá, Chocó (BTBC13207). To ensure the utility of this data for future phylogeographic and population genetic analyses, we selected three individuals from different localities. DNA was extracted using the innuPREP DNA Mini Kit 2.0 (Analytik Jena), with a final elution of 80 µl milliQ water.

The complete mitochondrial genome was amplified using long-range PCR followed by primer-walking procedures. Using primers from Fritz et al. (2012, 2023), two long-range PCR reactions (LR1 and LR2) were performed, yielding amplicons with lengths of 11824 bp and 6797 bp, respectively, which overlap by more than 1000 bp. Long-range PCR products were purified using the ExoSAP-IT enzymatic clean-up (USB Europe GmbH, Staufen, Germany). These amplicons cover most of the mitochondrial genome, except for

the first part of the tRNA-Phe and the final 3'-end of control region. These regions were amplified using standard PCR procedures and custom primers designed from the consensus sequence of the primer-walking PCR products. The cleaned products were PCR amplified using a set of primers compatible with the mitochondrial genome of *Trachemys scripta elegans* (MW019443; Supplementary Table A1). The initial batch of sequences were mapped onto a mitogenome sequence from *T. s. elegans* (MW019443) and curated using Geneious R7 (<http://geneious.com>). From this assembly, new pairs of primers were designed to close the gaps between the sequences. This process was repeated twice, resulting in a total of 14 newly designed primers (Supplementary Table A2). Detailed laboratory procedures are provided in the Supplementary Information. Mitogenomes were annotated with MITOS (Bernt et al. 2013) through the Proksee server (Grant et al. 2023).

Initiation and termination codons were identified using ORFfinder (<https://www.ncbi.nlm.nih.gov/orffinder>). AT skew and GC skew were calculated to describe base composition (Perna and Kocher 1995).

Phylogenetic analyses were conducted using two datasets. The first dataset included *Trachemys* mitogenomes obtained from Russel and Beckenbach (2008), Yu et al. (2014), Park et al. (2021), Ryu et al. (2021), and Fritz et al. (2023, 2024), alongside with mitogenomes from *Pseudemys concinna* (OM935747; Park et al. 2022) and *Graptemys ouachitensis* (OP115973) as outgroups. Protein-coding genes were screened for internal stop codons in Geneious R7 (<http://geneious.com>). Mitogenome alignment was performed

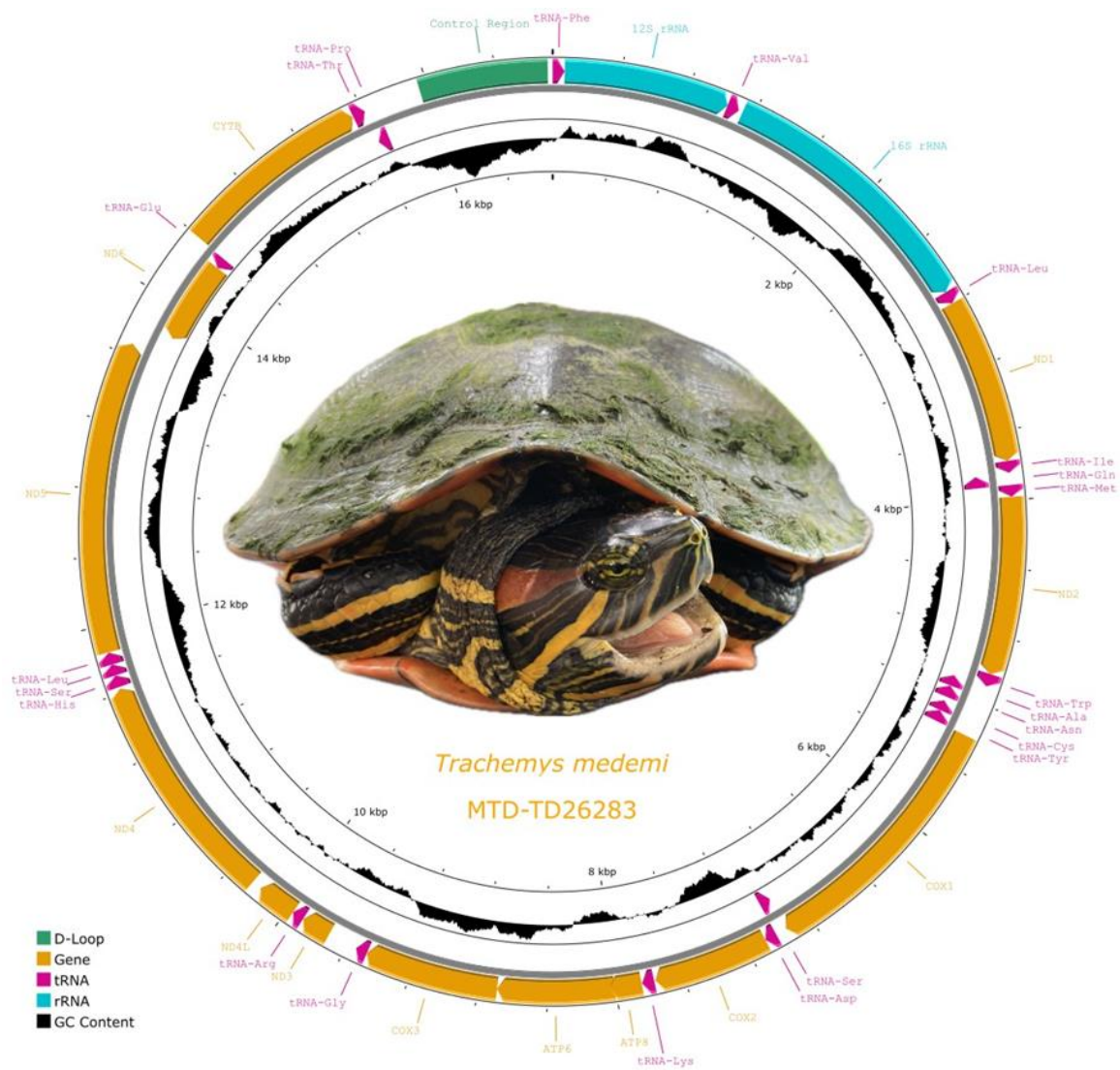


Figure 1. Circular view of the complete annotated Atrato Slider (*Trachemys medemi*) mitogenome, displaying 13 protein-coding genes, 22 tRNA genes, and the control region. Inset photo: *T. medemi*, adult female from Ciénaga de Napipí, Chocó, Colombia (BTBC12643). Photo: Sebastián Cuadrado-Ríos.

with MUSCLE (Edgar 2004) and manually curated to correct alignment errors. Problematic sequence features, such as internal stop codons in coding genes, genes or tRNAs overlapping regions, frame shift inducing in coding region, and non-coding spacer DNA were identified and 232 positions were removed from the alignment. For the second dataset, our dataset was concatenated with the mtDNA alignment used by Fritz et al. (2024) covering almost all continental *Trachemys* taxa, except for *T. hartwegi*. Their dataset comprised 3221 bp and included sequences of the partial 12S gene and the complete ND4L, ND4, and *cyt b* genes plus part of the adjacent tRNA-Thr gene. Missing data was coded as Ns. The best partitioning schemes and evolutionary models (Supplementary Tables S4, S5) were determined by using a greedy search and the Bayesian Information Criterion in PartitionFinder 2 (Lanfear et al. 2017). For the extended mtDNA dataset, PartitionFinder 2 detected a single partitioning scheme with HKY+G+I model as the best fit. This partition scheme was used for the complete mitogenome dataset.

Phylogeny was analyzed using maximum likelihood (ML) and Bayesian inference (BA). ML analyses were performed in IQ-Tree 1.6.12 (Nguyen et al. 2015) with 1000 non-parametric thorough bootstrap replicates to assess node support, plotted against the best tree obtained. Bayesian analyses were conducted in MrBayes 3.2.6 (Ronquist et al. 2012), with two simultaneous runs and four chains each, sampling every 1000th generation for 50 million generations. After a burn-in of 25%, convergence was assessed by verifying that the average standard deviation of split frequencies were below 0.01 and the effective sample size exceeded 200 for all parameters using Tracer 1.7.1 (Rambaut et al. 2018).

1.3 Results

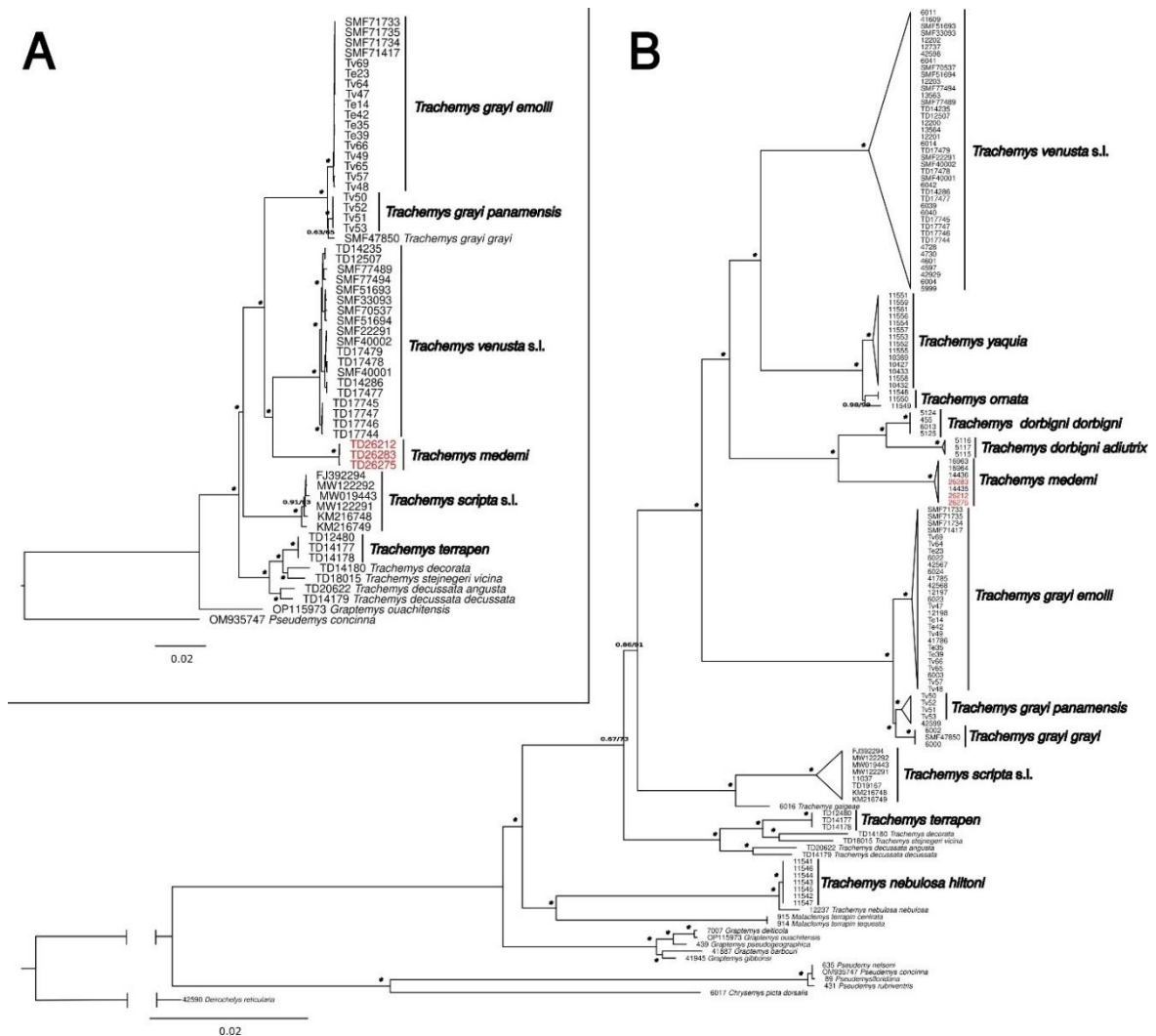
We obtained three complete mitogenomes of *Trachemys medemi*, with lengths ranging from 16711 to 16810 bp. The length variation was due to the absence of a 92 bp repetitive region in the Control Region in one of the mitogenomes. BLAST searches of the mitochondrial genes previously reported for the species (Vargas-Ramírez et al. 2017) confirmed the species identification. The mitogenome of *T. medemi* is composed of 13 protein-coding genes, 2 ribosomal RNA genes, 22 transfer RNA genes and a non-coding control region (Figure 1) as in other *Trachemys* species and other species of the family



Figure 2. Part of the alignment of the three *Trachemys medemi* mitogenomes obtained in the present study, illustrating the absence of a 92-bp-long duplicated sequence in the control region of MTD-TD26283, spanning from position 15861 to 15952.

Emydidae (Fritz et al. 2023, 2024; Ren et al. 2024). Nine of the 37 genes are coded on the light strand, and the remaining genes on the heavy strand. These mitogenomes differ in two mutations in the ND1 gene, one mutation in the ND2 gene, two mutations in the ND4L gene, two mutations in the ND4 gene, one mutation in the ND5 gene, one mutation in ND6, and one mutation and the absence/presence of a duplicated region in the Control Region (Figure 2). The genome composition is as follows: A: 34.3%, C: 26.1%, G: 12.8%, T: 26.8%, showing a slight AT bias (61.0%), with a positive AT skew (0.12274) and a negative GC skew (-0.34190). The AT skew falls within the range of other *Trachemys* species (Supplementary Table A3) but is higher than most emydid mitogenomes. The GC skew is also higher than most of the other previously sequenced emydid mitogenomes (Ren et al. 2024). Like other *Trachemys* species, the GC content was consistent across the mitogenome, ranging from 38.1% to 39.4%, except by the control region, which had a GC content of 30.4%.

From the 22 tRNA genes in the mitogenome of *T. medemi*, 14 are coded on the heavy strand and eight are coded on the light strand (Figure 1). tRNA genes ranged between 67 and 75 bp, exhibiting a positive AT skew (0.12804) and A+T bias (61.9%; Supplementary Table S2). The 12S rRNA is located between the initial tRNA-Phe and tRNA-Val, while the 16S rRNA is positioned between tRNA-Val and tRNA-Leu (Figure 1). They are 976 bp and 1635 bp long, respectively, and show A+T bias (60.6%), positive AT skew (0.28239) and a negative GC skew (-0.1608; Supplementary Table A2). The control region (CR) is situated between the tRNA-Pro and tRNA-Phe genes. In the mitogenome of MTD-TD26283, the CR exhibits a different length due to the absence of a 92 bp repetitive region (Figure 2). The CR has positive AT skew (0.00193) and a negative GC skew (-0.26605). The mitogenomic



characteristics of *T. medemi* and other species are detailed in the Supplementary Table S2.

Figure 3. Maximum likelihood trees estimated with (A) near-complete mitogenomes and (B) complete mitogenomes and data assembled by Fritz et al. (2024), which covers most of the Neotropical *Trachemys* species. The newly sequenced *T. medemi* mitogenomes are highlighted in red. Numbers at nodes represent posterior probabilities from a Bayesian phylogeny (left) and bootstrap values from the ML tree (right). Asterisks indicate maximum support from both approaches. Some clades are collapsed in (B) for clarity.

Phylogenetic analyses using complete mitogenomes resulted in similar topologies with robust support for the main clades (Figure 3A). Both analyses placed *T. medemi* as sister to *T. venusta* s.l., with both species forming a clade that is sister to *T. grayi*. When the mitogenomes were aligned with the 3221-bp-long mtDNA dataset from Fritz et al. (2024), the resulting phylogenies showed similar topologies with also robust support for main clades (Figure 3B). These phylogenies placed now *T. medemi* as sister species of *T. dorbigni*, with this clade being sister to a clade composed of *T. venusta* s.l., *T. yaquia* and *T. ornata* (Figure 3B). The phylogenetic position of *T. terrapen*, *T. decorata*, *T. stejnegeri*, and *T. decussata* were weakly resolved (Figure 3B).

1.4 Discussion

In this study, we sequenced, assembled and characterized three new mitogenomes of *Trachemys medemi*, representing the first complete mitogenomes for the species. One mitogenome (MTD-TD26283) lacks a duplicated region between positions 15861 and 15952 (Figure 2) in the right domain of the Control Region according to Bernacki and Kilpatrick (2020); the turtle was captured in the Ciénaga de Napipí Bojayá, Chocó, Colombia, at the southernmost edge of the species' distribution range (Vargas-Ramírez et al. 2017). Furthermore, an examination of the mitogenome alignment revealed that this phenomenon also occurs in *T. venusta*: the only mitogenome published for *T. v. iversoni* (OZ038161), which does not significantly differ from the mitogenomes of *T. v. venusta* and *T. v. uhrigi* (Fritz et al. 2024), misses the same duplicated region as MTD-TD26283. It remains to be tested whether this phenomenon has population genetic or phylogenetic implications.

Phylogenetic analyses based on these complete mitogenomes (Figure 3A), recovered *T. medemi* as sister to *T. venusta* s.l in a maximally supported clade. When additional mtDNA sequences are included, *T. medemi* was sister to *T. dorbigni*, with these two species together representing the sister clade of another clade composed of *T. venusta* s.l., *T. yaquia*, and *T. ornata* (Figure 3B). Thus, our results confirm previous phylogenetic analyses based on shorter mtDNA sequences (Vargas-Ramírez et al. 2017; Fritz et al. 2023, 2024). Previous phylogenies based on mtDNA sequences (Vargas-Ramírez et al. 2017, Fritz et al. 2023, Fritz et al. 2024) recovered *T. medemi* and *T. dorbigni* as sister to a clade

comprised by *T. venusta* s.l., *yaquia*, *T. ornata* and *T. grayi*. The phylogenetic networks recovered by Fritz et al. (2023) and Fritz et al. (2024), based on nuclear DNA sequences, failed to resolve the phylogenetic position of *T. venusta*, with several subspecies scattered across the cluster “South” and *T. v. callirostris*, *T. v. chichiriviche* and *T. v. uhrigi* in the same subcluster as *T. medemi* and *T. dorbigni*. The phylogenetic relationships between species from the cluster “South” remain unclear, and more genomic evidence, such as mitogenomes and genome-wide nuclear markers, is needed to resolve the taxonomy and evolutionary history of the genus.

Acknowledgements

We thank Anke Müller and Andrea Criado-Flórez (Senckenberg Natural History Collections Dresden) for their guidance and assistance with laboratory work.

1.5 References

- Bernacki, L. E., and Kilpatrick, C. W. (2020). Structural variation of the turtle mitochondrial control region. *Journal of Molecular Evolution*, 88(7), 618-640.
- Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsche, G., ... and Stadler, P. F. (2013). MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular phylogenetics and evolution*, 69(2), 313-319.
- Fritz, U., Herrmann, H. W., Rosen, P. C., Auer, M., Vargas-Ramírez, M., and Kehlmaier, C. (2024). *Trachemys* in Mexico and beyond: Beautiful turtles, taxonomic nightmare, and a mitochondrial poltergeist (Testudines: Emydidae). *Vertebrate Zoology*, 74, 435-452.
- Fritz, U., Stuckas, H., Vargas-Ramírez, M., Hundsdörfer, A. K., Maran, J., and Päckert, M. (2012). Molecular phylogeny of Central and South American slider turtles: implications for biogeography and systematics (Testudines: Emydidae: *Trachemys*). *Journal of Zoological Systematics and Evolutionary Research*, 50(2), 125-136.
- Grant, J. R., Enns, E., Marinier, E., Mandal, A., Herman, E. K., Chen, C. Y., ... and Stothard, P. (2023). Proksee: in-depth characterization and visualization of bacterial genomes. *Nucleic Acids Research*, 51(W1), W484-W492.

Paez, V. P., Bock, B. C., Alzate-Estrada, D. A., Barrientos-Munoz, K. G., Cartagena-Otalvaro, V. M., Echeverry-Alcendra, A., ... and Vallejo-Betancur, M. M. (2022). Turtles of Colombia: an annotated analysis of their diversity, distribution, and conservation status. *Amphibian and Reptile Conservation*, 16(1), 106-135.

Perna, N. T., and Kocher, T. D. (1995). Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *Journal of molecular evolution*, 41(3), 353-358.

Ren, Y., Zuo, Z., Xing, X., Zhai, X., Wang, T., and Ma, G. (2024). The characterization of the mitochondrial genome of *Graptemys ouachitensis*. *Mitochondrial DNA Part B*, 9(5), 563-567.

TTWG [Turtle Taxonomy Working Group]. (2017). Turtles of the world. Annotated checklist and atlas of taxonomy, synonymy, distribution, and conservation status (8th edition). *Chelonian Research Monographs*, 7, 1–292. doi:10.3854/crm.7.checklist.atlas.v8.2017

TTWG [Turtle Taxonomy Working Group]. (2021). Turtles of the world. Annotated checklist and atlas of taxonomy, synonymy, distribution, and conservation status (9th edition). *Chelonian Research Monographs*, 8, 1–472. doi:10.3854/crm.8.checklist.atlas.v9.2021

Vargas-Ramírez, M., del Valle, C., Ceballos, C. P., and Fritz, U. (2017). *Trachemys medemi* n. sp. from northwestern Colombia turns the biogeography of South American slider turtles upside down. *Journal of Zoological Systematics and Evolutionary Research*, 55(4), 326-339.

2. Capítulo 2

Phylogeography of *Kinosternon leucostomum* (Testudines: Kinosternidae) in the Chocó: Biogeographic Barriers, Connectivity, and Isolation-by-Distance

Capítulo en preparación para ser sometido a *Zookeys*.

Abstract

The Biogeographic Chocó (a humid lowland corridor in northwestern South America) hosts exceptional biodiversity but its remains underrepresented in genetic studies of its fauna and flora. We examined the phylogeographic affinities of the Chocoan populations of the white-lipped mud turtle (*Kinosternon leucostomum*), combining newly generated mitochondrial and nuclear DNA sequence data with recently published datasets for the species' distribution range. Using analyses of molecular variance (AMOVA), redundancy analysis (RDA), Bayesian clustering (FastBAPS), and haplotype networks, we identified shallow mitochondrial divergence primarily shaped by isolation-by-distance. Two major genetic breaks were detected: the Hess Escarpment in Central America and the Río Prado in the Middle Magdalena Valley. Chocoan individuals belong to a widely distributed clade that extends from the Lower Magdalena Valley to mid-Central America. The species' ecological traits likely promote gene flow across fragmented lowland habitats. These results support recognizing *K. leucostomum* as a single, continuously distributed species and highlight geographic distance as the primary driver of its phylogeographic structure.

Key words: Central America, Colombia, South America, taxonomy, box turtle

2.1 Introduction

The Biogeographic Chocó, extending along the Pacific coast of northwestern South America into southeastern Central America, represents one of the Earth's most diverse regions, characterized by exceptional species endemism and distinctive ecological

communities. Like the Atlantic Forest (Mata Atlántica) of South America, this region combines complex topography with persistent climatic stability, including the world's highest rainfall levels (Yepes et al. 2019), to create unique evolutionary arenas that have spawned remarkable biodiversity (Myers et al. 2000; Mittermeier et al. 2005; Antonelli and Sanmartín 2011; FEOW 2021). Recent research increasingly recognizes the Chocó's dual role as both a Pleistocene refuge and a modern cradle of speciation (Jaramillo-Vivanco et al. 2010; Arteaga et al. 2016; Pérez-Escobar et al. 2019; Rangel et al. 2019; González-Orozco 2023), with molecular studies revealing intricate patterns of diversification across elevational gradients (Arteaga et al. 2016), landscape heterogeneity (Cucalón et al. 2022) and historical connections with adjacent biomes (Elías et al. 2023; Sedano-Cruz et al. 2024; van der Kamp & Ortega-Andrade 2024). These findings not only elucidate fundamental diversification processes but also imply conservation needs as the region faces escalating threats from habitat conversion and climate change.

Sampling efforts remain insufficient to capture this vast diversity. Challenges such as terrestrial and riverine transportation, limited access to remote areas, and sociopolitical barriers persist across the Chocó (Ortiz Lancheros 2022; Otálvaro-Marín et al. 2023). These logistical constraints impede inventorying the local biodiversity, creating knowledge gaps that undermine any conservation initiatives.

A recent study by Hurtado-Gómez et al. (2024) examined the phylogeography of the white-lipped mud turtle *Kinosternon leucostomum* Duméril and Bibron, 1851 and presented a robust dataset covering much of its distribution range spanning from Central America to northern South America. Their work included key regions such as the inter-Andean valleys, the Caribbean coastal plains of Colombia, and the humid lowlands of Central America. However, samples from the Chocó region were lacking. The Chocó is as a transition zone between Central and South America, harboring unique freshwater systems that likely drive divergent evolutionary trajectories in aquatic and semi-aquatic taxa. Targeted sampling in this region is thus essential to determine whether the Chocoan populations of *K. leucostomum* represent distinct genetic lineages, as observed in other taxa (references, some examples).

Kinosternon leucostomum is a semi-aquatic turtle species that exhibits broad ecological plasticity. It inhabits diverse freshwater environments, including slow-moving rivers, creeks, swamps and artificial ponds, often in areas with varying degrees of forest coverage

(Cogălniceanu et al. 2015). The species occurs from Veracruz, Mexico, to northern South America, including the inter-Andean valleys of Colombia and the Pacific coast of Perú (de la Fuente et al. 2014; Cogălniceanu et al. 2015; Hurtado-Gómez et al. 2024; TTWG 2025). Two genetically weakly differentiated subspecies are currently recognized: *K. l. leucostomum*, distributed along the Atlantic coast of northern Central America, and *K. l. postinguinale* Cope, 1887, which ranges from southern Nicaragua to northwestern Perú (de la Fuente et al. 2014; Hurtado-Gómez et al. 2024; TTWG 2025). The present study supplements the genetic dataset of Hurtado-Gómez et al. (2024) to assess the taxonomy of Chocoan *K. leucostomum* populations.

2.2 Materials and Methods

Sampling and laboratory methods

Samples of *Kinosternon leucostomum* were collected from 2022 to 2023 across nine sites, including the upper and lower Atrato River and the estuary of the lower Baudó Basin (Fig. B1). Turtles were captured with baited funnel crab traps, after visual inspection of swamps, small creeks and artificial ponds. Blood samples were extracted from the caudal vein and preserved in 70% ethanol. All turtles were released at the capture site after sampling. A total of 11 *K. leucostomum* were sampled at nine sites corresponding to three localities. Five specimens originated from Surikí, municipality of Turbo, department of Antioquia (7.78°N, 76.8°W; Fig. B1- A, site 1); three individuals were from the Ciénaga de Napipí, municipality of Bojayá, department of Chocó (6.62°N, 76.94°W; Fig. B1 - A, site 2); and three additional turtles were trapped along the Pizarro Ecological Trail, municipality of Bajo Baudó, Chocó (4.97°N, 77.36°W; Fig. B1A, site 3).

The samples were deposited in the Banco de Tejidos de la Biodiversidad Colombiana (BTBC) at the Instituto de Genética, Universidad Nacional de Colombia, Bogotá. Genomic DNA was extracted using the innuPREP DNA Mini Kit 2.0 (Analytik Jena), with a final elution of 80 µl milliQ water. For PCR amplifications and Sanger sequencing, we followed the approaches described in Hurtado-Gómez et al. (2024) for the mitochondrial cytochrome b (CYTB), 12S and 16S rRNA genes, as well as for the nuclear marker R35 (Supplementary Tables B1, B2). PCR products were purified using the ExoSAP-IT enzymatic clean-up (USB Europe, Staufen) and resolved on an ABI 3500 Genetic Analyzer (Applied Biosystems).

Sequences were manually edited and aligned with the Geneious R7 sequence visualizer (Biomatters Ltd., Auckland) with the MAFFT algorithm (Kato and Toh 2010). For each marker, we aligned our sequences with those from Hurtado-Gómez et al. (2024), resulting in 39 mitochondrial 12S–16S and 36 CYTB sequences and 31 nuclear R35 sequences, covering the species' distribution range.

Samples from Kinosternoidea species, including Dermatemydidae and Kinosternidae, were used as outgroups for phylogenetic analyses (Supplementary Table B3). Analyses were performed with three datasets: (1) the protein-coding CYTB gene, (2) the genes coding for 12S and 16S (12S-16S), and (3) the nuclear marker R35.

Barrier detection, geographic structuring and phylogenetic analyses

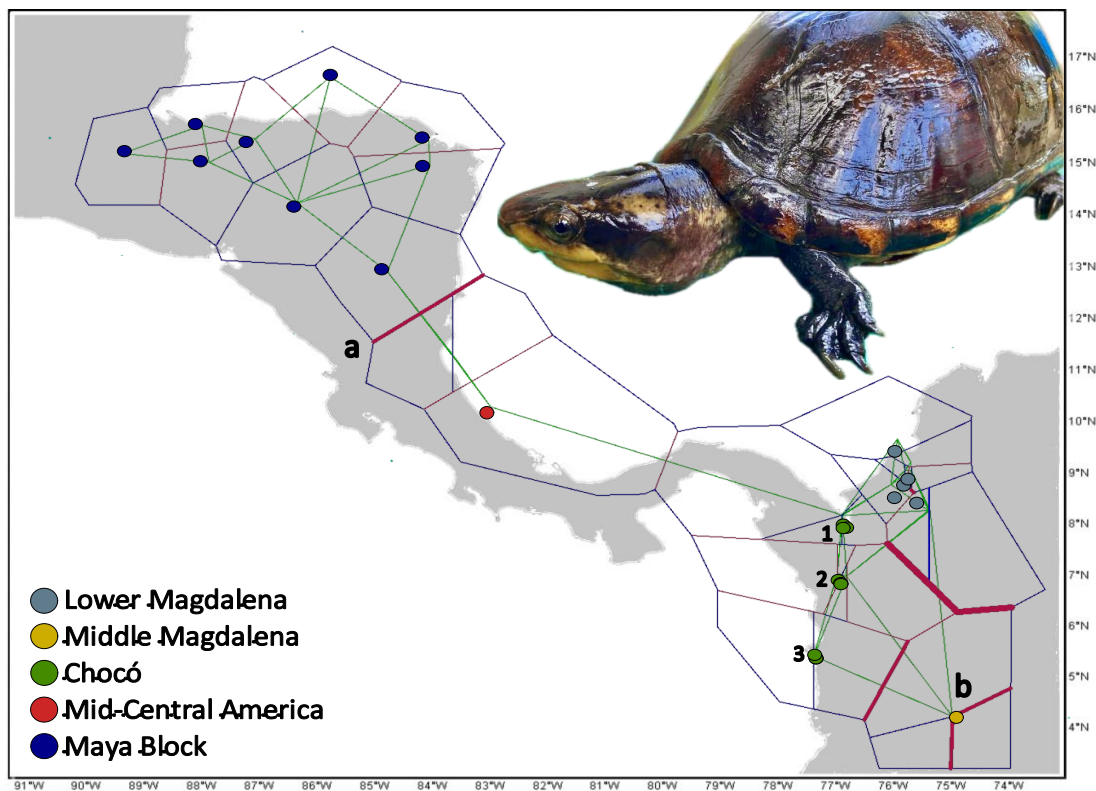


Figure B1. Results of Monmonier's algorithm identifying geographic barriers across the current distribution of *Kinosternon leucostomum*, overlaid with genetic sampling locations and clades analyzed in this study. Blue lines depict Voronoi tessellation of localities based on geographic provenance, while red lines indicate genetic barriers identified via bootstrap permutations. The two primary genetic barriers (thick red lines) correspond to the Hess Escarpment (a) and the Andes Mountain range (b). Numbers refer to sampling localities (see Materials and Methods section). In set photo: Female *K. leucostomum* from the Ciénaga de Napipí, municipality of Bojayá, Chocó - Colombia.

We performed independent analyses for each dataset to assess potential gene flow barriers among sampling sites. Barriers were identified using Monmonier's maximum difference algorithm in BARRIER v2.2 (Manni et al. 2004). To evaluate the robustness, we generated 100 replicates pairwise difference matrices for each dataset and overlapped the resulting barriers to identify concordant patterns. To better understand the geographic distribution of genetic diversity, we constructed parsimony networks for each dataset using PEGAS 1.2 (Paradis 2010) in R 3.2. For comparative analyses, individuals were assigned to five biogeographic regions based on geological and sampling criteria: (1) Middle Magdalena Valley, which includes samples from the upper Río Prado, a Magdalena River tributary draining the eastern slopes of the Central Cordillera, (2) Pacific lowlands of Colombia (Chocó Biogeographic Region), which corresponds to the humid tropical forest of western Colombia, (3) Lower Magdalena, encompassing the lowland floodplains and deltaic systems of the lower Magdalena River in northern Colombia, (4) mid-Central America (the Chorotega block), which includes populations from Costa Rica to northern Panama, characterized by active and inactive volcanic ranges, and (5) Maya Block, defined as the region between the Isthmus of Tehuantepec in Mexico and the Motagua-Polochic-Jocotán fault system, a tectonically active zone where complex uplift and transpression dynamics have shaped the landscape since the Eocene (Gutiérrez-García and Vásquez-Domínguez, 2013).

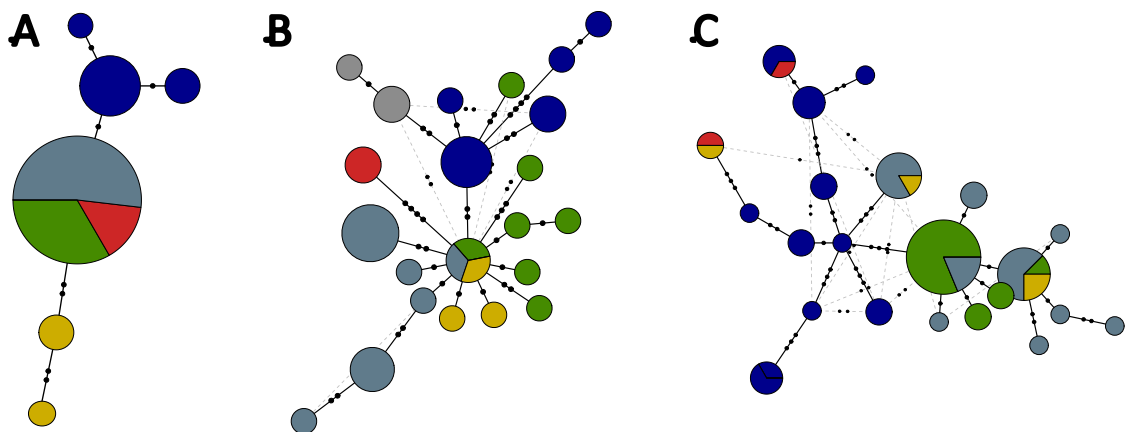


Figure B2. (A) Haplotype network derived from the concatenated 12S and 16S mitochondrial alignment. (B) Haplotype network based on the mitochondrial CYTB gene alignment. (C) Haplotype network constructed from the nuclear R35 fragment. Colors represent geographic regions (see Fig. B1).

We constructed phylogenetic networks using the NeighborNet algorithm in SplitsTree 4.14 (Huson and Bryant 2024) and performed Principal Component Analyses (PCA) of uncorrected p distances calculated in MEGA 11 (Tamura et al. 2021). To assess isolation-by-distance (IBD) patterns, we conducted a Redundancy Analysis (RDA) using genetic distances computed under the Tamura-Nei 93 (TN93) model in the ape package. The resulting matrix was then transformed using a Principal Coordinate Analysis (PCoA) with the “cmdscale” function. The RDA was performed using the “rda” function from the vegan package, with the PCoA axes as the response variable and geographic coordinates as explanatory variables. The significance of the model was assessed with 1,000 permutations using the “anova.cca” function. To assess genetic structure among geographic regions, we performed an Analysis of Molecular Variance (AMOVA) in Arlequin v3.5 (Excoffier and Lischer 2010), using each marker's genetic distance matrix. Statistical significance was tested using 10,000 permutations. To identify clusters of genetically similar individuals, we performed Bayesian clustering analyses using the R package fastbaps (Tonkin-Hill et al. 2019). Analyses were conducted on each alignment independently using both the standard hierarchical clustering algorithm (baps) and the optimized clustering approach (optimised.baps), which implements a model-based optimization of the hierarchical structure.

Furthermore, we calculated for our three datasets and a fourth one, comprised by the mtDNA markers, a phylogeny using the maximum likelihood optimization in IQ-TREE 2.2 (Minh et al. 2020) and determined the optimal partition scheme and substitution models for each dataset using ModelFinder (Kalyaanamoorthy et al. 2017) implemented in IQ-TREE. This approach evaluates models based on the Bayesian Information Criterion (BIC). Branch support was assessed with 1,000 Ultrafast Bootstrap replicates (UFBoot; Hoang et al. 2018) and 1,000 Shimodaira-Hasegawa-like approximate Likelihood Ratio Test (SH-ALrt; Guindon et al. 2010). Consensus trees were annotated and visualized using Figtree (<http://tree.bio.ed.ac.uk/software/figtree/>).

2.3 Results

The overlap of barriers identified independently for each dataset revealed two major zones of genetic discontinuity (Fig. B1). The first break is in Central America, delineating a boundary between the central region and the Maya Block, referred as the Hess escarpment (Fig. B1 – A), and the second one separates populations from the Middle Magdalena Valley

from those in the rest of Colombia (Fig. B1 – B), aligning with the previously recognized subspecies boundaries (Hurtado-Gómez et al. 2024).

Haplotype networks exhibited distinct patterns of diversity across markers (Fig. B2 - A-C). Mitochondrial 12S-16S displayed low haplotype diversity with weak structure, featuring a central haplotype shared among Chocó, Lower Magdalena, and Mid-Central America individuals, while Middle Magdalena and the Maya Block individuals maintained unique haplotypes. The *CYTb* network revealed greater differentiation (Fig. B2 - B), with multiple region-specific haplotypes radiating from two haplotypes shared by several regions, though no clear geographic clustering emerged. The nuclear *R35* network (Fig. B2 - C) showed

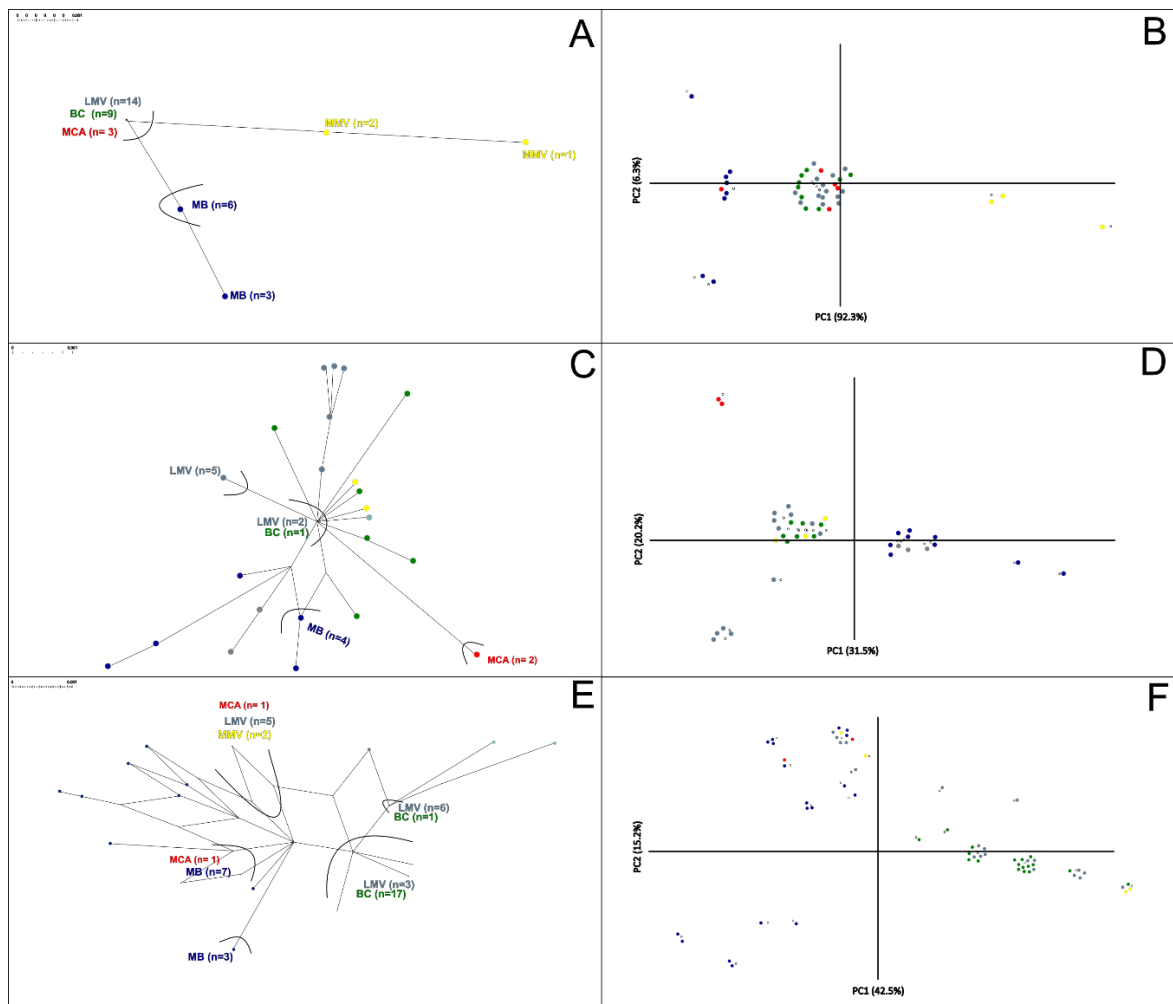


Figure B3. Neighbor-net networks for (A) concatenated 12S-16S, (C) *CYTb*, and (E) *R35*. Right panels: Principal coordinate analyses (PCoA) of (B) 12S-16S, (D) *CYTb*, and (F) *R35*, showing the first two axes (percentages indicate variance explained). Colors represent geographic regions: Maya Block (MB-blue), Middle Magdalena Valley (MMV-yellow), Biogeographic Chocó (BC-green), Lower Magdalena Valley (LMV-light blue), Mid-Central America (MCA-red), and unknown provenance (gray).

extensive reticulation and weak phylogeographic structure, with several haplotypes shared across regions; consistent with previous reports of limited nuclear differentiation between subspecies (Hurtado-Gómez et al. 2024). NeighborNets and PCoAs corroborated these patterns, with distinct clustering of Middle Magdalena and the Maya Block individuals for the 12S-16S, and no structure for the CYTB and R35 (Fig. B3).

Bayesian clustering analyses revealed congruent patterns of genetic structure (Fig. B4). The standard BAPS algorithm identified two primary clusters for both 12S–16S and, versus three for the R35, exhibiting substantial interregional haplotype sharing. The optimized BAPS algorithm detected finer-scale structure, recovering three clusters for 12S–16S, four for the CYTB, and five for R35. Notably, 12S-16S markers showed pronounced differentiation between the Middle Magdalena and those North/South of the Hess escarpment, while CYTB and R35 displayed more complex patterns of haplotype sharing among regions (Fig. B4).

Analyses of molecular variance (AMOVA) revealed a substantial among-population differentiation (Table B1). Mitochondrial markers showed 38.34% (CYTB) and 70.72% (12S-16S) of variation distributed among populations, compared to 30.24% for the nuclear R35 marker. This population structure corresponds with recognized subspecies boundaries in *K. leucostomum*. The redundancy analysis (RDA) indicates that geographic distance is a significant predictor of genetic differentiation across the CYTB and 12S-16S datasets (Table B1). For CYTB, the constrained ordination significantly explained 25.3% of the total genetic variance (adjusted $R^2 \approx 0.25$, $F = 5.07$, $P = 0.001$). While, the 12S-16S dataset showed that geographic distance significantly explained around 47.9% of the total variance (adjusted $R^2 \approx 0.48$, $F = 16.71$, $P = 0.001$). No significant IBD pattern was detected for the nuclear R35 marker ($P > 0.05$). These results demonstrated a significant pattern of IBD, where genetic similarity decreases with increasing geographic distance. Phylogenetic analyses did not recover any type of geographic structuring, instead resolving the ingroup branches as large polytomies (Supplementary Figures B1-B4).

2.4 Discussion

The biogeographic history of *Kinostemon leucostomum* and related species has been progressively clarified through molecular studies. Iverson et al. (2013) initially documented

a well-supported Miocene radiation of *K. leucostomum* (along with *K. dunnii* and *K. angustipons*) into southern Central America and South America. Subsequent work by Hurtado-Gómez et al. (2024), provided more precise divergence estimates, dating the split between *K. leucostomum* and *K. dunnii* to 8.8 million years ago [mya] (95% HDP 4.7-15.7 mya). Biogeographic analyses using Bayesian Binary MCMC analysis in Iverson et al. (2013) suggested a Central/South America distribution for the common ancestor of *K. leucostomum*, *K. angustipons* and *K. dunnii*, with Central American origins proposed for a

Table B1. Population genetic structure and isolation-by-distance patterns in *Kinostemon leucostomum* analyzed across mitochondrial (*CYTB*, 12S-16S) and nuclear (*R35*) markers. Asterisks (*) denote *Fst* values statistically significant at $P < 0.001$. RDA quantifies the proportion of genetic variance explained by geographic distance, with $P < 0.001$ indicated by *.

	DNA markers	CYTB	12S-16S	R35
AMOVA	Sum of Squares (Among populations)	24.711	11.865	24.281
	Variance Component (Among populations)	0.80399	0.38480	0.44964
	Percentage of Variation (Among populations)	38.34	70.72	30.24
	Sum of Squares (Within populations)	36.198	5.417	59.122
	Variance Component (Within populations)	129.280	0.15931	103.723
	Percentage of Variation (Within populations)	61.66	29.28	69.76
	Total Variance	209.679	0.54411	148.687
FST		0.38344*	0.70720*	0.30240*
RDA	Total Variance	0.00001028	0.000001536	0.00001039
	Constrained Variance	0.000002596	0.0000007363	0.000001598
	% Explained (R²)	25.3%	47.9%	15.4%
	Adjusted R² (≈)	0.25	0.48	0.15
	F-value	5.07*	16.71*	2.87
	Significant Constrained Axes	2	1	1
	% Var RDA1	82.9%	100%	100%
	% Var RDA2	17.1%	–	–

more inclusive clade (*Cryptochelys*) also containing two further Central American species (*K. creaseri*, *C. herrera*). However, these results might have been affected by a putative past mitochondrial capture detected by Hurtado-Gómez et al. (2024), who recovered *K. leucostomum* as sister to *K. dunni*, in contrast to the sister group relationship of *K. dunni* to *K. angustipons* inferred by Iverson et al. (2013).

The fossil record from the Panama Canal Basin documents early Miocene dispersals of Central/North American turtle clades —*Rhinoclemmys* (Geoemydidae), *Staurotypus* (Kinosternidae) and trionychids— into the Neotropics prior to the complete closure of the Isthmus of Panama (Cadena et al. 2012). Yet, our genetic data from *K. leucostomum* reveals shallow divergence across its range that aligns with patterns in other lowland reptiles (Jiménez-Alonso et al. 2023), lowland birds (Harvey and Brumfield 2015), and freshwater fishes (Bermingham and Martin 1998).

This weak phylogeographic structure suggests either ongoing gene flow or recent immigration across ecological convergent regions: Central America lowlands, the Biogeographic Chocó and the Magdalena River valley. In the Magdalena Valley and

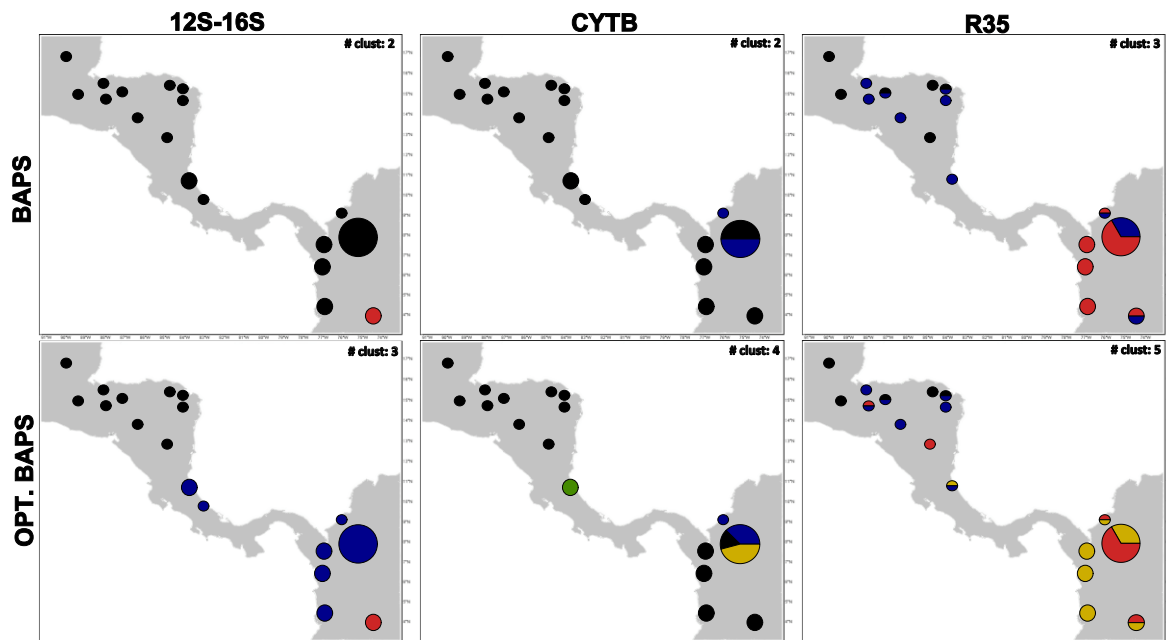


Figure B4. FastBAPS analysis of *Kinosternon leucostomum* population structure. Top row: Standard hierarchical BAPS clustering; bottom row: Optimized BAPS results. Sampling localities are shown with circle sizes proportional to sample size and pie charts representing cluster assignment proportions. Numbers in top-right corners indicate inferred clusters (colors are arbitrary and do not correspond to geographic regions).

the Biogeographic Chocó, long-term Andean orogenesis has generated broad floodplains and sediment-rich braided rivers, maintaining habitat continuity for riparian and forest-dwelling species (Hoorn et al. 2010; Montes-Rojas et al. 2024). Likewise, the lowland forests of Central America (from the Chorotega volcanic ranges to the Caribbean slopes) exhibit comparable vegetation structure and hydrological regimes, facilitating dispersal along both Pacific and Atlantic coasts (Condit et al. 2002; Fagua and Ramsey 2019). The convergence of these environmental features across what are today separate biogeographic corridors underpins the recurrent pattern of shallow genetic divergence observed in *K. leucostomum* and other Neotropical organisms.

The weak phylogeographic structure of *K. leucostomum* may reflect an interplay between ecological adaptability and landscape-mediated evolutionary processes. This species demonstrates remarkable dispersal capacity, including extensive overland movements during rainy seasons and estivation in riparian forests during dry periods (Rueda-Almonacid et al., 2007; Giraldo et al., 2012). In addition, it is a trophic generalist that can persist in diverse habitats from pristine streams to disturbed systems (Castaño-Mora et al., 2005; Ceballos et al., 2016). Despite these traits promoting connectivity, our analyses reveal two major genetic discontinuities: (1) at the Hess Escarpment in Central America, where the Motagua-Polochic-Jocotán fault system creates a topographic and ecological transition maintaining the subspecies boundary between *K. l. leucostomum* and *K. l. postinguinale* (Hurtado-Gómez et al., 2024), and (2) in the Middle Magdalena Valley, where Andean foothills and dynamic fluvial processes restrict gene flow (Hoorn et al., 2010; Figs. 1-2). Following the Miocene uplift of the Central and Eastern Andes, the Magdalena Basin experienced a major reorganization of its drainage systems, leading to the formation of independent sub-basins and persistent lowland discontinuities (Hoorn et al., 2010). These tectonic and geomorphological changes reshaped river courses, altered floodplain connectivity, and created alternating zones of sediment accumulation and incision that remain evident today. Such landscape transformations likely intensified the separation between the Upper and Middle Magdalena regions, reducing hydrological continuity and increasing the probability of localized genetic differentiation in freshwater and semi-aquatic organisms. For *K. leucostomum*, which relies on lowland riparian corridors for dispersal, these Miocene-derived drainage rearrangements would have acted as partial filters, constraining movement over evolutionary timescales while still allowing the shallow divergence patterns observed across its range.

In the studied mitochondrial markers, IBD patterns are evident, while the nuclear R35 loci shows an even weaker structure without any IBD signal, consistent with its slower mutation rate (Hurtado-Gómez et al., 2024). The genetic identity of the Chocó populations with cis-Andean/Middle American populations extending to the Hess Escarpment suggests gene flow along river corridors with periodic constraint at major topographic breaks. These patterns mirror those of co-distributed taxa, such as *Rhinoclemmys melanosterna* and *Caiman crocodilus* (Vargas-Ramírez et al., 2013; Díaz-Moreno et al., 2021), and suggests that *K. leucostomum* constitutes a single, widely distributed species in which the observed genetic structure primarily reflects distance-dependent dispersal rather than deep historical vicariance. Future genomic studies should aim to establish more precise migration times for *K. leucostomum* populations in relation to Miocene–Pleistocene geological and climatic events, and investigate whether populations inhabiting ecotones (e.g., Pacific–Atlantic transitions) show signatures of adaptive divergence. Such integrative approaches would clarify how landscape history and ecological gradients jointly shape phylogeographic patterns in Neotropical freshwater turtles.

Acknowledgements

Funding and support were provided by the Turtle Conservation Fund (TCF). SC-R is supported by the “Convocatoria del Fondo de Ciencia, Tecnología e Innovación del Sistema General de Regalías, en el marco del Programa de Becas de Excelencia Doctoral del Bicentenario – Corte II”, Department of Chocó, Universidad Nacional de Colombia and Ministerio de Ciencia, Tecnología e Innovación de Colombia and by a “Research Grant (bi-nationally supervised Doctoral Degree/Cotutelle, 2023/24)” from the German Academic Exchange Service (DAAD). Turtles were manipulated under University research permits granted by the Ministry of Environment and Sustainable Development of Colombia. During manipulation, we followed the “Guidelines for live amphibians and reptiles in field and laboratory research”, developed by the American Society of Ichthyologists and Herpetologists, 2004.

2.5 References

Antonelli, A., and Sanmartín, I. (2011). Why are there so many plant species in the Neotropics?. *Taxon*, 60(2), 403-414.

Arteaga, A., Pyron, R. A., Peñafiel, N., Romero-Barreto, P., Culebras, J., Bustamante, L., ... and Guayasamin, J. M. (2016). Comparative phylogeography reveals cryptic diversity and repeated patterns of cladogenesis for amphibians and reptiles in northwestern Ecuador. *PLoS One*, 11(4), e0151746.

Bermingham, E., and Martin, A. P. (1998). Comparative mtDNA phylogeography of neotropical freshwater fishes: testing shared history to infer the evolutionary landscape of lower Central America. *Molecular Ecology*, 7(4), 499-517.

Cadena, E., Bourque, J. R., Rincon, A. F., Bloch, J. I., Jaramillo, C. A., and Macfadden, B. J. (2012). New turtles (*Chelonia*) from the late Eocene through late Miocene of the Panama Canal Basin. *Journal of Paleontology*, 86(3), 539-557.

Castaño-Mora, O. V., Cárdenas-Arévalo, G., Gallego-García, N., and Rivera-Díaz, O. (2005). *Protección y conservación de los quelonios continentales en el departamento de Córdoba (Convenio No. 28)*. Universidad Nacional de Colombia, Instituto de Ciencias Naturales; Corporación Autónoma Regional de los Valles del Sinú y San Jorge CVS, Bogotá, Colombia. 185 pp.

Ceballos, C. P., Zapata, D., Alvarado, C., and Rincón, E. (2016). Morphology, diet, and population structure of the southern white-lipped mud turtle *Kinosternon leucostomum* postinguinale (Testudines: Kinosternidae) in the Nus River drainage, Colombia. *Journal of Herpetology*, 50(3), 374-380.

Cogălniceanu, D., Torres-Porras, J., Seoane, J. M., and Lascano, C. A. F. (2015). The southernmost known locality for *Kinosternon leucostomum* (Reptilia, Testudines, Kinosternidae), El Oro province, southern Ecuador. *Check List*, 11(1), 1549-1549.

Condit, R., Pitman, N., Leigh Jr, E. G., Chave, J., Terborgh, J., Foster, R. B., ... and Hubbell, S. P. (2002). Beta-diversity in tropical forest trees. *Science*, 295(5555), 666-669.

de la Fuente, M. S., Sánchez, D. J., and Torres, O. M. (2014). Range extension and natural history notes for *Kinosternon leucostomum* in Colombia. *Check List*, 10(2), 345–350. <https://doi.org/10.15560/10.2.345>

Díaz-Moreno, D. M., Hernández-Gonzalez, F., Moncada-Jimenez, J. F., Mora, C., Prada, C., Jiménez-Alonso, G., and Balaguera-Reina, S. A. (2021). Molecular characterization of

the spectacled caiman (*Caiman crocodilus*) in the upper Magdalena River basin, Colombia: demographic and phylogeographic insights. *Systematics and Biodiversity*, 19(8), 1040-1048.

Excoffier, L., and Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology resources*, 10(3), 564-567.

Fagua, J. C., and Ramsey, R. D. (2019). Geospatial modeling of land cover change in the Chocó-Darien global ecoregion of South America; One of most biodiverse and rainy areas in the world. *PloS One*, 14(2), e0211324.

FEOW (Freshwater Ecoregions of the World). (2021). *Freshwater ecoregions of the world: a global biogeographical regionalization of the Earth's freshwater biodiversity*. <https://feow.org/> (accessed 12/04/2025).

Giraldo, A., Garcés-Restrepo, M., and Carr, J. L. (2012). *Kinosternon leucostomum*. In P. V. Páez, M. A. Morales-Betancourt, C. A. Lasso, O. V. Castaño-Mora, and B. C. Bock (Eds.), *Biología y conservación de las tortugas continentales de Colombia* (pp. 332–339). Serie Editorial Recursos Hidrobiológicos y Pesqueros Continentales de Colombia.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, 59(3), 307-321.

Gutiérrez-García, T. A., and Vázquez-Domínguez, E. (2013). Consensus between genes and stones in the biogeographic and evolutionary history of Central America. *Quaternary Research*, 79(3), 311-324.

Harvey, M. G., and Brumfield, R. T. (2015). Genomic variation in a widespread Neotropical bird (*Xenops minutus*) reveals divergence, population expansion, and gene flow. *Molecular Phylogenetics and Evolution*, 83, 305-316.8

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Molecular biology and evolution*, 35(2), 518-522.

Hoorn, C., Wesselingh, F. P., Ter Steege, H., Bermudez, M. A., Mora, A., Sevink, J., ... and Antonelli, A. (2010). Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*, 330(6006), 927-931.

Hurtado-Gómez, J. P., Vargas-Ramírez, M., Iverson, J. B., Joyce, W. G., McCranie, J. R., Paetzold, C., and Fritz, U. (2024). Diversity and biogeography of South American mud turtles elucidated by multilocus DNA sequencing (Testudines: Kinosternidae). *Molecular Phylogenetics and Evolution*, 197, 108083.

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35(2), 518-522.

Huson, D. H., and Bryant, D. (2024). The SplitsTree App: interactive analysis and visualization using phylogenetic trees and networks. *Nature methods*, 21(10), 1773-1774.

Jiménez-Alonso, G., Balaguera-Reina, S. A., Hoyos, M., Ibáñez, C., Rangel, S. M. H., del Valle Useche, C. M., ... and Bloor, P. (2023). Phylogenetic and phylogeographic insights on Trans-Andean spectacled caiman populations in Colombia. *Marine and Freshwater Research*, 74(12), 1071-1080.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods*, 14(6), 587-589.

Katoh, K., and Toh, H. (2010). Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics*, 26(15), 1899-1900.

Manni, F., Guérard, E., and Heyer, E. (2004). Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. *Human biology*, 76(2), 173-190.

Medem, F. (1962). La distribución geográfica y ecológica de los Crocodylia y Testudinata en el departamento del Chocó. *Revista de la academia de Ciencias Exactas, Físicas y Naturales*, 11, 279-304.

Mendoza, A. M., Bolívar-García, W., Vázquez-Domínguez, E., Ibáñez, R., and Olea, G. P. (2019). The role of Central American barriers in shaping the evolutionary history of the

northernmost glassfrog, *Hyalinobatrachium fleischmanni* (Anura: Centrolenidae). *PeerJ*, 7, e6115.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37(5), 1530-1534.

Morales-Verdeja, S. A., and Vogt, R. C. (1997). Terrestrial movements in relation to aestivation and the annual reproductive cycle of *Kinosternon leucostomum*. *Copeia*, 123-130.

Montes-Rojas, A., Delgado-Morales, N. A. J., Escucha, R. S., Siabatto, L. C., and Link, A. (2024). Recovering connectivity through restoration corridors in a fragmented landscape in the magdalena river's valley in Colombia. *Biodiversity and Conservation*, 33(11), 3171-3185.

Ordóñez-Garza, N., Matson, J. O., Strauss, R. E., Bradley, R. D., and Salazar-Bravo, J. (2010). Patterns of phenotypic and genetic variation in three species of endemic Mesoamerican *Peromyscus* (Rodentia: Cricetidae). *Journal of Mammalogy*, 91(4), 848-859.

Ortiz Lancheros, C. A. (2022). Entre la fragilidad de la paz y la persistencia de la guerra: El caso de la subregión del Bajo Atrato, Chocó, Colombia. *Revista Ratio Juris*, 17(34), 319-342.

Ortega-Gutiérrez, F., Solari, L. A., Ortega-Obregón, C., Elias-Herrera, M., Martens, U., Morán-Icál, S., ... and Schaaf, P. (2007). The Maya-Chortís boundary: a tectonostratigraphic approach. *International Geology Review*, 49(11), 996-1024.

Otálvaro-Marín, B., Parra-López, M. Y., and Klinger-Cundumí, E. (2023). Análisis de las injusticias sociales, ambientales y territoriales del departamento del Chocó, Colombia. *Prospectiva*, (36).

Paradis, E. (2010). pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics*, 26(3), 419-420.

Pérez-Escobar, O. A., Lucas, E., Jaramillo, C., Monro, A., Morris, S. K., Bogarín, D., ... and Antonelli, A. (2019). The origin and diversification of the hyperdiverse flora in the Chocó biogeographic region. *Frontiers in Plant Science*, *10*, 1328.

Rangel, T. F., Edwards, N. R., Holden, P. B., Diniz-Filho, J. A. F., Gosling, W. D., Coelho, M. T. P., ... and Colwell, R. K. (2018). Modeling the ecology and evolution of biodiversity: Biogeographical cradles, museums, and graves. *Science*, *361*(6399), eaar5452.

Rueda-Almonacid, J. V., Carr, J. L., Mittermeier, R. A., Rodríguez-Mahecha, J. V., Mast, R. B., Vogt, R. C., Rhodin, A. G. J., De La Ossa-Velásquez, J., Rueda, J. N., and Mittermeier, C. G. (2007). *Las tortugas y los cocodrilianos de los países andinos del trópico*. Conservación Internacional, Bogotá, Colombia. 538 pp.

Saldarriaga-Córdoba, M., Parkinson, C. L., Daza, J. M., Wüster, W., and Sasa, M. (2017). Phylogeography of the Central American lancehead *Bothrops asper* (SERPENTES: VIPERIDAE). *PLoS One*, *12*(11), e0187969.

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, *38*(7), 3022-3027.

Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S. D., and Corander, J. (2019). Fast hierarchical Bayesian analysis of population structure. *Nucleic Acids Research*, *47*(11), 5539-5549.

Vargas-Ramirez, M., Carr, J. L., and Fritz, U. (2013). Complex phylogeography in *Rhinoclemmys melanosterna*: conflicting mitochondrial and nuclear evidence suggests past hybridization (Testudines: Geoemydidae). *Zootaxa*, *3670*(2), 238-254.

Yepes, J., Poveda, G., Mejía, J. F., Moreno, L., and Rueda, C. (2019). Choco-jex: A research experiment focused on the Chocó low-level jet over the far eastern Pacific and western Colombia. *Bulletin of the American Meteorological Society*, *100*(5), 779-796.

3. Capítulo 3

One river, one turtle: Population genetics of the Atrato slider turtle (*Trachemys medemi*) reveals a panmictic population and low genetic diversity

Capítulo en preparación para ser sometido a *Chelonian Conservation and Biology*.

Abstract

Trachemys medemi is a freshwater turtle endemic to the Atrato River basin in northwestern Colombia. The species faces increasing threats from habitat loss, overexploitation, and environmental change. To evaluate its genetic health, we analyzed mitochondrial control region sequences and 11 nuclear microsatellite loci from 96 individuals across seven localities through its distribution range to assess genetic diversity, population structure, demographic history, and effective population size. Both marker systems revealed shallow genetic structuring, with most variation occurring within populations. We found evidence of low and asymmetric gene flow among localities, and no evidence of recent bottlenecks was detected. Nuclear genetic diversity was low (mean $H_e = 0.479$), and the pooled effective population size was estimated at 245.7 (95% CI: 28.4–1304.5), above the short-term threshold to avoid inbreeding depression but below the benchmark required to maintain long-term adaptive potential. These results indicate that while *T. medemi* retains genetic signatures of historical connectivity, it is vulnerable to genetic drift and demographic instability. Our findings underscore the urgent need for conservation strategies that minimize overexploitation, maintain habitat quality, protect aquatic corridors, and monitor effective population sizes to ensure the species' long-term persistence.

Key words: Atrato slider, Colombia, Control Region, population genetics, Microsatellites.

3.1 Introduction

The Biogeographic Chocó (commonly referred to as Tumbes-Chocó-Magdalena, but called the Biogeographic Chocó hereafter), spans the Pacific lowlands of Ecuador, Colombia, and Panama. This region is one of the wettest on Earth and harbors exceptional levels of biodiversity and endemism (Poveda et al. 2004, Aguirre-C and Rangel-Ch 2008, Mittermeier et al. 2011, Cano et al. 2017, Christenhusz et al. 2017). Within this hotspot, the Atrato River basin in northwestern Colombia forms a complex hydrological network of rivers, streams, and floodplain wetlands embedded within tropical rainforests (Anaya-Acevedo et al. 2017, Palomino-Ángel et al. 2019). This unique ecosystem provides critical habitat for aquatic fauna but it is simultaneously subjected to intense anthropogenic pressures. Gold mining, deforestation, agricultural expansion, and contamination have substantially altered aquatic environments, threatening the region's ecological integrity (Hurtado-Gómez et al. 2015, Calzadilla 2019, Caicedo-Rivas et al. 2022, Bernal-Alviz et al. 2025). Consequently, the Chocó has become a conservation priority at both national and international levels.

Despite its global significance, the fauna of the Biogeographic Chocó remains poorly studied. Many species are severely threatened by overexploitation, habitat transformation, and pollution. This pattern holds true for its freshwater turtles, which play key roles in ecosystem functioning yet are highly vulnerable to hunting, egg collection, and habitat loss (Epperson and Heise 2003, Moll and Moll 2004, Rhodin et al. 2011, TCC 2011, Patino and Estupinan-Suarez 2016, Stanford et al. 2020). Several species in the region are harvested for local consumption or the pet trade, while their eggs are collected from nesting sites, leading to demographic declines (Morales-Betancourt et al. 2015; Páez et al., 2012, Lovich et al. 2018). Moreover, most species have not been the focus of detailed ecological or genetic studies. As a result, their population status, levels of connectivity, and conservation needs are largely unknown. This critical knowledge gap hinders the development of effective, evidence-based conservation strategies for these unique Chelonian lineages. *Trachemys medemi* is a turtle species endemic to the Atrato River basin within the Biogeographic Chocó (Vargas-Ramírez et al. 2017), a distribution notably narrower compared to other *Trachemys* species (TTWG 2025). Morphologically, this species can be distinguished by a brick-red postorbital stripe, and inhabits rivers, streams, and floodplain wetlands (Páez et al. 2022). *Trachemys medemi* is omnivorous, with a diet consisting primarily of aquatic plants and invertebrates (pers. obs.). Alarmingly, it is subject to intensive hunting for local consumption and commercial trade and has been identified as

one of the most frequently trafficked turtles in the region (Asprilla-Perea et al. 2013, Arroyave Bermudez et al. 2014). These anthropogenic pressures pose a significant threat to its already limited populations (Páez et al. 2022). Although not yet been globally assessed by the IUCN, the species' conservation status is a cause for concern.

The genetic profile of *T. medemi* remains largely unknown. Although population genetic studies on other Colombian freshwater turtles have uncovered patterns of diversity and structure influenced by both natural history and human impacts (Vargas-Ramírez et al. 2012, Gallego-García et al. 2018), no such research has been conducted on *T. medemi*. Consequently, critical parameters such as population connectivity, effective population sizes, and historical demography are entirely unexplored. Given the species' restricted range and the ongoing environmental degradation within the Atrato basin, this lack of genetic data represents a major obstacle for formulating conservation planning.

Here, we present the first population genetic study of *Trachemys medemi* utilizing nuclear microsatellites and mitochondrial control region sequences. Specifically, we aim to: (1) quantify genetic diversity within and among populations, (2) assess population structure and connectivity across the basin; and (3) detect signals of demographic expansion or contraction. By addressing this critical knowledge gap, our study provides essential information for the conservation of *T. medemi* and contributes to a broader understanding of freshwater turtle persistence in one of South America's most threatened ecosystems.

3.2 Materials and Methods

Sampling and DNA extraction

A total of 96 tissue samples of *Trachemys medemi* were collected from multiple sampling sites across the species' known distribution range (Fig. 1A). Samples originated through new fieldwork collections and from tissues previously deposited in the Banco de Tejidos de la Biodiversidad Colombiana (BTBC), Instituto de Genética, Universidad Nacional de Colombia. Live turtles were captured by hand, using dip nets, or with baited funnel traps. A small volume of blood (~1 ml) was drawn from each individual prior to its release at the capture site. Tissue samples were also salvaged from opportunistically encountered deceased specimens (e.g., road-killed individuals). All samples were preserved in 96% ethanol and deposited at the BTBC. Sampling was conducted at 12 sites in northwestern

Colombia, later grouped into seven localities based on geographic proximity (Fig. 1A ; Supplementary Table C1): 1) Las Brisas and Llano Rico, Belén de Bajirá, Chocó (n = 27); 2) Ríos Salaquí y Truandó, Riosucio, Chocó (n = 20); 3) Ríos Surikí y León, Turbo, Antioquia (n = 24), 4) Ciénaga de Napipí, Bojayá, Chocó (n = 5); 5) Unguía, Chocó (n = 2); 6) Río Turbo y Ciénaga de Necoclí, Turbo, Antioquia (n = 4), and 7) Caño Pedeguita, Pedega, Chocó (n = 14). Our sampling design encompasses nearly the entire known distribution of *T. medemi*, including the southernmost locality reported at Ciénaga de Napipí (TTWG 2025), the westernmost locality reported at Río Truandó, and close to the easternmost locality reported at Las Brisas (TTWG 2025). Despite extensive sampling effort at Ciénaga de Necoclí, one of the northernmost localities, only a single individual was captured, highlighting the reduced detectability of the species in wetland systems. We grouped individuals from Ciénaga de Necoclí and Río Turbo in the same locality due to their connection via a complex hydrological network along Colombia's Caribbean coast.

Genomic DNA was extracted from all samples using either the InnuPrep DNA Mini Kit or the InnuPrep Blood DNA Mini Kit (Analytik Jena AG, Jena, Germany), following the manufacturer's instructions. To investigate the population genetics of *Trachemys medemi*, we employed a complementary approach by analyzing both the maternally inherited, fast-evolving mitochondrial Control Region (D-loop), and nuclear microsatellites. This dual-marker approach is well-established for assessing genetic variation across a species' distribution and providing critical data for conservation and management (Komoroske et al. 2017).

Mitochondrial Control Region

We targeted the most variable portion of the mitochondrial control region (D-loop), previously identified by Cuadrado-Ríos et al. (2025), which contains segregating sites. This region was amplified using the primers medemitRNA-Phe (5'-TGGGTAACCAATATAACCCCA-3'), medemitRNA-Phe2 (5'-TTAGATTGCTAGGGCGTTTT-3'), and medemiCRend (5'-ACCCACGACAGTAATTTTCA-3') under the following PCR conditions: Initial denaturation at 94°C for 5 min, followed by 30 cycles of 94°C for 30s, 60°C for 45s, 72°C for 30 s, and final elongation at 72°C for 10 min. We successfully amplified a 962 pb fragment of length, encompassing the segregating sites. All sequences were imported, curated and aligned using GENEIOUS R7 (<http://geneious.com>).

A median-joining haplotype network was reconstructed using the R package *pegas* (Paradis 2010). To investigate genetic structure, we applied a Bayesian clustering approach using the R package *fastbaps* (Tonkin-Hill et al. 2019). We used the functions *optimize.baps*, *baps*, and *optimize.symmetric*, to implement hierarchical Bayesian partitioning algorithms to identify the most likely number of genetic clusters. Standard indices of genetic diversity and demographic history were identified using DnaSP v5 (Rozas et al. 2017). For each locality and for the complete dataset, we calculated the number of segregating sites (S), number of haplotypes (h), haplotype diversity (Hd), nucleotide diversity (π), and the average number of pairwise differences (K). To tests for departures from neutrality and infer demographic history, we computed Tajima's D (Tajima 1989) and Fu's FS (Fu 1997) using DnaSP v5 (Librado and Rozas, 2009). The significance of the neutrality statistics was assessed with 10,000 coalescent simulations, considered results significant at $P < 0.05$. Furthermore, we performed mismatch distribution analyses to evaluate signals of demographic expansion. For each analysis, we estimated Harpending's raggedness index (r), Ramos-Onsins and Rozas' R2 statistic, and the expansion parameter (τ). The Unguía locality (n = 1) was excluded from all demographic and diversity analyses.

Population structure was assessed using an analysis of molecular variance (AMOVA) based on pairwise differences between haplotypes in Arlequin v3.5 (Excoffier and Lischer 2010). Significance levels were determined with 1,000 permutations. We also computed pairwise Φ_{ST} values to quantify genetic differentiation between all locality pairs, with significance tested using the same permutation procedure. This framework allowed us to quantify the distribution of genetic variance among versus within localities and to evaluate differentiation at both global and pairwise levels.

Finally, we estimated female-mediated gene flow using the Bayesian coalescent framework in MIGRATE-n v.4.4.3 (Beerli and Felsenstein 2001). A Bayesian analysis was run using one long chain of 1,000,000 sampled steps, recording every 100th genealogy after discarding an initial burn-in of 100,000. Population sizes were estimated as $\Theta = Nef_{\mu}$, where μ is the mutation rate per locus. We assumed a mutation rate of 2.48×10^{-7} substitutions per site per year for the mtDNA control region (Chassin-Noria et al. 2004), multiplied by the effective sequence length (892 bp) to obtain the locus-specific mutation rate. Female effective population sizes (Nef) were then calculated as $\Theta/\mu_{\text{locus}}$. Pairwise gene flow was summarized as $4Nm = \Theta_j \times M_{i \rightarrow j}$, with Θ_j representing the effective size of the receiving population.

Microsatellite loci

To genotype *Trachemis medemi*, we tested 14 microsatellite loci originally developed for *Trachemys scripta* (Simison et al. 2013). The applicability of these loci in *T. medemi* was preliminary demonstrated by Balcerro-Deaquiz (2022), who successfully amplified 13 of the 14 loci using laboratory procedures established for *T. venusta callirostris*. Cross-amplification of these loci in other *Trachemys* is common, e.g. the same loci has been recently amplified in *T. hartwegi* (Becerra et al. 2025). We adapted the laboratory procedures and PCR conditions from Balcerro-Deaquiz (2022); loci were organized into three multiplex PCR reactions and amplified using the Qiagen Type-it Microsatellite PCR Kit (QIAGEN GmbH), following the protocol of Balcerro-Deaquiz (2022). Thermocycling conditions consisted of an initial denaturation at 95 °C for 5 min, followed by 30 cycles of 95 °C for 30 s, 55 °C for 90 s, and 72 °C for 30 s, and a final elongation step at 60 °C for 30 min (Standfuss et al. 2016).

To validate the amplifications, PCR products were first visualized to confirm success. For selected individuals, the presence of the expected repeat motifs was confirmed by bidirectional sequencing. Of the 14 loci tested, 11 were polymorphic in *Trachemys medemi* (see Supplementary material). PCR products were purified with ExoSAP-IT (Thermo Fisher Scientific, Waltham, USA) and sequenced using each locus' reverse primer with the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, USA) on an ABI 3130xl Genetic Analyzer. Amplified fragments were sized on the same instrument, and allele lengths were scored using PEAK SCANNER v1.0 (Applied Biosystems). All 11 loci were successfully amplified and confirmed as polymorphic in *T. medemi*.

Tests for Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) were conducted in GENEPOP v.4.8.3 (Raymond and Rousset 1995). Deviations from HWE for each locus were evaluated using Fisher's exact probability test with a Markov Chain Monte Carlo (MCMC) algorithm (Guo and Thompson 1992), applying 10,000 dememorization steps, 20 batches, and 5,000 iterations per batch. The same parameters were used to test for pairwise LD among loci. To account for multiple comparisons, Bonferroni corrections were applied to both HWE and LD tests, using a significance threshold of $p \leq 0.05$ (Rice, 1989).

Null allele frequencies were estimated with the Expectation–Maximization (EM) algorithm (Dempster et al., 1977) implemented in FreeNA (Chapuis and Estoup 2007), based on 25,000 replicates. To evaluate potential bias caused by null alleles, the ENA (Excluding Null Alleles) correction was applied to F_{ST} estimates (see below). Results were summarized in a locus-by-population matrix to visualize the distribution of null alleles across localities. We also calculated Cavalli–Sforza and Edwards’ chord distance (D_c) among all locality pairs using FreeNA, both with and without the Individual Null Allele (INA) correction. Global estimates were based on all 11 microsatellite loci, and pairwise matrices were obtained for corrected and uncorrected distances.

Genetic diversity and Population Structure

We calculated a suite of population genetic diversity indices using the R environment (v.4.3.2). Genetic data were initially imported and handled as a genind object using the

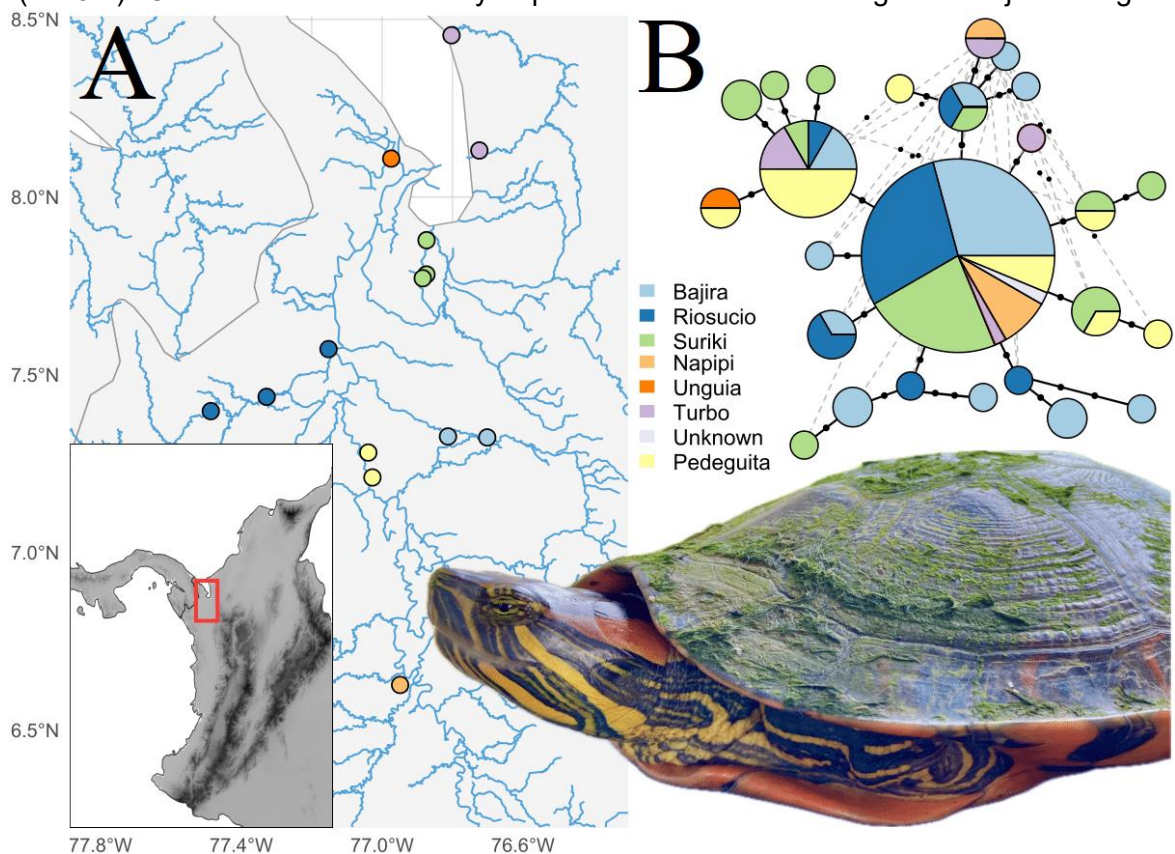


Figure 1. A) Geographic distribution of the sampling localities of *Trachemys medemi* in the Atrato River basin, northwestern Colombia. Colored circles indicate the geographic origin of the individuals included in the genetic analyses. B) Median-joining haplotype network of *Trachemys medemi* based on mtDNA control region sequences. In-set photo: adult male from Río Salaquí, Riosucio, Chocó (photo by Sebastián Cuadrado-Ríos).

package *adegenet* (Jombart and Ahmed 2011). For each locality we estimated the observed number of alleles per locus (A_{obs}), allelic richness (A_R , rarefied to the minimum sample size) using the function `allel.rich` from the *poppr* package (Kamvar et al. 2014). We also calculated observed (H_O) and expected heterozygosity (H_E), the number of private alleles per population (P_A) and their frequency range (P_{AF}). Shannon's (H) and Simpson's (λ) diversity indices were calculated for each population. To assess multilocus linkage and test for non-random association of alleles, we estimated the index of association (I_A) and the standardized index of association (\bar{r}_D) with 999 permutations in *poppr* (Kamvar et al. 2014).

Allele size-based measures of population structure were estimated in GENEPOP following Michalakis and Excoffier (1996). For each locality, we calculated the mean squared allele size difference within localities (MSD_{intra}) and between localities (MSD_{inter}), along with the allele size-based inbreeding coefficient (ρ_{IS}), an analogue to FIS. These statistics incorporate the evolutionary distance among alleles and are particularly informative when assessing recent divergence or mutation-driven differentiation. Values of ρ_{IS} near zero indicate panmixia, while positive or negative deviations suggest inbreeding or heterozygote excess, respectively.

Genetic structure among populations of *Trachemys medemi* was further assessed with an Analysis of Molecular Variance (AMOVA) in ARLEQUIN v3.5 (Excoffier and Lischer 2010), using the number of different alleles (F_{ST}) as the distance measure. The significance of variance components was evaluated with 1,000 permutations. A hierarchical AMOVA was also performed to partition variation within and among the seven sampled localities. The significance of the fixation index (F_{ST}) and the among-population variance component (V_a) was tested using 1,000 nonparametric permutations. In addition, an exact test of population differentiation was performed in ARLEQUIN based on genotype frequencies, using an MCMC algorithm with 100,000 steps and 10,000 dememorization steps ($\alpha = 0.05$).

Population structure was also evaluated with the Bayesian clustering algorithm in STRUCTURE v.2.3.4 (Pritchard et al. 2000). Analyses were run under the admixture model with correlated allele frequencies and without prior information on population origin. For each value of K (number of clusters), 10 independent runs were performed, each consisting of 100,000 burn-in iterations followed by 500,000 MCMC repetitions. K values from 1 to 9 were tested. The optimal number of clusters was inferred using StructureSelector (Li and

Liu 2018), which implements multiple criteria including the ΔK method (Evanno et al. 2005) and the MedMeaK, MedMedK, and MaxMeanK statistics, providing a conservative and robust assessment of K.

To explore genetic variation without prior assumptions, we conducted a Principal Component Analysis (PCA) using *adegenet* (Jombart and Ahmed 2011). Individuals were grouped by the previously defined localities, with 95% confidence ellipses superimposed. The first six principal components were retained to capture a substantial proportion of variance while avoiding overfitting.

Migration, Bottlenecks, and Effective Population Sizes

The effective number of migrants (N_m) among localities was estimated using the private allele method (Barton and Slatkin 1986) implemented in GENEPOP. This approach calculates N_m based on the frequency of alleles unique to a single population. To assess the sensitivity of this estimate to sample size we performed calculations using different average sample sizes ($N = 10, 25, \text{ and } 50$), including a correction for unequal sample sizes across localities. Inputs for these calculations were the mean private allele frequency and mean sample size. Subsequently, contemporary rates of gene flow between populations were estimated using a Bayesian framework in BAYESASS (Wilson and Rannala 2003). The analysis was run with default parameters for 10 million MCMC iterations, following a burn-in of 100,000 iterations.

Finally, we estimated contemporary effective population sizes (N_e) were estimated using the linkage disequilibrium method implemented in NeEstimator v.2.1 (Do et al. 2014), accessed via the R package *RLDNe* (Robinson et al., unpubl.). a wrapper for NeEstimator. Estimates were generated for populations with ≥ 5 individuals and for the entire dataset pooled as a single unit. We generated 95% confidence intervals using Jackknife resampling across loci. Negative or infinite N_e estimates were considered biologically uninformative and were excluded from final interpretation.

Because preliminary analyses indicated panmixia, we evaluated bottlenecks both for localities and at the species-wide level, treating all localities as a single population. First, we used BOTTLENECK (Piry et al., 1999) to test for heterozygosity excess relative to mutation–drift equilibrium expectations. Analyses were performed under three mutation models commonly applied to microsatellite data: the Infinite Allele Model (IAM), the

Stepwise Mutation Model (SMM), and the Two-Phase Model (TPM). We conducted One-tailed Wilcoxon signed-rank and sign tests with 1,000 iterations. A significant heterozygosity excess under TPM ($p \leq 0.05$) was considered evidence of a recent bottleneck. Secondly, we applied the M ratio test (Garza and Williamson 2001) which compares the number of alleles to the allelic range. As genetic drift preferentially removes rare alleles, the number of alleles (k) decreases more quickly than the allelic range (r); consequently, populations that have experienced a decline typically exhibit smaller M ratios (k/r) than stable populations. We computed M ratio using ARLEQUIN v3.5 (Excoffier and Lischer 2010). To determine the significance of the observed M ratio, we generated a threshold (CRITICAL_M) using coalescent Monte Carlo simulations in CRITICAL_M (Garza and Williamson 2001) under a Two-Phase Mutation model (TPM). We used the recommended parameters: a proportion of multi-step mutations (p_g) = 0.22 and a mean multi-step size (Δg) = 3.1 (Peery et al. 2012). Simulations were run for a range of mutation-drift parameters ($\theta = 4N_e\mu \in \{0.1, 0.5, 1, 5, 10\}$) to encompass plausible mutation-drift scenarios. The number of gene copies per locus in the simulations was set to the median of the observed non-missing diploid gene copies across loci. For each θ , we generated 5,000 replicates and defined the critical value as the 5th percentile of the resulting M distribution. A bottleneck was inferred if the observed M ratio was lower than the CRITICAL_M value (Hoban et al. 2013).

3.3 Results

Mitochondrial Control Region

A 92-bp-long duplicated sequence in the control region of *Trachemys medemi* reported by Cuadrado-Ríos et al. (2025) was absent in four individuals from our dataset ($n = 96$). These individuals were from Ciénaga de Napipí, Chocó (BTBC31207), Río Surikí (BTBC13181), Río Turbo (BTBC7056) and Río Truandó (BTBC13170). As no geographic pattern was evident for this deletion, the duplicate region was excluded from subsequent analyses.

Among the 94 correctly amplified sequences, we recovered 20 haplotypes. The haplotype network revealed a shallow, star-like structure, with most haplotypes separated by one or two mutational steps (Fig. 1B). This topology was consistent with recent demographic expansion. Several haplotypes were shared across localities, while others were private to single localities. Bayesian clustering analysis in *fastbaps* consistently identified a single

Table 1. Genetic diversity and demographic parameters of *Trachemys medemi* populations based on mitochondrial D-Loop sequences. N = number of individuals; S = number of segregating sites; h = number of haplotypes; Hd = haplotype diversity \pm standard deviation; K = average number of pairwise differences; π = nucleotide diversity; r = Harpending's raggedness index; R2 = Ramos-Onsins and Rozas statistic; τ = expansion parameter. Population Unguía was excluded from the estimations, since it was represented by only one individual.

Pop	mtDNA Control Region											nDNA Microsatellites										
	N	S	h	Hd \pm SD	K	π	r	R2	τ	Tajima's D	Fu's FS	N	Aob	AR	Ho	He	PA	PAF	H	λ	Ia	iD
Bajirá	27	13	9	0.607 \pm 0.12	1368	0.00139	0.0273	0.0623	0.869	-2.0251**	4.62086**	27	5,00	1,488	0,508	0,487	5	0.037-0.074	3.30	0.963	-0.0044	-0.0004
Riosucio	20	5	5	0.442 \pm 0.10	0.589	0.00066	0.1429	0.0822	0.589	-1.63814**	3.05269**	21	5,36	1,503	0,506	0,503	5	0.025-0.05	3.04	0.952	0.1712	0.0180
Surikí	24	8	-	-	1279	0.00143	0.0578	0.0740	1279	-1.42494**	5.90581**	22	5,55	1,514	0,474	0,514	12	0.3-0.09	3.09	0.955	0.5641	0.0581
Turbo	4	1	-	-	0.667	0.00075	0.5556	0.3333	0.667	0.0	0.0	5	3,73	1,534	0,467	0,537	0		1.61	0.800	0.2719	0.0337
Napipí	5	0	1	0.000 \pm 0.00	0.000	0.00000	-	-	-	163299	0.54002	6	2,18	1,538	0,455	0,454	2	0.045-0.048	1.79	0.833	-0.0198	-0.0022
Unguía	1	-	-	-	-	-	-	-	-	-	-	1	-	1,370	-	-	0		-	-	-	-
Pedeguita	14	6	7	0.802 \pm 0.09	1363	0.00138	0.0949	0.1001	1363	-0.99713	2.99979**	14	4,64	1,454	0,494	0,453	4	0.039-0.067	2.64	0.929	0.1199	0.0124
Total	95	18	20	0.654 \pm 0.08	1083	0.00121	0.0429	0.0318	1083	-0.63605	-229130	96	4,41	1,474	0,467	0,479	28	0.025-0.09	2.21	0.0932	0.228	0.02313

genetic cluster across all algorithms tested (*optimize.baps*, *baps*, and *optimize.symmetric*), indicating a lack of significant genetic subdivision supporting panmixia across the sampled range.

Haplotype diversity (H_d) ranged from 0.000 in Napipí, to 0.802 in Pedeguita, with an overall value of 0.654 for the total dataset. Nucleotide diversity (π) was generally low, varying between 0.00055 in Riosucio and 0.00153 in Pedeguita (Table 1). The mismatch distribution for the pooled dataset was unimodal with a low raggedness index ($r = 0.0429$) and a low R_2 (0.0318), further supporting a model of demographic expansion. At the locality scale, Bajirá and Surikí exhibited smooth mismatch distributions ($r < 0.06$), while Turbo showed a highly irregular distribution ($r = 0.5556$), a pattern likely attributable to its small sample size. Neutrality tests yielded significantly negative values of Tajima's D and Fu's F_S (Table 1).

Pairwise indirect estimates of gene flow (N_m) were very high (>4) between localities revealed that with a robust sampling size, indicating substantial contemporary gene flow. Key examples include Bajirá–Surikí ($N_m = 26.03$), Surikí–Pedeguita ($N_m = 16.54$), and Bajirá–Turbo ($N_m = 24.00$). The female effective population size (N_{ef}), derived from mitochondrial genetic diversity (Watterson's θ and Φ_{ST}), was estimated at approximately 7,988 individuals. Pairwise comparisons (Table 2) further revealed significant differentiation among several localities: Pedeguita exhibited strong divergence from Bajirá ($\Phi_{ST} = 0.153$, $P < 0.001$) and Riosucio ($\Phi_{ST} = 0.231$, $P < 0.001$), whereas Surikí showed moderate but significant differentiation with Bajirá ($\Phi_{ST} = 0.037$, $P < 0.01$) and Riosucio ($\Phi_{ST} = 0.045$, $P < 0.05$). In contrast, no significant differentiation was detected involving Napipí, Unguía, or Turbo, which clustered closely with other localities. The AMOVA analysis confirmed these patterns, attributing 91.65% of the total genetic variation to differences within populations (and 8.35% to differences among populations ($\Phi_{ST} = 0.084$, $P < 0.01$; Table 3).

Bayesian coalescent estimates of female-mediated migration ($4N_m$) revealed heterogeneous patterns of gene flow among (Supplementary Table C2). The highest levels migration rates were into Unguía, which received >7 migrants per generation from several source localities. Bajirá and Surikí showed moderate immigration rates ($4N_m \approx 1.5$ – 4.6). In contrast, Napipí and Pedeguita exhibited markedly lower immigration ($4N_m < 1$), indicating restricted female-mediated gene flow into some localities.

Microsatellite loci

After applying the Bonferroni correction for multiple tests (adjusted $\alpha = 0.05/11 \approx 0.0045$), no loci showed significant deviations from Hardy–Weinberg equilibrium (HWE) across localities (Supplementary Table C3). Prior to correction, only locus Tsc169 in Bajirá and locus Tsc328 in Bajirá yielded p-values below 0.05, but these were not significant after correction. Fisher’s combined probability test across all loci and populations indicated no global departure from equilibrium ($\chi^2 = 118.33$, $df = 124$, $p = 0.627$). Likewise, none of the pairwise tests for linkage disequilibrium were significant after Bonferroni correction (adjusted $\alpha = 0.05/55 \approx 0.00091$), indicating that alleles at different loci segregate independently.

Estimates of genetic diversity varied among the seven sampling localities (Table 1). The mean observed number of alleles per locus (A_{ob}) ranged from 2.18 (Napipí) to 5.55 (Surikí), with a global mean of 4.41. Allelic richness (A_R) was relatively consistent across populations, ranging from 1.454 (Pedeguita) to 1.538 (Napipí), with an overall value of 1.474. Observed heterozygosity (H_o) varied from 0.455 (Napipí) to 0.508 (Bajirá), while expected heterozygosity (H_E) ranged between 0.453 and 0.537. The number of private alleles (P_A) per population ranged from 0 (Necoclí and Unguía) to 12 (Surikí), with a total of 28 across all localities; private allele frequencies (PAF) spanned 0.025–0.09. Shannon’s diversity index (H) ranged from 1.61 (Necoclí) to 3.30 (Bajirá), and Simpson’s index (λ) ranged from 0.800 to 0.963. The index of association (I_A) was highest in Surikí (0.5641), followed by Necoclí (0.2719), while the standardized index of association (\bar{r}_D) reached its maximum also in Surikí (0.0581). The global \bar{r}_D was 0.0231, and the permutation test ($p = 0.1$) did not indicate significant multilocus linkage disequilibrium.

Table 2. Below the diagonal, pairwise genetic differentiation (Φ_{ST}) among sampling localities across the entire distribution of *Trachemys medemi*, based on mtDNA control region sequences; significant values are indicated as: $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***). Negative Φ_{ST} values were interpreted as zero, reflecting a lack of differentiation. Above the diagonal, indirect estimates of gene flow (N_m) based on the island-migration model. Values above four in bold.

	Bajirá	Riosucio	Surikí	Napipí	Unguía	Turbo	Pedeguita
Bajirá	–	-	26.03	-	1.04	24.00	5.54
Riosucio	-0.018	–	21.22	-	0.30	3.65	3.33
Surikí	0.037**	0.045*	–	-	1.40	-	16.54
Napipí	-0.088	-0.089	-0.038	–	0.0	1.40	5.37
Unguía	0.491	0.772	0.416	1.0	–	0.80	5.37
Turbo	0.040	0.215	-0.039	0.394	0.556	–	-
Pedeguita	0.153***	0.231**	0.057	0.158	0.157	-0.140	–

Null allele frequency estimates were low overall, with only a few loci exceeding the 0.05 threshold in specific localities (Supplementary Table C3). Loci Tsc252 and Tsc299 showed moderate frequencies (0.067–0.115) in some localities, but no locus exceeded the critical threshold of 0.20. The mean null allele frequency across loci and localities was 0.026. Global F_{ST} estimates from FreeNA were low, indicating minimal population differentiation (Supplementary Table C4). The uncorrected F_{ST} was 0.0036, while the ENA-corrected value (for null alleles) was slightly higher at 0.0050. Bootstrap 95% confidence intervals overlapped zero in both cases (–0.0036 to 0.0120 and –0.0023 to 0.0131, respectively). At the locus level, several markers (e.g., Tsc299, Tsc328, Tsc302, Tsc323) exhibited negative or near-zero values, whereas loci Tsc243 and Tsc252 showed moderate differentiation ($F_{ST} \approx 0.02$ –0.04). These results suggest that null alleles had a negligible influence on overall estimates of differentiation. Pairwise Cavalli–Sforza and Edwards chord distances (D_c) ranged from 0.2356 (Bajirá–Riosucio) to 0.4887 (Unguía–Napipí), with Unguía consistently showing the highest pairwise differentiation (Supplementary Fig. C1).

The AMOVA confirmed that nearly all genetic variation (99.94%) resided within populations, with only 0.06% partitioned among populations ($F_{ST} = 0.00063$, $P = 0.3519$; Table 3). These results indicate a near absence of genetic structuring across the distribution of *T. medemi*. Principal Component Analysis (PCA) further supported this conclusion, as Individuals from different localities largely overlapped in multivariate space, with no distinct clusters. Bayesian clustering analysis in STRUCTURE indicated that the most likely number of genetic clusters was $K = 2$ according to Puechmaille’s estimators, whereas Evanno

Table 3. Results of the analysis of molecular variance (AMOVA) based on microsatellite loci and the mitochondrial control region (D-Loop) in *Trachemys medemi*.

Source of variation	nDNA Microsatellite loci				mtDNA D-Loop			
	df	Sum of squares	Variance components	Percentage of variation	df	Sum of squares	Variance components	Percentage of variation
Among populations	6	15.726	0.00162	0.06	6	6.835	0.04867	8.35
Within populations	185	477.284	257.991	99.94	88	47.028	0.53441	91.65
Total	191	493.01	258.154	100.0	94	53.863	0.58308	100.0

method (ΔK) showed a secondary peak at $K = 1$. However, the bar plot for $K = 2$ revealed no biologically meaningful clustering, as all individuals and localities displayed broadly admixed genetic profiles (Fig. 2B). No evidence of recent bottlenecks was detected. Under the three mutation models evaluated (IAM, SMM, TPM), one-tailed Wilcoxon tests yielded no significant heterozygosity excess ($p > 0.80$). Consistently, the Garza–Williamson’s M ratio computed for the species-wide sample was 0.600. Coalescent simulations under the two-phased model (TPM) yielded critical values (CRITICAL_M) ranging from 0.368 to 0.444 across θ (Supplementary Table C5, Supplementary Fig. C2). In all cases, the observed M ratio was greater than the simulated critical value (one-tailed $p = 0.11$ – 0.26), providing no support for a population bottleneck.

Diversity indices and inbreeding coefficients, calculated following Weir and Cockerham (1984), are presented in Supplementary Table C6. Observed heterozygosity ($1 - Q_{intra}$) ranged from 0.4545 in Turbo to 0.6061 in Napipí, while expected heterozygosity ($1 - Q_{inter}$) ranged from 0.4864 in Turbo to 0.6106 in Suriki. F_{IS} values were generally low, from –

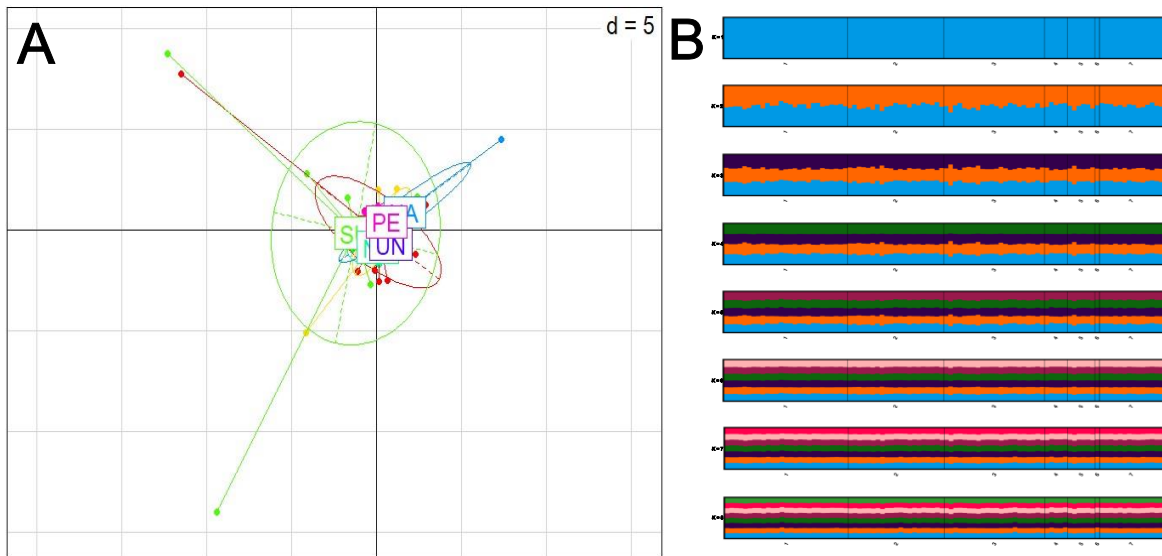


Figure 2. A) Principal Component Analysis (PCA) of *Trachemys medemi* individuals based on microsatellite genotypes, grouped by localities. Each dot represents an individual, colored according to its sampling locality. Ellipses represent 95% confidence intervals of multivariate dispersion for each population, while labels indicate the centroids of each group. BA corresponds to Bajirá (cyan), RI to Riosucio (red), SU to Sucre (green), NE to Necoclí (yellow), NA to Napipí (magenta), UN to Unguía (dark blue), and PE to Pedeguita (pink). This plot was constructed using the first five principal components ($d = 5$), which summarize most of the genetic variance (22.35%) in the dataset. B) Bayesian clustering analysis of *Trachemys medemi* populations based on microsatellite genotypes. Analyses were replicated across values of K from 1 to 9, with results summarized using StructureSelector. Puechmaille’s estimators identified $K = 2$ as the most likely partition, while the ΔK method of Evanno et al. (2005) peaked at $K = 1$. Despite these model-based signals, visual inspection of bar plots revealed widespread admixture across all populations, with no locality showing predominant assignment to a single cluster.

0.0256 in Napipí to 0.0654 in Necoclí, with most values close to zero or negative, suggesting generally high within-population diversity and an absence of inbreeding.

We estimated historical gene flow using the private alleles method in GENEPOP. Based on a mean sample size of 13.53 individuals and a mean private allele frequency of 0.036, N_m was estimated as 10.22, 3.81, and 2.42 for assumed average sample size of 10, 25, and 50 individuals, respectively. After correcting for unequal sample sizes, the adjusted estimate was 7.55 migrants per generation. These values are consistent with moderate historical gene flow. Contemporary migration rates estimated using BayesAss indicated negligible recent migration between the seven sampled localities. All pairwise migration rates were effectively zero, suggesting either very low levels of contemporary gene flow, complete reproductive isolation, or that the migration rates were below the detection limit of the method given our sample size and genetic resolution. This analysis also estimated inbreeding coefficients (F) for each locality; all values were 0.0000 with no associated variance, indicating no evidence of recent inbreeding or departures from Hardy-Weinberg expectations. These results align with the weak or absent genetic structure inferred from STRUCTURE and PCA analyses and are consistent with the panmictic population model. Estimates of contemporary effective population size (N_e) using the linkage disequilibrium method (LDNe) were characterized by high uncertainty. Most population-level estimates had infinitely wide confidence intervals. Surikí was the only locality with a finite confidence interval ($N_e = 75.2$; 95% CI: 45.6–188.6). When all individuals were pooled ($N = 96$), the species-wide N_e was estimated at 245.7 (95% CI: 28.4–1304.5). Localities with small sample sizes (e.g., Napipí, Turbo, Unguía; $N < 5$) were excluded from analysis, and their estimates were considered unreliable.

Table 4. Estimates of contemporary effective population size (N_e) for *Trachemys medemi* using the linkage disequilibrium method. Estimates were calculated for each population with sample size (N) ≥ 10 and for the entire pooled dataset. A minimum allele frequency (MAF) threshold of 0.05 was applied. N_e values are accompanied by 95% confidence intervals estimated via the jackknife method.

Locality	N	MAF	N_e	95% CI
Bajirá	27	0.05	173.4	63.7 – ∞
Riosucio	21	0.05	352.8	93.2 – ∞
Surikí	22	0.05	75.2	45.6 – 188.6
Pedeguita	14	0.05	161.2	24.7 – ∞
All pooled	96	0.05	245.7	28.4 – 1304.5

3.4 Discussion

The combined evidence from mitochondrial control region sequences and nuclear microsatellites reveals a consistent pattern of minimal genetic structuring among *Trachemys medemi* populations across the Atrato basin. For both marker types, most of the genetic variation is distributed within populations, with only a minor fraction attributable to differences among localities (Table 2). This pattern is consistent with a pattern of basin-wide connectivity and shallow population structure, as expected for an aquatic turtle species restricted to a single river network (Howeth et al. 2008). Although localities at the periphery of the species' distribution, such as Turbo, Napipí, and Unguía (Fig. 1A), were represented by relatively small sample sizes (Table 1), they exhibited a similar pattern of weak genetic differentiation when compared to well-sampled populations like Riosucio or Bajirá. This consistency across populations with contrasting sampling efforts indicates that the observed genetic homogeneity in *Trachemys medemi* is robust and not an artifact of uneven sampling.

The pattern of panmixia aligns with results reported for *Trachemys venusta callirostris* in the Colombian Caribbean region by Balcero-Deaquiz (2022), a species distributed across a considerably broader geographic range than that of *T. medemi*. The concordance between these studies suggests that extensive riverine networks and floodplain wetlands facilitate gene flow in *Trachemys* species, buffering the effects of geographic distance and maintaining genetic homogeneity even over large spatial scales. For *T. medemi*, this panmixia is likely enabled by the complex network of rivers, streams, and floodplain wetland within the Atrato River basin, which provides extensive opportunities for dispersal and connectivity (Anaya-Acevedo et al. 2017, Palomino-Ángel et al. 2019). The basin's dense dendritic hydrographic system, with its multiple interconnected channels, seasonal floodplains, and permanent wetlands, reduce effective barriers to movement for aquatic organisms. Such hydrological complexity promotes both maternal and biparental gene flow, counteracting the effects of geographic distance and promoting genetic homogeneity across populations. Moreover, the dynamic flooding regime of the Atrato may periodically redistribute individuals across habitats, further maintaining high connectivity and inhibiting the development of genetic structure (Palomino-Ángel et al. 2019).

Human-mediated translocations may also contribute to the observed genetic homogeneity observed in *Trachemys medemi*. Freshwater turtles are frequently transported by local communities for consumption or trade (Morales-Betancourt et al. 2015, Páez et al. 2016), and such movements can inadvertently facilitate gene flow among otherwise distant or semi-isolated populations (González-Porter et al. 2011, Converse et al. 2017). The Atrato Slider was the most frequently seized turtle species in the region between 2005 and 2010 (Asprilla-Perea et al. 2013). Moreover, this phenomenon might have occurred for millennia. Zooarchaeological evidence has documented remains of *T. callirostris* at more than a dozen archaeological sites across the Caribbean region, with radiocarbon dates extending back to nearly 6,000 years before present (Mendoza Roldán 2024). For *T. medemi*, such anthropogenic translocations could reinforce natural dispersal processes by introducing individuals into new localities, thereby enhancing genetic admixture and obscuring subtle patterns of population subdivision.

The observed levels of nuclear genetic diversity in *Trachemys medemi* were low, with allelic richness (AR) ranging from 1.45 to 1.54 and expected heterozygosity (H_e) between 0.45 to 0.54 (mean 0.48; Table 1). These values are consistent with those reported for other Neotropical freshwater turtles facing anthropogenic pressures, including *T. hartwegi* (Becerra et al. 2025), *T. taylori* (McGaugh 2012), *T. venusta venusta* (Recino-Reyes et al. 2020) and *T. venusta callirostris* (Balcerio-Deaquiz 2022). Similar levels of H_e have also been documented in other imperiled turtle species, such as *Graptemys geographica* (Bennett et al. 2010), *G. caglei* (Ward et al. 2013), *Emys orbicularis* (Velo-Antón et al. 2008) and *Astrochelys radiata* (Rioux Paquette et al. 2009). In contrast, some species like *Podocnemis expansa* in the Brazilian Amazon maintain higher genetic variation despite recent population decline and strong population fragmentation from hydroelectric development and overharvesting (Pearse et al. 2006). Although cross-study comparisons require caution due to differences in marker selection and sampling design the collective evidence indicates that *T. medemi* harbors comparatively low nuclear genetic diversity. This reduced variation highlights the species' potential vulnerability to demographic stochasticity, environmental change and the effects of genetic drift. The narrow endemism of *T. medemi* within the Atrato basin, coupled with intense anthropogenic pressures on its wetland habitats, underscore the critical importance of conserving its remaining genetic variation to ensure long-term evolutionary potential.

When all sampled individuals were pooled, the overall effective population size (N_e) of *Trachemys medemi* was estimated at 245.7, (95% CI: 28.4 - 1304.5). According to the conservation genetic thresholds outlined by Frankham et al. (2014), this value is well above the short-term minimum ($N_e = 50$) required to minimize the risk of inbreeding depression but remains below the long-term threshold ($N_e \geq 500$) considered necessary to retain evolutionary potential. This intermediate N_e suggests that while *T. medemi* is unlikely to suffer from immediate inbreeding effects, it may be vulnerable to the loss of adaptive variation through genetic drift over longer timescales (Frankham et al. 2014, Femerling et al. 2023, Aitken et al. 2024). A comparable pattern is observed in other freshwater turtles, such as *T. venusta callirostris* in the Magdalena River valley ($N_e = 251.8$, CI: 133.3- 1287.7; Balceró Deaquiz 2022), and *T. harwegi* in the Nazas River ($N_e = 307$, CI: 139- 8352.3; Becerra et al. 2025). These consistent findings reinforce the conclusion that *T. medemi*, while not currently at risk of inbreeding, persists at an effective size that is likely insufficient to ensure long-term evolutionary resilience in the face of deterministic and stochastic changes.

Evolutionary history of *Trachemys medemi*

Beyond contemporary processes, our results also provide insights into the evolutionary history of *T. medemi*. While both mitochondrial and microsatellite markers consistently reveal an overall pattern of weak population structure, analyses of the mtDNA control region suggest a species-wide demographic expansion. This signature, evidenced by the star-like haplotype network (Fig. 1B), significant negative neutrality tests, and unimodal mismatch distribution analyses (Table 1), is consistent with a relatively recent origin of mitochondrial diversity following historical bottlenecks or colonization events. A post-Last Glacial Maximum (Rohling et al. 2017) demographic expansion may account for these signatures, consistent with the idea that the Atrato bay and lower basin likely formed a broadly connected freshwater system during the LGM (Villegas et al. 2018). We proposed that rapid postglacial sea-level rise (Hewitt, 2000) following the LGM, subsequently reconnected aquatic corridors and floodplain systems, facilitating extensive dispersal and gene flow across the Atrato Basin. This post-LGM connectivity may have been so profound that it erased the genetic footprints of habitat fragmentation in the nuclear genome, while the mitochondrial genome retained the signature of the preceding demographic expansion. An alternative scenario is that local extinction events occurred during the LGM, followed by rapid recolonization from surviving populations; a process that could also explain the

observed patterns of diversity (Reid et al. 2019). Similar dynamics of postglacial expansion from a single refugium, erasing signals of past isolation, have been reported in the painted turtle *Chrysemys picta* (Reid et al. 2017). Distinguishing between these hypotheses will require genomic-scale data, which could provide the resolution needed to test whether historical bottlenecks, recolonization events, or a combination of both shaped the evolutionary history of *T. medemi*. Ultimately, these findings emphasize that the species' current genetic patterns cannot be fully understood without considering the legacy of Quaternary climatic oscillations, which likely structured—and subsequently reconnected—the aquatic landscapes of northwestern Colombia.

Conservation implications

Populations of *Trachemys medemi* face a high risk of extinction across the Atrato River Basin. Nuclear estimates of effective population size (N_e) reinforce this assessment. While the population in Riosucio approached the recommended threshold for long-term viability (Frankham et al. 2014), most populations had values well below the critical benchmark of $N_e = 500$. The species-wide estimate also yielded an intermediate value, which is likely insufficient to ensure long-term evolutionary potential. Our findings of shallow genetic structuring across the species' distribution suggest that *T. medemi* can be managed as a single management unit. This genetic homogeneity underscores the importance of maintaining connectivity. Previous studies on other freshwater turtles have demonstrated that maintaining large and connected populations through habitat conservation is critical for preventing extinction processes. Populations in fragmented habitats exhibit reduced genetic diversity and lower effective sizes, which can accelerate local extinctions (Gallego-García et al. 2018). Similarly, studies on other Neotropical turtles have highlighted that protecting nesting beaches and aquatic habitats is essential to prevent demographic bottlenecks and preserve adaptive capacity (Pearse et al. 2006, Escalona et al. 2009, Ferreira-Junior y Castro 2010). These findings are directly relevant for *T. medemi*, a narrow endemic restricted to northwestern Colombia, where anthropogenic pressures such as wetland conversion, overharvesting, and pollution may disproportionately affect small and isolated populations.

Beyond its genetic vulnerabilities, *T. medemi* multiplies threatened by multiple anthropogenic pressures across the Atrato basin. Habitat degradation from wetland drainage, agricultural expansion, and mining activities directly reduce available aquatic

environments and alters water quality (Hurtado-Gómez et al. 2015, Calzadilla 2019, Caicedo-Rivas et al. 2022, Bernal-Alviz et al. 2025). Overharvesting for local consumption and trade further exacerbates demographic pressures (Epperson and Heise 2003, Moll and Moll 2004, Rhodin et al. 2011, TCC 2011, Patino and Estupinan-Suarez 2016, Stanford et al. 2020). Furthermore, climate change is expected to alter rainfall regimes in the Chocó region, increasing the frequency of extreme floods and droughts that could destabilize the wetland systems upon which *T. medemi* depends. These cumulative pressures highlight the urgent need to integrate genetic insights with habitat-based management to mitigate the synergistic effects of demographic decline, habitat loss, and environmental change.

Given these threats, we recommend that conservation strategies prioritize the following actions: (i) preserving continuous aquatic corridors to maintain natural dispersal and gene flow, (ii) protecting critical habitats such as wetlands and nesting areas, and (iii) implementing long-term monitoring of effective population sizes to detect demographic instability before irreversible genetic erosion occurs. Specific populations such as Necoclí and Unguía require further studied, as they were underrepresented in this study. In line with global recommendations for freshwater turtles (Rhodin et al. 2018, Lovich et al. 2018), the survival of *T. medemi* depends on the proactive protection of its aquatic habitats to prevent further demographic and range reductions. By prioritizing the protection and restoration of natural habitats, conservation efforts can secure the persistence of large, genetically diverse, and resilient populations of *T. medemi*. Such actions will not only safeguard the evolutionary potential of this endemic species but also ensure the ecological integrity of the Chocó's wetland ecosystems, which sustain a vast array of biodiversity and essential ecological processes.

Conclusions

This study provides the first comprehensive genetic assessment of *Trachemys medemi*, an endemic freshwater turtle of the Atrato River basin. Our analyses of mitochondrial and nuclear markers revealed shallow population structure and signs of demographic expansion. These patterns are consistent with historical connectivity facilitated by the basin's complex hydrography, a process that may have been reinforced by human-mediated translocations. Despite this high connectivity, levels of nuclear genetic diversity were low, and effective population sizes (N_e) remained below the threshold considered necessary for long-term adaptive potential. These results indicate that while *T. medemi* is

not at immediate risk of inbreeding depression, it is vulnerable to the long-term effects of genetic drift and loss of evolutionary resilience. When coupled with intense anthropogenic pressures, including wetland degradation, overharvesting, mining activities, and the projected impacts of climate change, these findings highlight an urgent need for targeted conservation action. Protecting and restoring wetland habitats, maintaining aquatic connectivity, and implementing long-term monitoring of effective population sizes are critical to ensuring the persistence of genetically diverse and resilient populations of *T. medemi* and safeguarding the integrity of the wetland ecosystems of the Chocó region.

Acknowledgements

Funding and support were provided by the Turtle Conservation Fund (TCF). SC-R is supported by the “Convocatoria del Fondo de Ciencia, Tecnología e Innovación del Sistema General de Regalías, en el marco del Programa de Becas de Excelencia Doctoral del Bicentenario – Corte II”, Department of Chocó, Universidad Nacional de Colombia and Ministerio de Ciencia, Tecnología e Innovación de Colombia and by a “Research Grant (bi-nationally supervised Doctoral Degree/Cotutelle, 2023/24)” from the German Academic Exchange Service (DAAD). Turtles were manipulated under University research permits granted by the Ministry of Environment and Sustainable Development of Colombia. During manipulation, we followed the "Guidelines for live amphibians and reptiles in field and laboratory research", developed by the American Society of Ichthyologists and Herpetologists, 2004.

3.5 References

- Aguirre-C, J., and Rangel-Ch, O. (2008). *El Chocó biogeográfico*. In: *Colombia diversidad biótica VI: Riqueza y diversidad de los musgos y líquenes en Colombia*, 77-84.
- Aitken, S. N., Jordan, R., and Tumas, H. R. (2024). Conserving evolutionary potential: Combining landscape genomics with established methods to inform plant conservation. *Annual Review of Plant Biology*, 75(1), 707-736.
- Anaya-Acevedo, J. A., Escobar-Martínez, J. F., Massone, H., Booman, G., Quiroz-Londoño, O. M., Cañón-Barriga, C. C., ... and Palomino-Ángel, S. (2017). Identification of wetland areas in the context of agricultural development using Remote Sensing and GIS. *Dyna*, 84(201), 186-194.

- Arroyave-Bermúdez, F. J., Romero-Goyeneche, O. Y., Bonilla-Gómez, M. A., Hurtado-Heredia, R.G. (2015) Tráfico ilegal de tortugas continentales (Testudinata) en Colombia: una aproximación desde el análisis de redes. *Acta Biológica Colombiana*, 19(3), 381-392.
- Asprilla-Perea, J., Serna-Agudelo, J. E., and Palacios-Asprilla, Y. (2013). Diagnóstico sobre el decomiso de fauna silvestre en el departamento del Chocó (Pacífico Norte colombiano). *Revista UDCA Actualidad and Divulgación Científica*, 16(1), 175-184.
- Avila-Cervantes, J., and Larsson, H. C. (2023). Ice Age effects on genetic divergence of the American crocodile (*Crocodylus acutus*) in Panama: reconstructing limits of gene flow and environmental ranges: a reply to O’Dea et al. *Evolution*, 77(1), 329-334.
- Balcerro-Deaquiz, M. C. (2022). *Genética poblacional de la tortuga icotea colombiana Trachemys venusta callirostris (Testudines: Emydidae), proponiendo estrategias para su conservación* [Master’s thesis]. Universidad Nacional de Colombia. Repositorio Universidad Nacional. <https://repositorio.unal.edu.co/handle/unal/84508>
- Barton, N. H., and Slatkin, M. (1986). A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. *Heredity*, 56(3), 409-415.
- Becerra, E., Rodríguez López, B., Borja, M., and Rico, Y. (2025). Conservation genetics of a freshwater turtle (*Trachemys hartwegi*) in a threatened riverine ecosystem. *Molecular Biology Reports*, 52(1), 761.
- Beerli, P., and Felsenstein, J. (2001). Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of sciences*, 98(8), 4563-4568.
- Bennett, A. M., Keevil, M., and Litzgus, J. D. (2010). Spatial ecology and population genetics of northern map turtles (*Graptemys geographica*) in fragmented and continuous habitats in Canada. *Chelonian Conservation and Biology*, 9(2), 185-195.
- Bernal-Alviz, J., Córdoba-Tovar, L., Pastrana-Durango, D., Molina-Polo, C., Buelvas-Soto, J., Cruz-Esquivel, Á., ... and Díez, S. (2025). Influence of environmental and biological factors on mercury accumulation in fish from the Atrato River Basin, Colombia. *Environmental Pollution*, 366, 125345.

- Caicedo-Rivas, G., Salas-Moreno, M., and Marrugo-Negrete, J. (2022). Health risk assessment for human exposure to heavy metals via food consumption in inhabitants of middle basin of the Atrato river in the Colombian Pacific. *International journal of environmental research and public health*, 20(1), 435.
- Calzadilla, P. V. (2019). A Paradigm Shift in Courts' View on Nature: The Atrato River and Amazon Basin Cases in Colombia. *Law, Environment and Development Journal*, 15, 12.
- Cano, A., Manrique, H. F., Hoyos-Gomez, S. E., Echavarria, N., Upedui, A., Gonzalez, M. F., et al. (2017). Palms of the Darién Gap (Colombia-Panama). *Palms* 61, 5–20.
- Chapuis, M. P., and Estoup, A. (2007). Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, 24(3), 621-631.
- Chassin-Noria, O., Abreu-Grobois, A., Dutton, P. H., and Oyama, K. (2004). Conservation genetics of the east Pacific green turtle (*Chelonia mydas*) in Michoacan, Mexico. *Genetica*, 121(2), 195-206.
- Christenhusz, M. J., Fay, M. F., and Chase, M. W. (2017). *Plants of the world: an illustrated encyclopedia of vascular plants*. London: Kew Publishing, 792 p. doi: 10.7208/chicago/9780226536705.001.0001
- Converse, P. E., Kuchta, S. R., Hauswaldt, J. S., and Roosenburg, W. M. (2017). Turtle soup, Prohibition, and the population genetic structure of Diamondback Terrapins (*Malaclemys terrapin*). *PLoS One*, 12(8), e0181898.
- Cuadrado-Ríos, S., Vargas-Ramírez, M., Kehlmaier, C., and Fritz, U. (2025). Complete mitochondrial genome and phylogenetic analysis of the Atrato slider, *Trachemys medemi* (Testudines, Emydidae). *ZooKeys*, 1224, 253.
- Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J. and Ovenden, J. R. (2014). NeEstimator V2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*. 14, 209-214.
- Epperson, D. M., and Heise, C. D. (2003). Nesting and hatchling ecology of gopher tortoises (*Gopherus polyphemus*) in southern Mississippi. *Journal of Herpetology*, 37(2), 315-324.

- Escalona, T., Valenzuela, N., and Adams, D. C. (2009). Nesting ecology in the freshwater turtle *Podocnemis unifilis*: spatiotemporal patterns and inferred explanations. *Functional Ecology*, 826-835.
- Excoffier, L., and Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology resources*, 10(3), 564-567.
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14(8), 2611-2620.
- Femerling, G., Van Oosterhout, C., Feng, S., Bristol, R. M., Zhang, G., Groombridge, J., ... and Morales, H. E. (2023). Genetic load and adaptive potential of a recovered avian species that narrowly avoided extinction. *Molecular Biology and Evolution*, 40(12), 256.
- Ferreira Júnior, P. D., and Castro, P. D. T. (2010). Nesting ecology of *Podocnemis expansa* (Schweigger, 1812) and *Podocnemis unifilis* (Troschel, 1848)(Testudines, Podocnemididae) in the Javaés River, Brazil. *Brazilian Journal of Biology*, 70, 85-94.
- Frankham, R., Bradshaw, C. J., and Brook, B. W. (2014). Genetics in conservation management: revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation*, 170, 56-63.
- Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147(2), 915-925.
- Gallego-García, N., Vargas-Ramírez, M., Forero-Medina, G., and Caballero, S. (2018). Genetic evidence of fragmented populations and inbreeding in the Colombian endemic Dahl's toad-headed turtle (*Mesoclemmys dahli*). *Conservation Genetics*, 19(1), 221-233.
- Garza, J. C., and Williamson, E. G. (2001). Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, 10(2), 305-318.
- González-Porter, G. P., Hailer, F., Flores-Villela, O., García-Anleu, R., and Maldonado, J. E. (2011). Patterns of genetic diversity in the critically endangered Central American river turtle: human influence since the Mayan age?. *Conservation Genetics*, 12(5), 1229-1242.

- Guo, S. W., and Thompson, E. A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, 361-372.
- Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405(6789), 907-913.
- Hoban, S. M., Mezzavilla, M., Gaggiotti, O. E., Benazzo, A., van Oosterhout, C., and Bertorelle, G. (2013). High variance in reproductive success generates a false signature of a genetic bottleneck in populations of constant size: a simulation study. *BMC bioinformatics*, 14(1), 309.
- Howeth, J. G., McGaugh, S. E., and Hendrickson, D. A. (2008). Contrasting demographic and genetic estimates of dispersal in the endangered Coahuilan box turtle: a contemporary approach to conservation. *Molecular Ecology*, 17(19), 4209-4221.
- Jombart, T., and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3070-3071.
- Kamvar, Z. N., Tabima, J. F., and Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281.
- Komoroske, L. M., Jensen, M. P., Stewart, K. R., Shamblin, B. M., and Dutton, P. H. (2017). Advances in the application of genetics in marine turtle biology and conservation. *Frontiers in Marine Science*, 4, 156.
- Kuo, C. H., and Janzen, F. J. (2004). Genetic effects of a persistent bottleneck on a natural population of ornate box turtles (*Terrapene ornata*). *Conservation Genetics*, 5(4), 425-437.
- Li, Y. L., and Liu, J. X. (2018). StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology resources*, 18(1), 176-177.
- Lovich, J. E., Ennen, J. R., Agha, M., and Gibbons, J. W. (2018). Where have all the turtles gone, and why does it matter?. *BioScience*, 68(10), 771-781.
- McGaugh, S. E. (2012). Comparative population genetics of aquatic turtles in the desert. *Conservation Genetics*, 13(6), 1561-1576.

Mendoza Roldán J.S. (2024) Tortugas continentales en el registro arqueológico y etnozoológico del Caribe Colombiano: una reflexión sobre el uso sustentable como recurso alimenticio. *Jangwa Pana* 23. <https://doi.org/10.21676/issn.1657-4923>

Michalakis, Y., and Excoffier, L. (1996). A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics*, 142(3), 1061-1064.

Mittermeier, R. A., Turner, W. R., Larsen, F. W., Brooks, T. M., and Gascon, C., (2011). "Biodiversity hotspots," in *Global biodiversity conservation: the critical role of hotspots*. Eds. F. E. Zachos and J. C. Habel (London: Springer-Verlag), 3–22. doi: 10.1007/978-3-642-20992-5_1

Moll, D. and Moll EO. (2004) *The Ecology, Exploitation, and Conservation of River Turtles*. Oxford: Oxford University Press, 393 p.

Morales-Betancourt, M.A., Lasso, C.A., Páez, V.P. and Bock, B.C. (2015) *Libro rojo de los reptiles de colombia (2015)*. Instituto de Investigación de recursos biológicos Alexander von Humboldt (IAvH), Universidad de Antioquia. Bogotá, D.C., Colombia, 258 p.

Páez, V. P., Morales-Betancourt, M. A., Lasso, C. A., Castaño Mora, O. V., and Bock, B. C. (2016). *Biología y conservación de las tortugas continentales de Colombia. Serie Recursos Hidrobiológicos y Pesqueros Continentales de Colombia*. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Bogotá D.C., 528 p.

Paez, V. P., Bock, B. C., Alzate-Estrada, D. A., Barrientos-Munoz, K. G., Cartagena-Otalvaro, V. M., Echeverry-Alcendra, A., ... and Vallejo-Betancur, M. M. (2022). Turtles of Colombia: an annotated analysis of their diversity, distribution, and conservation status. *Amphibian and Reptile Conservation*, 16(1), 106-135.

Palomino-Ángel, S., Anaya-Acevedo, J. A., Simard, M., Liao, T. H., and Jaramillo, F. (2019). Analysis of floodplain dynamics in the Atrato River Colombia using SAR interferometry. *Water*, 11(5), 875.

Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics*, 26(3), 419-420.

- Patino, J. E., and Estupinan-Suarez, L. M. (2016). Hotspots of wetland area loss in Colombia. *Wetlands*, 36(5), 935-943.
- Pearse, D. E., Arndt, A. D., Valenzuela, N., Miller, B. A., Cantarelli, V., and Sites Jr, J. W. (2006). Estimating population structure under nonequilibrium conditions in a conservation context: continent-wide population genetics of the giant Amazon river turtle, *Podocnemis expansa* (Chelonia; Podocnemididae). *Molecular Ecology*, 15(4), 985-1006.
- Peery, M. Z., Kirby, R., Reid, B. N., Stoelting, R., Doucet-Bëer, E. L. E. N. A., Robinson, S., ... and Palsbøll, P. J. (2012). Reliability of genetic bottleneck tests for detecting recent population declines. *Molecular Ecology*, 21(14), 3403-3418.
- Piry, S., G. Luikart, and J. Cornuet. (1999). BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, 90, 502–503.
- Poveda, I. C., Rojas, C. A., Rudas, A., and Rangel, O. (2004). El Chocó biogeográfico: ambiente físico. *Colombia diversidad biótica IV, El Chocó biogeográfico/Costa Pacífica*, 1-22.
- Pritchard, J. K., Stephens, M., Rosenberg, N. A., and Donnelly, P. (2000). Association mapping in structured populations. *The American Journal of Human Genetics*, 67(1), 170-181.
- Raymond, M. and Rousset, F. (1995). Genepop (version 1.2), population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86, 248–249.
- Recino-Reyes, E. B., Leshner-Gordillo, J. M., Machkour-M'Rabet, S., Gallardo-Alvarez, M. I., Zenteno-Ruiz, C. E., Olivera-Gómez, L. D., ... and Martínez, R. H. (2020). Conservation and Management of *Trachemys venusta venusta* in Southern Mexico: A Genetic Approach. *Tropical Conservation Science*, 13, 1940082920961506.
- Reid, B. N., Kass, J. M., Wollney, S., Jensen, E. L., Russello, M. A., Viola, E. M., ... and Naro-Maciel, E. (2019). Disentangling the genetic effects of refugial isolation and range expansion in a trans-continentally distributed species. *Heredity*, 122(4), 441-457.

Reid, B. N., Naro-Maciel, E., Hahn, A. T., FitzSimmons, N. N., and Gehara, M. (2019). Geography best explains global patterns of genetic diversity and postglacial co-expansion in marine turtles. *Molecular Ecology*, 28(14), 3358-3370.

Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution* 43: 223–225.

Rioux Paquette, S., Ferguson, B. H., Lapointe, F. J., and Louis Jr, E. E. (2009). Conservation genetics of the radiated tortoise (*Astrochelys radiata*) population from Andohahela National Park, southeast Madagascar, with a discussion on the conservation of this declining species. *Chelonian Conservation and Biology*, 8(1), 84-93.

Rhodin, A. G., Stanford, C. B., Van Dijk, P. P., Eisemberg, C., Luiselli, L., Mittermeier, R. A., ... and Vogt, R. C. (2018). Global conservation status of turtles and tortoises (order Testudines). *Chelonian Conservation and Biology*, 17(2), 135-161.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., and Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, 34(12), 3299-3302.

Simison, W. B., Sellas, A. B., Feldheim, K. A., and Parham, J. F. (2013). Isolation and characterization of microsatellite markers for identifying hybridization and genetic pollution associated with red-eared slider turtles (*Trachemys scripta elegans*). *Conservation genetics resources*, 5(4), 1139-1140.

Stanford, C. B., Iverson, J. B., Rhodin, A. G., van Dijk, P. P., Mittermeier, R. A., Kuchling, G., ... and Walde, A. D. (2020). Turtles and tortoises are in trouble. *Current Biology*, 30(12), R721-R735.

Standfuss, B., Lipovšek, G., Fritz, U., and Vamberger, M. (2016). Threat or fiction: is the pond slider (*Trachemys scripta*) really invasive in Central Europe? A case study from Slovenia. *Conservation Genetics*, 17(3), 557-563.

Tajima, F. (1989). The effect of change in population size on DNA polymorphism. *Genetics*, 123(3), 597-601.

TCC [Turtle Conservation Coalition], Rhodin, A.G.J., Walde, A.D., Horne, B.D., van Dijk, P.P., Blanck, T. and Hudson, R. (2011) *Turtles in Trouble: The World's 25+ Most Endangered Tortoises and Freshwater Turtles—2011*. Lunenburg, MA: IUCN/SSC Tortoise

and Freshwater Turtle Specialist Group, Turtle Conservation Fund, Turtle Survival Alliance, Turtle Conservancy, Chelonian Research Foundation, Conservation International, Wildlife Conservation Society, and San Diego Zoo Global, p. 54

Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S. D., and Corander, J. (2019). Fast hierarchical Bayesian analysis of population structure. *Nucleic Acids Research*, 47(11), 5539-5549.

TTWG [Turtle Taxonomy Working Group: Rhodin, A. G. J., Iverson, J. B., Fritz, U., Gallego-García, N., Georges, A., Shaffer, H. B., and van Dijk, P. P.] (2025). *Turtles of the World: Annotated Checklist and Atlas of Taxonomy, Synonymy, Distribution, and Conservation Status (10th Ed.)*. Chelonian Research Monographs, 10, 1–575.

Vargas-Ramírez, M., Stuckas, H., Castano-Mora, O. V., and Fritz, U. (2012). Extremely low genetic diversity and weak population differentiation in the endangered Colombian river turtle *Podocnemis lewyana* (Testudines: Podocnemididae). *Conservation Genetics*, 13(1), 65-77.

Vargas-Ramírez, M., del Valle, C., Ceballos, C. P., and Fritz, U. (2017). *Trachemys medemi* n. sp. from northwestern Colombia turns the biogeography of South American slider turtles upside down. *Journal of Zoological Systematics and Evolutionary Research*, 55(4), 326-339.

Velo-Antón, G., García-París, M., and Cordero Rivera, A. (2008). Patterns of nuclear and mitochondrial DNA variation in Iberian populations of *Emys orbicularis* (Emydidae): conservation implications. *Conservation Genetics*, 9(5), 1263-1274.

Villegas, P., Paredes, V., Betancur, T., Taupin, J. D., & Toro, L. E. (2018). Groundwater evolution and mean water age inferred from hydrochemical and isotopic tracers in a tropical confined aquifer. *Hydrological Processes*, 32(14), 2158-2175.

Ward, R., Babitzke, J. B., and Killebrew, F. C. (2013). Genetic Population Structure of Cagle's Map Turtle (*Graptemys caglei*) in the Guadalupe and San Marcos Rivers of Texas—A Landscape Perspective. *Ichthyology and Herpetology*, 2013(4), 723-728.

Wilson, G. A., and Rannala, B. (2003). Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, 163(3), 1177-1191.

4. Capítulo

Phylogenomics and divergence times of *Rhinoclemmys* turtles support multiple Miocene invasions of South America

Capítulo en preparación para ser sometido a *Ecology and Evolution*.

Abstract

The genus *Rhinoclemmys* Fitzinger, 1835 is the only New World lineage of the otherwise predominantly Old World turtle family Geoemydidae, making it a key model for investigating Neotropical biogeography and diversification. Distributed from Mexico to northern South America, *Rhinoclemmys* occupies diverse habitats and its evolutionary history has been linked to repeated dispersal events across Central and South America, particularly during the Miocene. However, phylogenetic relationships within the genus remain incompletely resolved, and evidence of mito–nuclear discordance complicates interpretations of lineage diversification. Here, we present complete mitochondrial genomes and a robust dataset of nuclear BUSCO genes from multiple *Rhinoclemmys* species, including representatives of both eastern and western mitochondrial lineages of *R. melanosterna*, to clarify phylogenetic relationships, estimate divergence times, and evaluate the timing of major dispersal events. Our analyses confirm the monophyly of *Rhinoclemmys*, with *R. nasuta* as a long-independent South American lineage and the first colonization event of South America, well before previously estimated. Both molecular dating frameworks support a scenario of multiple Miocene invasions into South America, predating the final closure of the Panamanian Isthmus. Complete mitogenomes reveal clear mito–nuclear discordance within the *R. melanosterna* complex occurred during the Miocene, consistent with at least two independent cases of mitochondrial capture involving *R. funerea* and *R. diademata*. By integrating genomic datasets with fossil-calibrated molecular dating, this study refines the evolutionary framework of *Rhinoclemmys* and demonstrates how geological processes, repeated dispersals, and potential hybridization shaped its diversification. These findings

highlight the evolutionary significance of *Rhinoclemmys* lineages and provide a robust phylogenetic foundation for conservation planning in biodiversity hotspots such as Mesoamerica and the Chocó.

Key words: *Rhinoclemmys*, Phylogenomics, Mitogenomics, BUSCO, Beast, MCMCTree.

4.1 Introduction

The genus *Rhinoclemmys* Fitzinger, 1835 is the only New World extant lineage within the otherwise predominantly Old-World turtle family Geoemydidae, which has substantial taxonomic and ecological breadth across Asia, the Palearctic, and Africa (Spinks et al. 2004, Le and McCord 2008, De La Fuente et al. 2014). Extant *Rhinoclemmys* species range from northwestern Mexico through Central America to northern South America, inhabiting multiple biomes and biogeographic provinces in the Neotropics (Le and McCord 2008, De La Fuente et al 2014). This unique distribution makes *Rhinoclemmys* a compelling model for the study of how colonizations and subsequent regional diversification shaped lineage history. As the sole extant geoemydid radiation in the Americas, the study of *Rhinoclemmys* diversification is key for understanding the tempo and mode of turtle evolution and for testing macroevolutionary hypotheses regarding range expansion, niche breadth, and trait evolution (Spinks et al. 2004, Vargas-Ramírez et al. 2013).

The evolutionary history of *Rhinoclemmys* is consistently linked to an intercontinental dispersal from Asia into North America (Le and McCord 2008), followed by radiation events across Mesoamerica and tropical South America (Ernst 1978, Hirayama 1984, Vargas-Ramírez et al. 2013). Earlier hypotheses proposed a scenario of Molecular and paleontological evidence supports a scenario of early entry into North America via Beringian region, with migration across the Bering Strait during the early Eocene, with subsequent dispersal and vicariant fragmentation throughout the Middle to Late Cenozoic (Le and McCord 2008, Spinks et al. 2004, Claude et al. 2012, De La Fuente et al. 2014, Ascarrunz et al. 2021). Within the Neotropics, the radiation of *Rhinoclemmys* broadly aligns with major geological events, including the uplift of the Sierra Madre mountains in Mexico, the formation of the Nuclear Central America Highlands (southern Mexico and northern Central America), and the emergence of the Panamanian Isthmus. These events provide a mechanistic backdrop for cladogenesis and range disjunctions (Le and McCord 2008, Cadena 2009, Cadena et al. 2012). Multiple lines of evidence further indicate repeated

colonization of South America by *Rhinoclemmys* lineages, with at least four invasion events, several of which likely predate the final closure of the Panama Isthmus (Carr 1991, Le and McCord 2008, Cadena et al. 2012). Fossil occurrences from the Early–Middle Miocene of the Panama Canal Basin add critical temporal and geographic resolution to these dispersals, underscoring an early cryptodiran presence in Central America (Cadena et al. 2012).

Despite this biogeographic framework, interspecific relationships within *Rhinoclemmys*, remain inconsistently resolved across studies. Conflicts between mitochondrial and nuclear markers suggest complex evolutionary dynamics, potentially involving historical hybridization and gene flow, particularly among *R. melanosterna*, *R. diademata*, *R. funerea*, and *R. punctularia* (Vargas-Ramírez et al. 2013). These uncertainties are compounded by limited genomic coverage, as most previous studies have relied on only a few mitochondrial loci or partial nuclear markers (Spinks et al. 2004, Le and McCord 2008, Vargas-Ramírez et al. 2013). Although fossil records from the Miocene of Panama and the Pleistocene of Venezuela confirms the long-standing presence of *Rhinoclemmys* in the Neotropics (Cadena 2014, Cadena et al. 2012, Cadena and Carrillo-Briceño 2019), the biogeographic history of the genus remains to be studied from a molecular dating framework. These gaps highlight the need for broader phylogenomic analyses to reconcile phylogenetic conflicts and provide a more complete picture of the evolutionary history of this unique Neotropical turtle lineage.

The rise of next-generation sequencing (NGS) has fundamentally reshaped evolutionary biology by turning previously inaccessible questions into tractable genomic problems. Beyond simply increasing the volume of data, NGS has made it possible to interrogate whole genomes of non-model organisms (Ekblom and Galindo 2011), recover thousands of orthologous markers across deep evolutionary timescales (Faircloth et al. 2012), and resolve phylogenetic relationships that were previously intractable with mitochondrial or limited nuclear loci (Shaffer et al. 2017, Selvatti et al. 2023). Importantly, NGS has transformed our understanding of species boundaries, historical demography, and genomic divergence, especially in groups with large, repetitive genomes such as reptiles (Schönenberg et al. 2023). This genomic era has opened the door for comprehensive comparative analyses across lineages, enabling robust tests of biogeographic hypotheses, clarifying cryptic diversity, and revealing the tempo and mode of diversification with unprecedented resolution. In *Rhinoclemmys*, several aspects of evolutionary history remain

difficult to resolve because the group presents a combination of recent divergences, broad geographic distributions, and instances of mito-nuclear discordance that are common in many Neotropical vertebrates (Vargas-Ramírez et al. 2013). Traditional markers have provided important frameworks, but their limited genomic scope makes it challenging to distinguish among alternative evolutionary processes such as incomplete lineage sorting, historical gene flow, or deep population structure. By incorporating genome-scale nuclear data, our study provides a more comprehensive and finely resolved view of genetic relationships across the genus. This expanded genomic perspective enhances our ability to clarify species boundaries, test biogeographic hypotheses, and recover the tempo and mode of diversification in *Rhinoclemmys* with a level of detail that was previously out of reach.

In addition to providing genome-wide nuclear information, the high-coverage Illumina datasets also allowed the recovery of complete mitochondrial genomes for all sampled taxa, offering an independent line of evidence to complement and contextualize the nuclear phylogenomic inferences. Mitochondrial DNA (mtDNA) has been widely employed in phylogenetic and phylogeographic studies due to its relatively small size, conserved gene content, rapid evolutionary rate, and strict maternal inheritance (Avice 2000, Boore 1999). In turtles, mitogenomes have provided crucial insights into evolutionary relationships, historical demography, and patterns of adaptation across ecological gradients (Parham et al. 2012, Duchêne et al. 2012). Although previous studies of *Rhinoclemmys* relied on multilocus mitochondrial datasets and successfully identified signals of mitochondrial capture, these inferences were necessarily limited by the restricted genomic scope of the markers available. With complete mitochondrial genomes now recoverable from the same NGS datasets used for nuclear analyses, it will be possible to examine these introgression events with far greater resolution.

From an ecological and conservation standpoint, *Rhinoclemmys* occupies biodiversity-rich regions—including Mesoamerican forests and the Chocó biogeographic corridor—where complex landscape have historically acted as both conduits and barrier for vertebrate dispersal (De La Fuente et al. 2014). Several species face pressures from habitat alteration and local exploitation, underscoring the value of robust phylogenetic and population-level inference for effective conservation planning (Vargas-Ramírez et al. 2013). In this study, we assemble and analyze complete mitochondrial genomes and nuclear BUSCO genes from several *Rhinoclemmys* species to clarify evolutionary relationships, estimate

divergence times, and elucidate the biogeographic processes that guided the diversification of this unique lineage in the Americas.

4.2 Materials and Methods

Tissue sampling and DNA extraction

Aiming to analyse the evolutionary history of extant *Rhinoclemmys* species, we utilized Illumina paired-end short read sequencing technology to expand the currently limited genomic resources for the genus (TTWG 2025). We included seven of the nine currently recognized species of *Rhinoclemmys*, with *R. rubida* and *R. areolata* remaining unsampled. For six of the included species, we processed a single individual per species, thus preventing us of any assessment of intraspecific divergence, such as the currently recognized subspecies for *R. pulcherrima* and *R. punctularia*. *R. melanosterna* was represented by two individuals corresponding to the Eastern and Western phylogeographic lineages (Vargas-Ramírez et al. 2013, Arias-Sosa et al. 2025), which are associated with the mitochondrial capture event reported for the genus. By representing both lineages, our sampling strategy aimed to study the extent of intraspecific mitochondrial variation and to explore the potential role of historical processes of mitochondrial capture as a major cause of mito-nuclear discordance in the genus. Despite the absence of two currently recognized species, our dataset represents the principal biogeographic episodes that shaped *Rhinoclemmys* diversity: an early southward invasion during the early Miocene is attributed to the lineage leading to the extant *R. nasuta*, whose present distribution in Colombia and Ecuador is consistent with a pre-Isthmus dispersal into northwestern South America (Vargas-Ramírez et al. 2013, Le and McCord 2008). Subsequent independent colonizations into South America—comprising at least three waves during the Miocene—are inferred for lineages represented by *R. punctularia*, *R. funerea*, and *R. annulata* (Le and McCord 2008, Vargas-Ramírez et al. 2013).

Tissues were obtained from the tissue collection of the Museum of Zoology, Senckenberg Dresden, and from the Banco de Tejidos de la Biodiversidad Colombiana (BTBC), Instituto de Genética, Universidad Nacional de Colombia. Genomic DNA was extracted using the innuPREP DNA Mini Kit 2.0 (Analytik Jena), following the manufacturer's protocol, with a final elution of 80 µl of Milli-Q water.

Preliminary data processing

Libraries were prepared using the NEB Next Ultra II DNA Library Prep Kit (New England Biolabs Inc., Ipswich, MA, USA) with an average insert size of 350 bp. Sequencing was performed with 150 bp paired end reads to a target depth of approximately 40 Gb per sample. Library quality was evaluated by Qubit fluorometry and fragment analysis prior to pooling and sequencing. Raw paired-end Illumina reads generated by Novogene were first subjected to adapter and quality trimming using Trimmomatic v0.39 (Bolger et al. 2014)

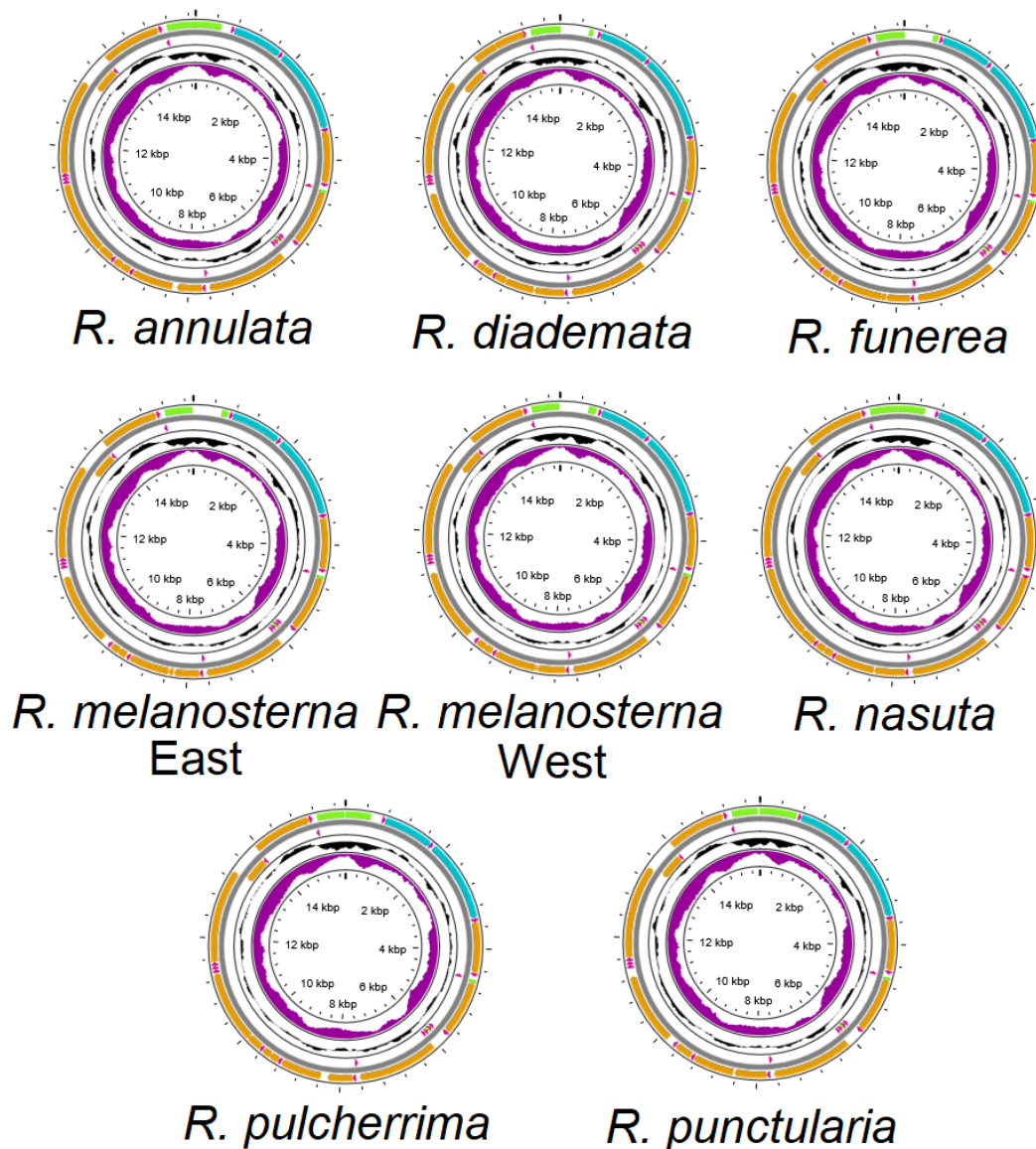


Figure D1. Circular maps of the complete mitochondrial genomes of eight *Rhinoclemmys* species reconstructed with MITOS. Each map depicts the gene organization and annotation, with protein-coding genes (orange), rRNAs (light blue), tRNAs (Pink) and repetitive regions (green) shown. The inner black plot represents GC content, and the inner purple plot indicates GC skew across the genome.

with the TruSeq3 adapter set to remove sequencing adapters and low-quality bases. Quality control was performed on all datasets both before and after trimming using FastQC v0.11 (Andrews 2010) to confirm the effectiveness of adapter removal and the overall quality profile of the reads. To further improve data quality, we removed potential contaminants by mapping trimmed reads against a curated contaminant reference database with BMap (Bushnell 2014), retaining only read pairs with $\geq 95\%$ identity that did not match contaminant sequences. The resulting “clean” paired-end datasets were used for subsequent mitochondrial assembly and nuclear genome profiling, assembly, and mapping analyses.

Mitochondrial genome assembly and alignment

The resulting high-quality read pairs were subsequently used for mitochondrial genome reconstruction with MITObim v1.9.1 (Hahn et al. 2013), a tool specifically designed for organellar genome assembly through a baiting and iterative mapping strategy. This method uses an initial reference sequence (the “bait”) to recruit homologous reads from the dataset, which are then iteratively extended and reassembled to improve contiguity and coverage. The process continues until read recruitment reaches a stationary phase, typically after fewer than ten iterations, at which point a consensus sequence is generated. Assemblies were inspected for completeness, continuity, and circularity, exporting the final consensus mitogenome from the last iteration directory.

For *Rhinoclemmys punctularia*, we initiated the assembly using the only available mitochondrial genome of the genus (*R. punctularia*, GenBank accession JN999706, Spinks et al. 2012) as a preliminary bait, despite its incomplete coverage. From the resulting assembly, we extracted the cytochrome b (Cyt-b) gene and employed it as a secondary bait to maximize recovery of the highly variable control region (D-loop). For all other *Rhinoclemmys* species, the complete mitogenome of *R. punctularia* assembled in this study—rather than the incomplete GenBank sequence—was used as the primary bait. This was followed by a second round of iterative mapping using the species-specific Cyt-b gene to optimize the assembly of the D-loop. This two-step baiting strategy allowed us to balance the use of conserved mitochondrial regions for robust read recruitment with the power of a

species-specific marker to resolve the control region, ensuring comprehensive recovery of the mitochondrial genome while improving accuracy in its most variable segment.

The assembled mitochondrial genomes were annotated using the MITOS Web Server (Bernt et al. 2013), which provides automated prediction of protein-coding genes, ribosomal RNAs, and transfer RNAs based on sequence similarity and RNA secondary structure inference. We specified the vertebrate mitochondrial genetic code and restricted the search to chordate models to improve annotation accuracy. Each annotation was manually inspected to confirm gene boundaries, particularly in variable regions such as the control region (D-loop). Open reading frames were verified for completeness and checked against expected mitochondrial gene repertoires, while tRNA predictions were validated using structural criteria. This pipeline ensured the identification and accurate delimitation of all canonical mitochondrial genes and structural features across species, thereby generating high-confidence reference mitogenomes for comparative and phylogenetic analyses.

We assembled a dataset comprised by the 13 mtDNA protein-coding genes from 16 taxa, including nine *Rhinoclemmys* species and seven outgroups (*Mauremys sinensis* – KC333650, *Cuora mouhotii* – NC010973, *Heosemys grandis* – NC032297, *Chelonoidis darwini* – NC051474, *Emys orbicularis* – JN999703, *Graptemys pseudogeographica* – PP661414, and *Platysternon megacephalum* – NC007970), yielding a final concatenated alignment of 14,836 bp. Nucleotide composition across the alignment was biased toward adenine and thymine (A+T = 59.1%), with individual base frequencies of A = 33.6%, T = 25.5%, C = 26.4%, and G = 12.9%. Phylogenetic analyses were performed with the 13 protein coding genes (PCGs), which were aligned individually with the automatic algorithm of MAFFT 7 (Katoh and Standley 2013).

Reference Genome Assembly and Ortholog-Based Phylogenomic Matrix Construction

De novo genome assembly was performed for *Rhinoclemmys funerea*, the species with the largest dataset, using SPAdes v3.15.5 (Bankevich et al. 2012) on a high-memory cluster node (1950 GB RAM, 63 threads). The resulting contigs were evaluated with QUAST to obtain assembly statistics and quality metrics. Clean paired-end reads for each species were then mapped to the *R. funerea* de novo reference genome using BWA-MEM v0.7.17 (Li 2013) with per-lane read groups. To root the phylogenomic analyses and provide broader taxonomic context within Geoemydidae, we incorporated two additional outgroup genomes from the subfamily Geoemydinae. These consisted of Illumina paired-end

sequencing datasets for *Pangshura tecta* (SRA accession: SRX26945621) and *Mauremys reevesii* (SRA accession: DRX461630). Per-lane alignments were coordinate-sorted, merged into species-level BAMs, and indexed with samtools (Li et al. 2009). We retained alignments with mapping quality ≥ 30 for downstream analyses. We evaluated alignment quality using Qualimap v2.3 in bamqc mode, executed via automated PBS scripts. For each species, Qualimap produced standard reports including coverage distribution, insert size statistics, and GC bias profiles. These qualitative metrics complemented the samtools-based mapping statistics. To obtain a consistent set of orthologous nuclear markers across the genus, we applied the BUSCO-to-Phylogeny pipeline (Wolf et al. 2023) to the set of consensus genomes. BUSCO v5.x (lineage: Sauropsida) was used to identify single-copy orthologs, which were then extracted, aligned with MAFFT, trimmed, and concatenated into a supermatrix suitable for maximum-likelihood phylogenomic inference. We randomly subsampled 2,000 single-copy BUSCO genes from the full ortholog set, yielding a final concatenated alignment of 5,149,000 bp suitable in computational capacity for downstream phylogenomic analyses.

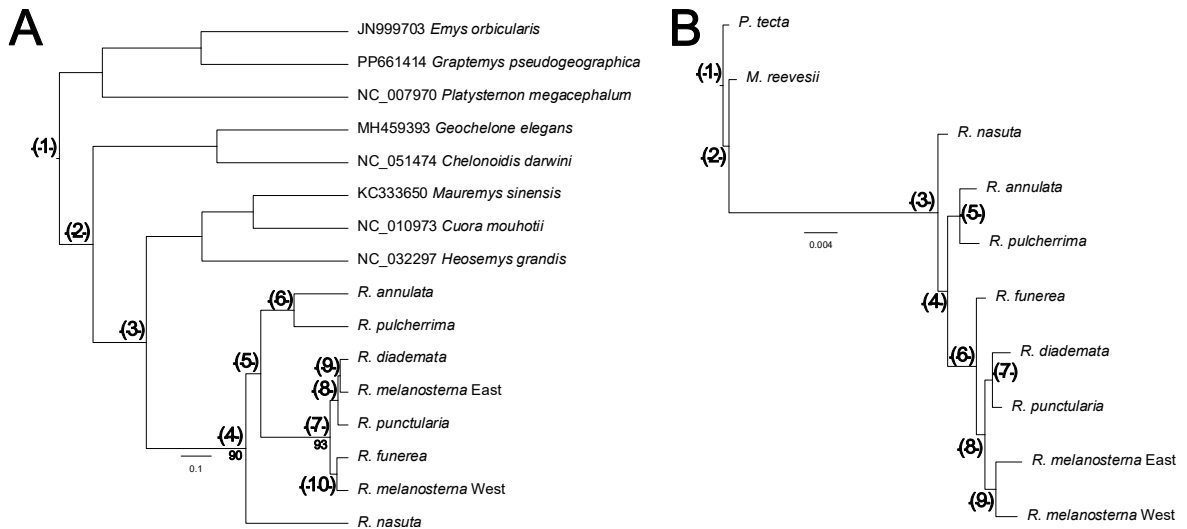


Figure D2. Phylogenetic relationships of the genus *Rhinoclemmys* based on mitochondrial and nuclear genomes. (A) Maximum-likelihood phylogeny inferred from complete mitochondrial genomes using IQ-TREE. (B) Maximum-likelihood phylogeny inferred from concatenated nuclear markers using IQ-TREE. Numbers above nodes correspond to dated nodes in Table D1 and D2, numbers below nodes represent bootstrap support values below 95.

Fossil calibration strategy

Fossil calibrations were incorporated following best-practice recommendations (Parham et al. 2012, Joyce et al. 2013) and the framework applied in Thomson et al. (2021). Based on the phylogenetic placement of Geoemydidae in Joyce et al. (2013), and the robust fossil data the authors include, we selected the next set of fossils to calibrate our trees:

1. The crown Testudinoidea (Fig. 1 - Node 17, Joyce et al. 2013), encompassing Testudinidae, Geoemydidae, and Emydidae, was constrained with a minimum of 50.3 Ma, derived from early Eocene pan-testudinoids, and a soft maximum of 149.5 Ma, corresponding to the absence of crown testudinoids in older Mesozoic strata.
2. *Platysternon* (Fig. 1 - Node 18, Joyce et al. 2013). We applied a conservative calibration to the split between Emydidae and *Platysternon*. We constrained the divergence between Emydidae and *P. megacephalum* using the oldest unambiguous emydid fossil, setting a minimum bound of 34 Ma, and applied a soft maximum of 100.5 Ma corresponding to the base of the Late Cretaceous, consistent with calibrations used across Testudinoidea.
3. *Graptemys* + *Emys* (Fig. 1 - Node 19, Joyce et al. 2013). This node was calibrated using the Oligocene fossil “*Chrysemys*” *antiqua*, widely regarded as the oldest confidently assignable crown emydid. Its morphology places it within Deirochelyinae, and the best-preserved specimen (SDSM 2754) derives from the Orellan Protoceras channels of the White River Group, establishing a minimum age of 32.0 Ma for crown Emydidae. Following the conservative approach adopted for other testudinoid groups, we set a soft maximum bound at 100.5 Ma, corresponding to the base of the Early Cretaceous, predating the appearance of stem testudinoids (“lindholmemydids”) but consistent with the absence of unambiguous crown emydids in older deposits.
4. Testuguria (Fig. 1 - Node 21, Joyce et al. 2013). Given the poor stratigraphic resolution associated with its holotype, we follow Joyce et al. (2013) in adopting the end of the Wasatchian (50.3 Ma) as a reliable minimum bound for the origin of Testuguria. To establish a soft maximum, we relied on the absence of diagnostic testugurians among the well-sampled Late Cretaceous testudinoid faunas of Asia, and set the maximum at 100.5 Ma, corresponding to the base of the Late Cretaceous, prior to the appearance of unambiguous stem testudinoids.

5. *Heosemys spinosa*–*Mauremys reevesii* (Fig. 1 - Node 22, Joyce et al. 2013) was calibrated using a minimum bound of 5.3 Ma, based on the earliest fossil representatives attributable to crown geoemydids, and a maximum bound of 65.8 Ma, corresponding to the upper limit for this clade proposed by Joyce et al. (2013).
6. Crown *Rhinoclemmys* (not calibrated in Joyce et al. 2013). We also calibrated the crown *Rhinoclemmys* node following the best-practice guidelines for fossil calibrations proposed by Parham et al. (2012) and Joyce et al. (2013), which emphasize explicit stratigraphic justification, phylogenetic assessment of fossil assignments, and the combined use of hard minimum and soft maximum bounds. As a hard minimum, we used the rhinoclemmydine specimen from the Las Cascadas Formation (~21 Ma, early Miocene) of the Panama Canal Basin, identified as *Rhinoclemmys* (sensu lato) by Bourque (2021). This fossil represents the oldest occurrence of the genus in southernmost North America and the earliest Panamanian record morphologically consistent with the *Rhinoclemmys* lineage and was chosen as a hard minimum. Establishing a soft maximum date for the clade, however, is harder, considering the scarcity of Geoemydid fossil records. There is a clear absence of rhinoclemmydines in the extensively sampled Gatuncillo (late Eocene-early Oligocene) and Culebra (early Miocene) formations of the Panama Canal Basin (Cadena et al. 2012). These formations contain rich podocnemidid, trionyhid, and testudinid assemblages but no members of Rhinoclemmydinae. Fossils of *Bridgeremys pusilla* (Hutchinson 2006, Adrian et al. 2022) and *Notapachemys oglaga* (Bourque 2021) from the middle to late Eocene of western and midwestern North America have been identified as crown rhinoclemmids. Both paleontological data and biogeographic analyses support the hypothesis that the Asian rhinoclemmids migrated to North America across the Bering Land Bridge during warmer periods in the Eocene (Bowen et al. 2002). This strait was formed about 100 mya, and as Bowen et al. (2002) and Beard (2002) argue, short global warming periods facilitated the dispersal through this strait during the late Tiffanian (57 mya), early Clarkforkian (56 mya) and early Wasatchian (55 mya). Our soft maximum was set to 66 Ma, grounded on the combination of broader evolutionary history of rhinoclemmydines and geologic eras, at the base of the Paleocene.

The fossil record of Testudinoidea provided a set of well-established calibration points that we incorporated into all divergence-time analyses. However, because mitochondrial and

nuclear datasets differ in size and structure, these calibrations were implemented in distinct ways for each dataset. The specific calibration procedures for each molecular partition are described in detail in the corresponding methodological sections.

Phylogenetic reconstruction and divergence time estimation – Complete mitogenomes

We reconstructed mitochondrial phylogeny for the genus *Rhinoclemmys* based on complete mitochondrial genomes. Partitioning schemes and substitution models were determined with PartitionFinder2 (Lanfear et al. 2017) under the Bayesian Information Criterion (BIC), allowing gene-combination strategies to minimize over-parameterization. Phylogenetic inference was conducted in IQ-TREE v2 (Minh et al. 2020) under a maximum likelihood framework with 100,000 ultrafast bootstrap replicates to assess branch support. The resulting topology, with branch lengths, was subsequently used as a guide tree for divergence time estimation in MCMCTree, implemented in PAML v4.10.7 (Yang 2007).

Table D1. Comparison of node ages estimated with MCMCTree and BEAST for the mitogenomes dataset. Values are in millions of years before present (MYa). Between parenthesis, the highest posterior density interval (95%). Fossil dating as

Node	MCMCTree				Beast				Temporal Range
	1	2	3	4	1	2	3	4	
1	92.85 (68.32-116.23)	95.88 (68.59-116.01)	95.47 (67.61-117.05)	95.20 (66.31-116.88)	90.45 (67.74-114.27)	86.18 (61.43-110.44)	88.11 (65.45-112.32)	82.62 (58.34-108.43)	Early-Late Cretaceous
2	84.69 (61.02-100.12)	84.73 (61.12-102.33)	84.34 (60.16-102.73)	84.12 (59.16-102.56)	74.00 (54.42-96.83)	70.29 (50.30-93.02)	72.36 (52.27-95.18)	67.19 (50.30-91.35)	Late Cretaceous
3	66.99 (47.52-81.85)	67.01 (47.88-82.50)	66.76 (46.66-82.08)	66.84 (45.93-82.15)	55.19 (40.19-72.22)	52.20 (36.01-71.20)	53.98 (39.09-71.28)	49.78 (34.31-69-90)	Cenozoic
4	33.94 (23.69-42.36)	33.94 (23.72-42.76)	33.89 (23.12-42.23)	33.66 (22.94-42.54)	27.49 (21.00-35.22)	25.18 (16.34-35.54)	26.62 (21.34-34.87)	23.94 (15.28-34.59)	Oligocene
5	28.98 (20.16-36.30)	28.99 (20.12-36.42)	28.94 (19.70-36.34)	28.75 (19.64-36.50)	26.32 (17.08-31.33)	21.83 (13.94-31.08)	23.08 (16.84-31.07)	20.75 (12.80-30.12)	Late Oligocene - Early Miocene
6	17.94 (12.16-22.90)	17.97 (12.34-23.11)	17.93 (12.03-23.01)	17.80 (11.82-23.05)	14.34 (8.58-19.83)	12.95 (7.24-12.95)	13.66 (8.62-19.88)	12.27 (6.65-18.56)	Earl-Middle Miocene
7	5.97 (4.09-7.65)	5.98 (4.04-7.64)	5.95 (3.94-7.61)	5.93 (3.93-7.68)	5.42 (3.40-7.35)	4.95 (3.05-7.12)	5.12 (3.23-7.22)	4.69 (2.83-6.91)	Late Miocene - Early Pliocene
8	3.35 (2.20-4.34)	3.34 (2.25-5.04)	3.34 (2.23-5.44)	3.32 (2.16-4.37)	3.21 (1.73-4.10)	2.70 (1.56-3.99)	2.79 (1.65-4.05)	2.56 (1.43-3.95)	Late Pliocene - Early Pleistocene
9	2.63 (1.69-3.47)	2.63 (1.75-3.52)	2.62 (1.70-3.03)	2.60 (1.65-3.47)	2.32 (1.26-3.28)	2.09 (1.14-3.13)	2.16 (1.19-3.22)	1.98 (1.05-3.07)	Early-Late Pleistocene
10	3.70 (2.43-4.91)	3.70 (2.41-4.93)	3.69 (2.33-4.85)	3.67 (2.35-4.88)	3.14 (1.79-4.71)	2.96 (1.61-4.35)	3.06 (1-68-4.16)	2.81 (1.48-4.40)	Late Pliocene - Early Pleistocene

follows: (1) all fossil calibrations, (2) all but crown *Rhinoclemmys* calibration, (3) all but *Graptemys* + *Emys* calibration, (4) all but crown *Rhinoclemmys* and *Graptemys* + *Emys* calibrations.

To evaluate the sensitivity of divergence-time estimates to fossil calibration choice, we conducted four parallel analyses using the complete mitochondrial dataset. The first analysis incorporated all six fossil constraints; the second excluded the newly proposed *Rhinoclemmys* calibration; the third removed calibration point 3 while retaining all others; and the fourth excluded both calibrations 3 and 6. This design allowed us to assess the influence of our newly integrated *Rhinoclemmys* fossil constraint, which was introduced here for the first time in a phylogenetic context, and to test whether the close similarity between calibrations 2 and 3 produced detectable effects on inferred node ages.

Divergence time estimation in MCMCTree was conducted in two stages. In the first step (*usedata* = 3), the gradient and Hessian matrices for branch lengths and substitution parameters were estimated from the full partitioned alignment. The Markov chain Monte Carlo (MCMC) analysis was run with a burn-in of 50,000 iterations, sampling every five steps, and retaining 20,000 samples. In the second step (*usedata* = 2), dating of the fixed topology was carried out using the likelihood approximations generated in Step 1 together with the fossil calibrations as priors. Two independent MCMC chains were run and combined after verifying convergence. Convergence was assessed in Tracer v1.7.2 (Rambaut et al. 2018), with effective sample sizes exceeding 20,000 for all parameters.

We also conducted molecular dating in BEAST v2.7 (Bouckaert et al. 2019) under the uncorrelated lognormal relaxed clock model (Drummond et al. 2006). A Birth–Death tree prior was specified, combined with the GTR+ Γ substitution model partitioned by mitochondrial genes. Two independent Markov chain Monte Carlo (MCMC) analyses were run for 50 million generations each, sampling parameters and trees every 5,000 generations. Convergence of parameters and effective mixing was assessed with Tracer v1.7 (Rambaut et al. 2018), confirming that all effective sample sizes (ESSs) exceeded 200 after discarding the first 10% of samples as burn-in. Post-burn-in trees were summarized in TreeAnnotator v2.7 to generate a maximum clade credibility (MCC) tree with mean node heights.

Phylogenetic reconstruction and divergence time estimation – BUSCO genes

For the nuclear dataset, we inferred phylogenetic relationships using a concatenated alignment of 2,000 randomly selected single-copy BUSCO genes (totaling 5,149,000 bp

after trimming). Maximum likelihood phylogeny was reconstructed in IQ-TREE v2 (Minh et al. 2020). Node support was assessed using 100,000 ultrafast bootstrap replicates. The resulting topology served as the fixed species tree for subsequent divergence-time estimation in MCMCTree. Because the nuclear dataset included only two external geoemydine lineages (*Pangshura tecta* and *Mauremys reevesii*), we employed a reduced calibration scheme using two fossil constraints, calibration 4 and calibration 6. To assess the impact of root constraints on nuclear divergence-time estimates, we performed two independent analyses that differed only in the prior specified for the root age (RootAge = 100 Ma and RootAge = 150 Ma, respectively).

Divergence-time estimation in MCMCTree followed the standard two-step procedure. First, the gradient and Hessian matrices for branch lengths and substitution parameters were estimated from the full 2,000-gene alignment (usedata = 3). Then, dating was performed with usedata = 2, using the likelihood approximation together with the fossil priors. Each analysis used a burn-in of 50,000 iterations, sampling every five steps and retaining 20,000 posterior samples. Posterior chronograms were summarized and visualized to compare temporal estimates across calibration conditions.

4.3 Results

Mitogenome organization

The complete mitochondrial genomes of the analyzed *Rhinoclemmys* species ranged between 15,800 to 16,100 bp in length and displayed the typical gene content and organization found in vertebrate mitogenomes (Fig. 1). Each mitogenome comprises 13 protein-coding genes (PCGs), 22 transfer RNA genes (tRNAs), two ribosomal RNA genes (12S and 16S), and a noncoding control region. Most genes were encoded on the heavy (H) strand, except for ND6 and eight tRNA genes (tRNA-Gln, tRNA-Ala, tRNA-Asn, tRNA-Cys, tRNA-Tyr, tRNA-Ser (UCN), tRNA-Glu, and tRNA-Pro), which were located on the light (L) strand (Fig. 1). The majority of PCGs were initiated by an ATG start codon, with occasional use of GTG in ND1 or COI. Termination codons varied across genes: TAA was used by ND1, ND2, COI, ATP6, ATP8, ND4L, ND5, and ND6, TAG was observed in ND3, while COII, COIII, ND4, and CYTB employed incomplete stop codons (T– or TA–), a common feature in vertebrate mitochondrial genomes.

The base composition of the *Rhinoclemmys* mitogenomes was biased toward adenine and thymine (A+T = 58% - 60%), consistent with other turtles. The overall AT-skew was slightly positive, while the GC-skew was negative, reflecting the typical strand asymmetry of vertebrate mitochondrial DNA. Codon usage across the 13 PCGs revealed that Leucine was encoded by the greatest number of synonymous codons, whereas amino acids such as Methionine and Tryptophan were represented by a single codon. The control region (D-loop) was the most variable portion of the mitogenomes, exhibiting length variation among species. The order of genes was conserved and identical to the canonical vertebrate mitochondrial arrangement.

Nuclear Genome Assembly

After adapter trimming and contaminant removal, more than 95–99% of reads were retained across all libraries, resulting in clean datasets ranging from approximately 300 to 490 million paired-end reads per species (based on MultiQC summaries). Read-level quality remained high after trimming, with Q30 values consistently above 93% and GC content ranging from 42–44% across *Rhinoclemmys* samples. These high-quality datasets provided sufficient depth and complexity for k-mer–based genome profiling, de novo assembly, and reference-guided comparative analyses.

The *Rhinoclemmys funerea* assembly generated with SPAdes served as the nuclear reference genome for the genus. QUAST analyses reported a total assembly length of 1.86

Table D2. Comparison of node ages estimated between two Root ages (Ma) in MCMCTree, for the BUSCO genes dataset. Values are in millions of years before present (Ma). Between parenthesis, the highest posterior density interval (95%).

Node	Root - 100	Root - 150	Temporal Range
1	58.85 (31.42-96.93)	60.33 (32.533-98.33)	Late Mesozoic - Paleocene
2	58.41 (31.26-96.35)	59.88 (32.05-97.38)	Middle Mesozoic - Oligocene
3	33.38 (18.86-55.17)	32.97 (18.55-53.94)	Eocene-Early Miocene
4	28.30 (15.31-46.57)	27.87 (15.41-45.72)	Eocene - Miocene
5	16.40 (7.30-28.39)	16.00 (7.29-27.81)	Oligocene - Miocene
6	19.48 (10.29-32.26)	18.94 (9.96-31.56)	Oligocene - Miocene
7	10.12 (4.83-17.30)	9.87 (4-71-16.66)	Miocene - Pliocene
8	15.30 (7.84-25.65)	14.959 (7.75-21.16)	Oligocene - Miocene
9	6.57 (2.43-12.49)	8.45 (6.36-18.12)	Miocene - Pliocene

Gb, comprising 281,721 contigs, with a GC content of 43.6% and an N50 of 11,803 bp (11.8 kb). The largest contig reached 121,283 bp, and the contig-length distribution showed a strong skew toward short fragments, as expected for short-read contig-level assemblies lacking long-range scaffolding. Despite this fragmentation, the assembly recovered gene-rich regions with sufficient completeness to support accurate downstream analyses. Mapping rates were exceptionally high for all species, with 100% of reads aligning to the reference and 90.5–92.8% of pairs mapping properly, while singleton alignments remained below 0.3% and duplicates were absent due to prior filtering. Consensus genomes derived from these alignments supported robust recovery of single-copy orthologs through the BUSCO-to-Phylogeny workflow.

Phylogenetics

Maximum likelihood analyses of the complete mitochondrial alignment and the 2,000-gene BUSCO nuclear matrix produced fully resolved and highly supported phylogenetic trees for *Rhinoclemmys* (Fig. 2). Both topologies showed absolute bootstrap support (UFBoot = 100) for all internal nodes. Mitochondrial and nuclear trees were broadly congruent with previous topologies (Vargas-Ramírez et al. 2013, Thomson et al. 2021), although some differences in the placement of deeper nodes were observed. In both topologies, the genus was recovered as monophyletic, *Rhinoclemmys annulata* and *R. pulcherrima* formed a sister clade and *nasuta* was resolved as the earliest-diverging lineage. Mitochondrial and nuclear phylogenies recovered markedly different evolutionary relationships within the *Rhinoclemmys melanosterna* complex. The mitochondrial tree places *R. funerea* as the sister lineage to the West clade of *R. melanosterna*, while *R. diademata* emerges as the sister group to the East clade—mirroring the well-known east–west structuring of *R. melanosterna* and yielding a non-monophyletic arrangement for the species (Fig. 2A). In contrast, the phylogeny inferred from BUSCO nuclear genes resolves an entirely different scenario: *R. funerea* represents the earliest diverging lineage within the *melanosterna* complex, and *R. diademata* and *R. punctularia* form a strongly supported sister-pair (Fig. 2B). Under this nuclear framework, *R. melanosterna* is recovered as a monophyletic species without the deep east–west split observed in mitochondrial data (Fig. 2B). This phylogenetic discordance between mitochondrial and nuclear genomes is consistent with past mitochondrial introgression or lineage-specific mitochondrial capture, processes that have shaped the evolutionary history of this group.

Divergence Times – Complete Mitogenomes

Across the 13 mitochondrial partitions, mean substitution rates (μ) ranged between 0.12 and 0.23 substitutions/site/million years, with most loci clustering between 0.16 and 0.20 (Supplementary Table D1). This relatively narrow distribution of rates suggests a high degree of homogeneity across mitochondrial regions, reinforcing the reliability of divergence time estimates. In contrast, posterior estimates of σ^2 , which reflect rate variation among lineages, were generally low to moderate ($\sigma^2 \approx 0.07$ – 0.23). The modest values indicate limited rate heterogeneity in substitution rates across the phylogeny, which is consistent with expectations under a clock-like mitochondrial genome evolution.

Divergence-time estimates generated with MCMCTree and BEAST were remarkably consistent across calibration schemes, converging on the same broad temporal framework for the evolutionary history of *Rhinoclemmys* (Table D1). Despite methodological differences between the two approaches, both recovered congruent geological placements for all major nodes, from Late Cretaceous divergences at the base of the tree to Neogene and Quaternary splits within the *melanosterna* complex. Divergence-time analyses conducted in MCMCTree yielded a temporally coherent framework for both deep and shallow divergences across the *Rhinoclemmys* phylogeny. The root of the tree (Fig. 2A, Node 1), representing the split between Emydidae + Platysternidae and all remaining taxa, was consistently placed in the Mesozoic, within the Late Cretaceous, and the subsequent divergence between Geoemydidae and Testudinidae (Fig. 2A, Node 2) also fell within the same era. The separation of *Rhinoclemmydinae* from the Asian geoemydines (Fig. 2A, Node 3) was inferred to have occurred near the Cretaceous–Paleogene transition, with the crown radiation of *Rhinoclemmys* (Fig. 2A, Node 4) emerging in the Paleogene, specifically the Oligocene. Deeper splits within the genus followed a coherent temporal sequence: the divergence of the *R. melanosterna* complex from *R. annulata* + *R. pulcherrima* (Fig. 2A, Node 5) occurred in the Late Oligocene to Early Miocene, while the MRCA of *R. annulata* and *R. pulcherrima* (Fig. 2A, Node 6) was placed firmly within the Miocene. More recent divergences—including the origin of the *R. melanosterna* complex (Fig. 2A, Node 7), the split among *R. diademata*, *R. melanosterna East*, and *R. punctularia* (Fig. 2A, Node 8), and the subsequent divergences of the eastern and western lineages of *R. melanosterna* (Fig. 2A, Nodes 9 and 10)—were consistently dated to the Miocene and Pliocene, culminating in Pleistocene divergences toward the tips. Collectively, these patterns indicate that the

temporal signal recovered by MCMCTree is highly structured and geologically consistent across the phylogeny.

The BEAST2 chronogram recovered a broadly similar temporal structure. The root (Fig. 2A, Node 1) was again placed in the Late Cretaceous, while the divergence between Geoemydidae and Testudinidae (Fig. 2A, Node 2) and the split between Rhinoclemmydinae and Geoemydinae (Fig. 2A, Node 3) were assigned to the Cretaceous–Paleogene boundary, in agreement with the MCMCTree estimates. The crown radiation of *Rhinoclemmys* (Fig. 2A, Node 4) was placed within the Paleogene, specifically the Oligocene. Subsequent splits within the genus also followed the same geological pattern: the diversification of the *R. melanosterna* complex + *R. annulata* + *R. pulcherrima* (Fig. 2A, Node 5) occurred between the Late Oligocene and Early Miocene, and the MRCA of *R. annulata* and *R. pulcherrima* (Fig. 2A, Node 6) fell within the Miocene, identical to the temporal placement recovered by MCMCTree. Likewise, the origin of the *R. melanosterna* complex (Fig. 2A, Node 7) and the successive divergences involving *R. diademata*, *R. melanosterna East*, *R. melanosterna West*, *R. funerea*, and *R. punctularia* (Fig. 2A, Nodes 8–10) unfolded through the Miocene, Pliocene, and into the Pleistocene, fully matching the temporal range inferred under the MCMC framework.

Divergence Times – BUSCO genes

At deep nodes, divergence-time estimates obtained from BUSCO nuclear genes were broadly consistent with those inferred from mitochondrial genomes, as both data sets place the major splits of the *Rhinoclemmys* lineage in the Late Cretaceous and Paleogene (Table D2). Despite differences in absolute ages, the two approaches converge on the same geological intervals for the root, the separation of Geoemydidae and Testudinidae, and the crown diversification of *Rhinoclemmys*. The MRCA of *Mauremys reevesii* and *Rhinoclemmys* (Fig. 2B, Node 2) spans from the Middle Mesozoic to the Oligocene, indicating that the split between the Asian geoemydid lineage and *Rhinoclemmys* is consistently placed deep in time, regardless of the root prior. The crown group of *Rhinoclemmys* (Fig. 2B, Node 3) is inferred to have originated in the Paleogene, between the Eocene and the early Miocene, with the subsequent split between the *R. annulata* + *R. pulcherrima* clade and the *R. melanosterna* complex (Fig. 2B, Node 4) also falling within the Eocene–Miocene interval. Within the genus, both the MRCA of *R. annulata* and *R. pulcherrima* (Fig. 2B, Node 5) and the MRCA of the *R. melanosterna* complex (Fig. 2B,

Node 6) are robustly placed in the Oligocene–Miocene, as is the deeper node uniting *R. diademata*, *R. punctularia* and *R. melanosterna* (Fig. 2B, Node 8). The more recent splits, between *R. diademata* and *R. punctularia* (Fig. 2B, Node 7) and between the eastern and western lineages of *R. melanosterna* (Fig. 2B, Node 9), are consistently dated to the Miocene–Pliocene, indicating that diversification within the *melanosterna* complex and its closest relatives is a Neogene event. However, these shallow nodes reveal a clear discrepancy with mitochondrial estimates, which place the same divergences substantially more recently. This contrast suggests that mitochondrial rates or lineage-specific introgression may have compressed the temporal signal at the tips, whereas the nuclear BUSCO framework retains an older, more gradual history of diversification.

4.4 Discussion

Our mitogenomic and nuclear BUSCO gene analyses refine the evolutionary history of *Rhinoclemmys*, the only New World lineage of the otherwise Old-World Geoemydidae, by integrating DNA sequences with two independent molecular dating frameworks (MCMCTree and BEAST), and different dating schemes. Both datasets support *Rhinoclemmys* as monophyletic (Supplementary Fig. D1), with an early radiation of *R. nasuta*, followed by the split between *R. annulata* + *R. pulcherrima* and the *melanosterna* complex (Fig. 2), a pattern consistent with previous molecular analyses (Le and McCord 2008; Spinks et al. 2004) but contrasting with morphological assessments that suggest paraphyly (Hirayama, 1985, Yasukawa et al. 2001). Within Geoemydidae, *Rhinoclemmys* is consistently allied with Asian genera (*Cuora*, *Heosemys*, *Mauremys*), though its exact position varies across studies (Spinks et al. 2004, Thomson et al. 2021). Importantly, both mitogenome frameworks yielded well-supported topologies and narrow highest posterior densities (HPDs), suggesting robust rate behavior across mitochondrial partitions (mean rate $\mu = 0.12\text{--}0.23$ substitutions/site/Ma, rate variation $\sigma^2 = 0.07\text{--}0.23$). These values reflect a relatively clock-like mitochondrial genome, consistent with prior assessments in chelonians (Duchêne et al. 2012).

A central finding of this study is the broad concordance in divergence-time estimates across mitochondrial, nuclear (BUSCO), and mixed calibration frameworks (Tables D1 and D2). Despite methodological differences, all approaches place the deepest nodes of the phylogeny in the Late Cretaceous, including the root of the tree (Node 1) and the split between *Mauremys reevesii* and *Rhinoclemmys* (Node 2). Likewise, the crown radiation of

Rhinoclemmys (Node 3) is consistently inferred to originate in the Paleogene, generally between the Eocene and early Miocene, regardless of dataset or calibration strategy. Comparable temporal patterns have been documented in other Neotropical freshwater turtles. For example, the early radiation of Kinosternidae is dated to the early Eocene (~53.4 Ma), closely matching the temporal depth we infer for the origin of the *Rhinoclemmys* crown lineage (Hurtado-Gómez et al. 2024). We agree with Carr (1991) in its hypothesized biogeographic pattern of dispersion into South America prior to its closure, as the inferred divergence times for the major lineages of the genus substantially predate the emergence of the land bridge (Montes et al. 2015). In this scenario, during the Great American Biological interchange (GABI), *Rhinoclemmys* radiation between the Americas started millions of years earlier to the closure of the Panama land bridge. Within the genus, the subsequent divergences also show substantial cross-method agreement. *R. nasuta* represents the earliest divergence from crown *Rhinoclemmys*, estimated during the Eocene-Early Miocene (Table D2). *R. nasuta*, thus, represents the earliest colonization of South America, possibly across the Late Eocene island across central Panama and the Azuero Peninsula (Baumgartner-Mora et al. 2008). Independent evidence from *Kinosternon* further supports the plausibility of pre-Isthmian dispersal in *Rhinoclemmys*. The southward invasion of *K. dunnii* may well predate the Great American Biotic Interchange (GABI; ~3 Ma), which has traditionally been linked to the final emergence of the Panamanian Isthmus (O’Dea et al. 2016). Pre-GABI migrations have likewise been proposed for other turtle lineages, including the *Trachemys dorbigni* / *T. medemi* clade (Fritz et al. 2012, 2023) as well as for diverse groups (Winston et al. 2017, Jaramillo et al. 2020, Doyle et al. 2021).

Subsequent Miocene uplift driven by the subduction of the newly formed Cocos Plate (~23 Ma) and the surfacing of key terranes such as the Baudo block fundamentally reorganized lower Central American topography (Marshall 2007). Competing geological models propose either a continuous peninsula connecting Chortis to northwestern South America between ~25 and 16 Ma (Kirby et al. 2008) or, alternatively, a Mid-Miocene to Pliocene volcanic archipelago separated by persistent seaways (Coates et al. 2004). Both scenarios imply repeated opportunities for over-water dispersal followed by periods of terrestrial connectivity, conditions that align well with our sequence of colonizations of South America by subsequent invasions of *R. annulata* (Table D2). The split between the *R. annulata* + *R. pulcherrima* clade and the *R. melanosterna* complex (Node 4) consistently falls between the Eocene and Miocene, and both the MRCA of *R. annulata* and *R. pulcherrima* (Node 5)

and the origin of the *R. melanosterna* complex (Node 6) are reliably recovered in the Oligocene–Miocene in all analyses. Moreover, lower Central American topography reached its modern configuration after a major collision with South America between 12.8 and 7.1 Ma, an event that established strike-slip systems along the Darién region and substantially reduced the Atrato marine gateway (Coates et al. 2004). This temporal window coincides closely with the divergences among the *melanosterna* complex, suggesting that Miocene tectonic accretion and the progressive emergence of land between Central and South America played a central role in structuring the early diversification of the genus. A Miocene concentration of cladogenetic events is not unique to *Rhinoclemmys*: independent time-calibrated phylogenies show that major divergences within Staurotypinae, *Kinosternon*, and *Sternotherus* also date to ~18–14 Ma, and that similar Miocene-centred radiations characterize several American turtle clades (Hurtado-Gómez et al. 2024; Le and McCord 2008; Fritz et al. 2012a; Spinks et al. 2016; Thomson et al. 2021).

Only nodes involving *R. diademata*, *R. punctularia*, and *R. melanosterna* (Nodes 7–9) do not align across methods, with divergence events placed between the Miocene and Pliocene, but this might be ruled as minor shifts toward slightly younger ages in mitochondrial estimates. These differences are restricted to shallow nodes and do not alter the broader temporal structure recovered across methods. Comparable patterns of repeated pre-Isthmian colonizations have also been documented in slider turtles (*Trachemys*). Using combined mitochondrial and nuclear loci, Fritz et al. (2012, 2023, 2024) showed that South America was colonized twice independently from southern Central America, with basal *Trachemys* lineages originating in the Late Miocene–Early Pliocene and the ancestor of *T. dorbigni* diverging ~7–8.6 Ma, well before final closure of the Panamanian Isthmus. This independent evidence from another Neotropical emydid lineage coincides with our inference that *Rhinoclemmys* underwent multiple Miocene dispersal events into South America, rather than a single post-Pliocene expansion.

Both chronograms indicate Miocene diversification for the clade containing *R. diademata*, *R. funerea*, *R. melanosterna*, and *R. punctularia*, though the precise timing differs. Critically, all scenarios predate the late Pliocene completion of the Panamanian land bridge (~3 Ma), providing strong support for the pre-Isthmian dispersal events. These findings corroborate the hypothesis of multiple independent invasions into South America. Le and McCord (2008) suggested at least four: an early Miocene invasion (represented by *R. nasuta*) and three subsequent Miocene dispersals (represented by *R. punctularia*, *R. funerea*, and *R.*

annulata). Our results align with this model, particularly when integrated with fossil evidence. The discovery of †*R. panamaensis* in the Miocene of Panama (Cadena et al. 2012) provides direct evidence of the genus in Central America before final isthmian closure. Early–Middle Miocene records from Panama (Cadena 2014, Cadena et al. 2012) validate pre-Pliocene presence of *Rhinoclemmys* near South America, while Pleistocene fossil from Venezuela (Cadena and Carrillo-Briceño 2019) demonstrates its persistence in semi-arid habitats, highlighting its ecological breadth. These fossils independently corroborate with the multi-invasion model and are more congruent with Miocene divergence times than with the implausibly recent ages recovered by some mitochondrial-only analyses. In contrast, scenarios invoking exclusively post-Pliocene invasions (Duellman 1979, Vanzolini and Heyer 1985) are inconsistent with both our chronograms and the paleontological record. When contrasted with *Trachemys*, which exhibits Late Miocene–Early Pliocene basal splits and primarily Early Pleistocene crown radiations in its Central and South American lineages (Fritz et al. 2012), *Rhinoclemmys* emerges as a substantially older Neotropical geoemydid radiation.

The divergence times we infer for the *melanosterna* complex (Nodes 7–9, Table D2) coincide with a period of intense geological and hydrological reorganization across northern South America during the late Miocene and early Pliocene. Sedimentological and paleontological evidence from western Amazonia indicates that 9–6.5 Ma environments were dominated by megafan systems and highly avulsive rivers (e.g., Solimões Formation), producing unstable drainage networks and repeated episodes of isolation and reconnection among lowland basins (Latrubesse et al. 2010). Such geodynamic conditions, coupled with the tectonic reactivation of Andean foreland structures around ~5 Ma, are consistent with scenarios that promote both lineage divergence and secondary contact. Thus, the diversification of the *melanosterna* complex occurred on par with the diversity explosion during the Miocene, a period marked by rapid tectonic uplift, volcanic accretion, and major reorganizations of drainage systems throughout Lower Central America and northern South America. Geological evidence indicates that by the middle to late Miocene (ca. 15–7 Ma), emerging landmasses, narrowing seaways such as the Atrato corridor, and the progressive collision between These shifts likely generated alternating episodes of geographic isolation and secondary contact among *Rhinoclemmys* lineages, conditions that are fully consistent with the divergence times we infer for *R. melanosterna*, *R. diademata*, and *R. punctularia* (Table D3). Comparable temporal patterns have been documented in other

Neotropical freshwater turtles: podocnemidid river turtles show a Late Eocene to Middle Miocene radiation associated with global cooling, aridification, and major phases of Andean uplift (Vargas-Ramírez et al. 2008). A similar temporal pattern has been documented in *Kinosternon*, whose major diversification also occurred during the Miocene (Hurtado-Gómez et al. 2024). This convergence across distantly related lineages suggests that Neogene geodynamic processes broadly structured the diversification of multiple South American turtle clades, including *Rhinoclemmys*.

Our data suggest that mito–nuclear discordance observed in *R. melanosterna* may have arisen in a landscape undergoing rapid geomorphological change, which would have facilitated ancient episodes of mitochondrial introgression among neighboring lineages. Vargas-Ramírez et al. (2013) showed that *R. melanosterna* is non-monophyletic with respect to *R. diademata*, *R. funerea*, and *R. punctularia* in mtDNA, yet forms a monophyletic clade in nuclear DNA. Our expanded sampling—including both the eastern and western mitochondrial lineages of *R. melanosterna*—recovers the same two patterns (Fig. D2). Mitogenomes allow us to place the discordant signals within a temporal framework: one event in which the western lineage carries mitochondria closely related to *R. funerea* during the middle–late Miocene, and a second involving the eastern lineage carrying mitochondria related to *R. diademata* during the late Miocene–early Pliocene (Tables D2–D3). However, despite these discordant mitochondrial histories, the BUSCO nuclear dataset recovers *R. melanosterna* as fully monophyletic, with all relevant divergences placed in the Miocene–Pliocene. This congruent nuclear signal strongly suggests that the mito–nuclear conflict reflects ancient mitochondrial introgression, rather than incomplete lineage sorting.

Conclusions

By integrating genomes with fossil calibrations and dual dating methods, we have reconstructed a comprehensive evolutionary framework for *Rhinoclemmys*. Our results support monophyly of the genus, identify *R. nasuta* as an early-divergent South American lineage, and confirm multiple Miocene invasions into the continent. Although MCMCTree and BEAST yield contrasting chronologies, both favor pre-Isthmian diversification scenario, a conclusion reinforced by Miocene fossils from Panama. Mito–nuclear discordance in the *R. melanosterna* complex underscores the need for nuclear genomic analyses to unravel its evolutionary history. Conservation priorities emerge from recognizing deeply divergent lineages. *Rhinoclemmys* serves as a model for understanding how tectonics, seaway

transitions, and ecological complexity shape Neotropical turtle diversification, highlighting the urgent need to conserve these evolutionarily significant lineages.

4.5 References

Adrian, B., Smith, H. F., Hutchison, J. H., & Townsend, K. B. (2022). Geometric morphometrics and anatomical network analyses reveal ecospace partitioning among geoemydid turtles from the Uinta Formation, Utah. *The Anatomical Record*, 305(6), 1359-1393.

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Babraham Bioinformatics. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>

Arias-Sosa, L. A., Rodríguez-Castro, K. G., Agudelo-González, M. H., Ramos-Villalba, B., Cuadrado-Ríos, S., Brieva, C., del Valle-Useche, C., Balcero-Deaquiz, C., and Vargas-Ramírez, M. (2025). Mitochondrial sequencing to guide the management of endangered turtles in Colombia. *Revista de Biología Tropical*, 73(1), e60604-e60604.

Ascarrunz, E, Joyce, W. G, and Smith, T. (2021). The phylogenetic relationships of geoemydid turtles from the Eocene Messel Pit Quarry. *PeerJ*, 9, e11805. <https://doi.org/10.7717/peerj.11805>

Avise, J. C. (2000). *Phylogeography: The history and formation of species*. Harvard University Press.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., ... and Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology*, 19(5), 455-477.

Baumgartner-Mora, C., Baumgartner, P. O., Buchs, D. M., Bandini, A. N., & Flores, K. (2008). Palaeocene to Oligocene Foraminifera from the Azuero Peninsula (Panama): The timing of seamount formation, accretion and forearc overlap, along the Mid American Margin. Available from: https://serval.unil.ch/resource/serval:BIB_C2D31C7C6363.P001/REF (last accessed August 30, 2017).

Beard, C. (2002). East of Eden at the Paleocene/Eocene boundary. *Science*, 295(5562), 2028-2029.

- Bernt, M, Donath, A, Jühling, F, Externbrink, F, Florentz, C, Fritzscht, G, ... Stadler, P. F. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, 69(2), 313–319. <https://doi.org/10.1016/j.ympev.2012.08.023>
- Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Research*, 27(8), 1767–1780. <https://doi.org/10.1093/nar/27.8.1767>
- Bolger, A. M, Lohse, M, and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120.
- Bouckaert, R, Vaughan, T. G, Barido-Sottani, J, Duchêne, S, Fourment, M, Gavryushkina, A, ... Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 15(4), e1006650. <https://doi.org/10.1371/journal.pcbi.1006650>
- Bourque, J. (2021). A new geoemydid (Testudines, aff. Rhinoclemmydinae) from the upper Eocene Chadron Formation (White River Group) of northwestern Nebraska. *Bulletin of the Florida Museum of Natural History*, 58(5), 86-101.
- Bowen, G. J., Clyde, W. C., Koch, P. L., Ting, S., Alroy, J., Tsubamoto, T., ... & Wang, Y. (2002). Mammalian dispersal at the Paleocene/Eocene boundary. *Science*, 295(5562), 2062-2065.
- Bushnell, B. (2014). BBMap: a fast, accurate, splice-aware aligner.
- Cadena, E. A. (2014). The fossil record of turtles in Colombia; a review of the discoveries, research and future challenges. *Acta Biológica Colombiana*, 19(3), 333-339.
- Cadena, E, Bourque, J. R, Rincon, A. F, Bloch, J. I, Jaramillo, C. A, and Macfadden, B. J. (2012). New turtles (Chelonia) from the late Eocene through late Miocene of the Panama Canal Basin. *Journal of Paleontology*, 86(3), 539-557.
- Cadena, E. A, and Carrillo-Briceño, J. D. (2019). First Fossil of *Rhinoclemmys* Fitzinger, 1826 (Cryptodira, Geoemydidae) East of the Andes. *South American Journal of Herpetology*, 14(1), 19-23.
- Carr, J. L. (1991). Phylogenetic analysis of the Neotropical turtle genus *Rhinoclemmys* (Testudines: Bataguridae). *Herpetologica*, 47(2), 191–201.

- Claude, J., Zhang, J. Y., Li, J. J., Mo, J. Y., Kuang, X. W., and Tong, H. (2012). Geoemydid turtles from the Late Eocene Maoming basin, southern China. *Bulletin de la Société géologique de France*, 183(6), 641-651.
- Coates, A. G., Collins, L. S., Aubry, M. P., and Berggren, W. A. (2004). The geology of the Darien, Panama, and the late Miocene-Pliocene collision of the Panama arc with northwestern South America. *Geological Society of America Bulletin*, 116(11-12), 1327-1344.
- De La Fuente, M., Sterli, J., and Maniel, I. (2014). *Origin, evolution and biogeographic history of South American turtles* (p. 168). Heidelberg: Springer.
- Doyle, E. D., Prates, I., Sampaio, I., Koiffmann, C., Silva Jr, W. A., Carnaval, A. C., and Harris, E. E. (2021). Molecular phylogenetic inference of the howler monkey radiation (Primates: *Alouatta*). *Primates*, 62(1), 177-188.
- Drummond, A. J., Ho, S. Y. W., Phillips, M. J., and Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4(5), e88. <https://doi.org/10.1371/journal.pbio.0040088>
- Duchêne, S., Archer, F. I., Vilstrup, J., Caballero, S., and Morin, P. A. (2011). Mitogenome phylogenetics: the impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. *PloS One*, 6(11), e27138.
- Duellman, W. E. (1979). *The South American herpetofauna: its origin, evolution, and dispersal* (Vol. 7, No. 1979). Lawrence, KS: Museum of Natural History, University of Kansas.
- Ekblom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, 107(1), 1-15.
- Ernst, C. H. (1978). A revision of the Neotropical turtle genus *Callopsis* (Testudines: Emydidae: Batagurinae). *Herpetologica*, 113-134.
- Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, 61(5), 717-726.

Fritz, U., Stuckas, H., Vargas-Ramírez, M., Hundsdörfer, A. K., Maran, J., & Päckert, M. (2012). Molecular phylogeny of Central and South American slider turtles: implications for biogeography and systematics (Testudines: Emydidae: *Trachemys*). *Journal of Zoological Systematics and Evolutionary Research*, 50(2), 125-136.

Fritz, U., Herrmann, H. W., Rosen, P. C., Auer, M., Vargas-Ramírez, M., & Kehlmaier, C. (2024). *Trachemys* in Mexico and beyond: Beautiful turtles, taxonomic nightmare, and a mitochondrial poltergeist (Testudines: Emydidae). *Vertebrate Zoology*, 74, 435-452.

Fritz, U., Herrmann, H. W., Rosen, P. C., Auer, M., Vargas-Ramírez, M., & Kehlmaier, C. (2024). *Trachemys* in Mexico and beyond: Beautiful turtles, taxonomic nightmare, and a mitochondrial poltergeist (Testudines: Emydidae). *Vertebrate Zoology*, 74, 435-452.

Hahn, C, Bachmann, L, and Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research*, 41(13), e129-e129.

Hirayama, R. (1984). Cladistic analysis of batagurine turtles (Batagurinae: Emydidae: Testudinidae); a preliminary result. *Studia geologica salmanticensia*, 1, 140-157.

Hutchison, J. H. (2006). *Bridgeremys* (Geoemydidae, Testudines), a new genus from the middle Eocene of North America. *Fossil Turtle Research*, 1, 63-83.

Joyce, W. G, Parham, J. F, Lyson, T. R, Warnock, R. C, and Donoghue, P. C. (2013). A divergence dating analysis of turtles using fossil calibrations: an example of best practices. *Journal of Paleontology*, 87(4), 612-634.

Katoh, K, and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. <https://doi.org/10.1093/molbev/mst010>

Kirby, M. X., Jones, D. S., & MacFadden, B. J. (2008). Lower Miocene stratigraphy along the Panama Canal and its bearing on the Central American Peninsula. *PLoS One*, 3(7), e2791.

Lanfear, R., Frandsen, P. B., Wright, A. M, Senfeld, T., and Calcott, B. (2017). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular

and morphological phylogenetic analyses. *Molecular Biology and Evolution*, 34(3), 772-773.

Latrubesse, E. M., Cozzuol, M., da Silva-Caminha, S. A., Rigsby, C. A., Absy, M. L., & Jaramillo, C. (2010). The Late Miocene paleogeography of the Amazon Basin and the evolution of the Amazon River system. *Earth-Science Reviews*, 99(3-4), 99-124.

Le, M., and McCord, W. P. (2008). Phylogenetic relationships and biogeographical history of the genus *Rhinoclemmys* Fitzinger, 1835 and the monophyly of the turtle family Geoemydidae (Testudines: Testudinoidea). *Zoological Journal of the Linnean Society*, 153(4), 751–767. <https://doi.org/10.1111/j.1096-3642.2008.00413.x>

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... and 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079.

Marshall, J. S. (2007). *The geomorphology and physiographic provinces of Central America. Central America: geology, resources and hazards*, 1, 75-121.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., and Lanfear, B. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37(5), 1530–1534. <https://doi.org/10.1093/molbev/msaa015>

O’Dea, A., Lessios, H. A., Coates, A. G., Eytan, R. I., Restrepo-Moreno, S. A., Cione, A. L., ... and Jackson, J. B. (2016). Formation of the Isthmus of Panama. *Science advances*, 2(8), e1600883.

Parham, J. F., Donoghue, P. C. J., Bell, C. J., Calway, T. D., Head, J. J., Holroyd, P. A., ... Benton, M. J. (2012). Best practices for justifying fossil calibrations. *Systematic Biology*, 61(2), 346–359. <https://doi.org/10.1093/sysbio/syr107>

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, 67(5), 901–904. <https://doi.org/10.1093/sysbio/syy032>

Selvatti, A. P., Moreira, F. R. R., Carvalho, D. C., Prosdocimi, F., de Moraes Russo, C. A., & Junqueira, A. C. M. (2023). Phylogenomics reconciles molecular data with the rich fossil record on the origin of living turtles. *Molecular Phylogenetics and Evolution*, *183*, 107773.

Shaffer, H. B., McCartney-Melstad, E., Near, T. J., Mount, G. G., & Spinks, P. Q. (2017). Phylogenomic analyses of 539 highly informative loci dates a fully resolved time tree for the major clades of living turtles (Testudines). *Molecular Phylogenetics and Evolution*, *115*, 7-15.

Schöneberg, Y., Winter, S., Arribas, O., Di Nicola, M. R., Master, M., Owens, J. B., ... & Fritz, U. (2023). Genomics reveals broad hybridization in deeply divergent Palearctic grass and water snakes (*Natrix* spp.). *Molecular phylogenetics and evolution*, *184*, 107787.

Spinks, P. Q., Shaffer, H. B., Iverson, J. B., and McCord, W. P. (2004). Phylogenetic hypotheses for the turtle family Geoemydidae. *Molecular phylogenetics and evolution*, *32*(1), 164-182.

Spinks, P. Q., Thomson, R. C., Zhang, Y., Che, J., Wu, Y., and Shaffer, H. B. (2012). Species boundaries and phylogenetic relationships in the critically endangered Asian box turtle genus *Cuora*. *Molecular Phylogenetics and Evolution*, *63*(3), 656-667.

Thomson, R. C., Spinks, P. Q., and Shaffer, H. B. (2021). A global phylogeny of turtles reveals a burst of climate-associated diversification on continental margins. *Nature Communications*, *12*, 6220. <https://doi.org/10.1038/s41467-021-26398-3>

TTWG [Turtle Taxonomy Working Group: Rhodin, A. G. J., Iverson, J. B., Fritz, U., Gallego-García, N., Georges, A., Shaffer, H. B., and van Dijk, P. P.] (2025). *Turtles of the World: Annotated Checklist and Atlas of Taxonomy, Synonymy, Distribution, and Conservation Status (10th Ed.)*. Chelonian Research Monographs, *10*, 1–575.

Vanzolini, P. E., and Heyer, W. R. (1985). The American herpetofauna and the interchange. In *The great American biotic interchange* (pp. 475-487). Boston, MA: Springer US.

Vargas-Ramírez, M., Carr, J. L., and Fritz, U. (2013). Complex phylogeography in *Rhinoclemmys melanosterna*: Conflicting mitochondrial and nuclear evidence suggests past hybridization (Testudines: Geoemydidae). *Zootaxa*, *3670*(2), 238–254. <https://doi.org/10.11646/zootaxa.3670.2.8>

Winston, M. E., Kronauer, D. J., and Moreau, C. S. (2017). Early and dynamic colonization of Central America drives speciation in Neotropical army ants. *Molecular Ecology*, 26(3), 859-870.

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8), 1586–1591. <https://doi.org/10.1093/molbev/msm088>

Yasukawa, Y, Hirayama, R, and Hikida, T. (2001). Phylogenetic relationships of geoemydine turtles (Reptilia: Bataguridae). *Current Herpetology*, 20(2), 105-133.

5. Conclusions

5.1 Conclusions

The Biogeographic Chocó stands as one of the planet's most biologically rich and unique areas, a hotspot of endemism and evolutionary novelty shaped by extreme rainfall, climatic stability, and complex topography (Poveda et al. 2004, Mittermeier et al. 2011). Its dual role as a Pleistocene refuge and as a long-term center of diversification since at least the Miocene (Jaramillo-Vivanco et al. 2010; Arteaga et al. 2016; Pérez-Escobar et al. 2019) makes it an invaluable natural laboratory for exploring evolutionary processes. However, its fauna remains poorly studied due to remoteness, sociopolitical challenges, and limited sampling (Ortiz Lancheros 2022, Otálvaro-Marín et al. 2023). In this context, this dissertation sought to deepen our understanding of chelonian biodiversity in the Chocó by generating new genetic and genomic resources, evaluating patterns of population connectivity and demographic history, and situating regional taxa in broader phylogenomic contexts. By integrating mitochondrial genomics, multilocus phylogeography, population genetics, and phylogenomics, this work provides insights into evolutionary and conservation processes that extend well beyond the focal species, while also contributing practical knowledge to directly inform their management and ensure their long-term persistence.

Conservation biology increasingly recognizes that effective strategies must integrate molecular data, as ecological observations alone often fail to reveal the demographic and evolutionary forces shaping species persistence (Moritz 1994, Avise 2000, Funk et al. 2012). In regions such as the Chocó, where complex geological and climatic dynamics have produced both ancient radiations and recent diversification, molecular approaches offer a unique window into the processes that maintain or erode biodiversity. These studies are urgently needed, as anthropogenic pressures, including habitat destruction and wildlife trade (Asprilla-Perea et al. 2013, Páez et al. 2022), negatively impact turtles populations at

an alarming rate. By combining molecular and ecological perspectives, this dissertation contributes to identifying evolutionary units, inferring historical connectivity, and highlighting genetic vulnerabilities, thereby addressing critical data deficiencies for chelonians in one of the world's most threatened biodiversity hotspots.

Chapter 1. Phylogenetic analyses based on complete mitochondrial genomes place *Trachemys medemi* within the South American *Trachemys* clade and recover it as sister to *T. venusta* when only complete mitogenomes are considered; analyses that expand taxon coverage with additional mtDNA data recover *T. medemi* as sister to *T. dorbigni*. The newly assembled mitogenomes (16,711–16,810 bp; 13 protein-coding genes, 22 tRNAs, two rRNAs, and a control region) show the conserved 37-gene architecture, AT-biased composition, and expected skews. Beyond topology, these references provide a practical anchor to standardize and validate cost-effective population markers such as the control region (D-loop), improving cross-study comparability and inference on diversity, connectivity, and demographic history.

Chapter 2. Phylogeographic analyses of the white-lipped mud turtle (*Kinosternon leucostomum*) reveal shallow mitochondrial structure dominated by isolation-by-distance, with two consistent breaks: across the Hess Escarpment in Central America and between the Middle and Low Magdalena Valley. Chocoan samples are nested within a widespread clade extending from the Lower Magdalena Valley into mid-Central America, consistent with the species' ecology and previously recognized subspecies patterns. Overall, lowland corridors appear permeable, and geographic distance—not discrete hard barriers—best explains genetic variation across the sampled range.

Chapter 3. Population-genetic analyses of *T. medemi* indicate largely panmictic populations throughout the Atrato basin: most variance occurs within, rather than among, localities, and mitochondrial and microsatellite markers concordantly show shallow structure. The pooled effective population size is ~246 (95% CI: 28–1305); nuclear diversity is low (mean $H_e = 0.479$); gene flow is asymmetric and limited in places, and there is no signal of recent bottlenecks. Together, these results point to historical connectivity coupled with current vulnerability to drift and demographic fluctuations.

Chapter 4. Phylogenomic inference within *Rhinoclemmys* resolves uncertainties that might change the phylogenetic history of the Genus. The genus likely originated in Central America during the Eocene, with early radiations into South America, with the Chocoan

endemic *R. nasuta* was recovered as the first radiation earlier than previously expected. The concentration of multiple *Rhinoclemmys* species in the Chocó, resolved under our phylogenomic analysis, provides an explicit comparative context to interpret coexistence, lineage divergence, and trait evolution within the genus.

Taken together, the results underscore immediate conservation needs. Endemic turtles in the Chocó face intense hunting, egg collection, habitat loss, and illegal trade; *T. medemi* is among the most trafficked. Constrained ranges, low nuclear diversity, and moderate *Ne* make populations susceptible to environmental shocks. Management should prioritize maintaining aquatic corridors, safeguarding nesting and floodplain habitats, and establishing long-term genetic monitoring to detect early declines.

Collectively, these four chapters demonstrate the power of integrating genomic, phylogeographic, and population-level approaches to illuminate biodiversity in one of the world's most complex and threatened regions. The new mitogenomes provide essential references for both evolutionary inference and practical population studies, enabling the standardization of cost-effective markers such as the D-loop. The phylogeographic analysis of *K. leucostomum* demonstrates how ecological traits mediate responses to geographic barriers, while the population genetic study of *T. medemi* underscores the precarious balance between short-term persistence and long-term vulnerability in an endemic species facing severe anthropogenic pressures. Finally, the phylogenomic evaluation of *Rhinoclemmys* situates regional diversity within a broader evolutionary framework, highlighting the Chocó's role as both refuge and engine of diversification. Collectively, these results reaffirm the global evolutionary significance of the Chocó (Arteaga et al. 2016, Pérez-Escobar et al. 2019, Sedano-Cruz et al. 2024) while also underscoring the urgency of conservation interventions (Myers et al. 2000, Lovich et al. 2018, Stanford et al. 2020).

5.2 Recommendations

The broader implications of this dissertation extend beyond turtles. It illustrates how molecular approaches can fill critical knowledge gaps in poorly studied taxa, inform conservation assessments, and provide the scientific foundation for adaptive management in biodiversity hotspots. In the Chocó, where political instability, habitat loss, and climate change converge (Otálvaro-Marín et al. 2023, Caicedo-Rivas et al. 2022), understanding genetic connectivity and diversity is a practical necessity for ensuring species persistence.

Moreover, these findings highlight the importance of genomic baselines for reassessing threat categories and informing conservation policies, from delineating management units to identifying candidates for genetic rescue. Turtles, with their combination of evolutionary distinctiveness and ecological vulnerability, exemplify why genetic monitoring should be a cornerstone of conservation planning (Amos and Balmford 2001, Frankham et al. 2014, Hohenlohe et al. 2021).

Looking ahead, the transition to conservation genomics offers unprecedented opportunities. Reduced-representation methods such as RAD-seq and whole-genome sequencing will enable finer-scale resolution of adaptive variation, demographic history, and genome-wide connectivity (Chattopadhyay et al. 2019, Hohenlohe et al. 2021). Integrating these genomic data with ecological, spatial, and sociopolitical information will be essential for designing holistic conservation strategies. In the Chocó, this means coupling molecular insights with habitat restoration, community-based management, and policy interventions that address both ecological and human dimensions of biodiversity loss. Future research should also expand taxonomic coverage, as many chelonian species in the region remain unstudied genetically, leaving critical blind spots in our understanding of their status and resilience.

In conclusion, this dissertation contributes novel genetic and genomic resources, clarifies evolutionary relationships, and provides the first population genetic insights for an endemic and threatened chelonian of the Chocó. These findings demonstrate the profound scientific and conservation value of molecular studies in hotspots like the Chocó, where evolutionary uniqueness and conservation urgency intersect. Preserving these turtles is not only about protecting species but also about safeguarding evolutionary processes and ecological functions that sustain one of the world's richest yet most imperiled regions. Strengthening the genomic foundation for conservation in the Chocó is therefore a critical investment in our global biological heritage and vital step toward ensuring the resilience of its extraordinary biodiversity.

5.3 References

Amos, W., and Balmford, A. (2001). When does conservation genetics matter?. *Heredity*, 87(3), 257-265.

- Arroyave Bermúdez, A., Henao, M., and Patiño, A. (2014). Illegal trade of freshwater turtles in the Colombian Pacific. *Herpetological Conservation and Biology*, 9(2), 456–463.
- Arteaga, A., Pyron, R. A., Peñafiel, N., Romero-Barreto, P., Culebras, J., Bustamante, L., Yáñez-Muñoz, M. H., and Guayasamin, J. M. (2016). Comparative phylogeography reveals cryptic diversity and repeated patterns of cladogenesis for amphibians and reptiles in northwestern Ecuador. *PLoS One*, 11(4), e0151746.
- Asprilla-Perea, J., Serna-Agudelo, J. E., and Palacios-Asprilla, Y. (2013). Diagnóstico sobre el decomiso de fauna silvestre en el departamento del Chocó (Pacífico Norte colombiano). *Revista UDCA Actualidad and Divulgación Científica*, 16(1), 175–184.
- Avice, J. C. (2000). *Phylogeography: the history and formation of species*. Harvard University Press.
- Caicedo-Rivas, G., Salas-Moreno, M., and Marrugo-Negrete, J. (2022). Health risk assessment for human exposure to heavy metals via food consumption in inhabitants of middle basin of the Atrato River in the Colombian Pacific. *International Journal of Environmental Research and Public Health*, 20(1), 435.
- Chattopadhyay, B., Garg, K. M., and Rheindt, F. E. (2019). Conservation genomics in the era of population genomics: insights from avian studies. *Avian Research*, 10, 30.
- Cucalón, J., Reyes-Puig, J. P., Yáñez-Muñoz, M. H., and Arteaga, A. (2022). Phylogeographic patterns in Ecuadorian reptiles. *Molecular Phylogenetics and Evolution*, 170, 107405.
- Frankham, R., Bradshaw, C. J., and Brook, B. W. (2014). Genetics in conservation management: revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation*, 170, 56–63.
- Funk, W. C., McKay, J. K., Hohenlohe, P. A., and Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology and Evolution*, 27(9), 489–496.
- Gibbons, J. W. (1987). Why do turtles live so long?. *BioScience*, 37(4), 262–269.

Hohenlohe, P. A., Funk, W. C., and Rajora, O. P. (2021). Population genomics for wildlife conservation and management. *Molecular Ecology*, 30(1), 62-82.

Jaramillo-Vivanco, D., Ríos, F., and Bustamante, L. (2010). Pleistocene refugia in northwestern South America: molecular evidence from amphibians and reptiles. *Journal of Biogeography*, 37, 123–135.

Jaramillo, A. F., De La Riva, I., Guayasamin, J. M., Chaparro, J. C., Gagliardi-Urrutia, G., Gutierrez, R. C., ... & Castroviejo-Fisher, S. (2020). Vastly underestimated species richness of Amazonian salamanders (Plethodontidae: *Bolitoglossa*) and implications about plethodontid diversification. *Molecular Phylogenetics and Evolution*, 149, 106841.

Lovich, J. E., Ennen, J. R., Agha, M., and Gibbons, J. W. (2018). Where have all the turtles gone, and why does it matter? *BioScience*, 68(10), 771–781.

Mittermeier, R. A., Turner, W. R., Larsen, F. W., Brooks, T. M., and Gascon, C. (2011). Biodiversity hotspots. In: Zachos, F. E., and Habel, J. C. (eds) *Global biodiversity conservation: the critical role of hotspots*. Springer, London, pp. 3–22.

Moritz, C. (1994). Defining 'Evolutionarily Significant Units' for conservation. *Trends in Ecology and Evolution*, 9(10), 373–375.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., and Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403(6772), 853-858.

Ortiz Lancheros, C. A. (2022). Entre la fragilidad de la paz y la persistencia de la guerra: El caso de la subregión del Bajo Atrato, Chocó, Colombia. *Revista Ratio Juris*, 17(34), 319–342.

Otálvaro-Marín, B., Parra-López, M. Y., and Klinger-Cundumí, E. (2023). Análisis de las injusticias sociales, ambientales y territoriales del departamento del Chocó, Colombia. *Prospectiva*, 36, e20212476.

Poveda, G., Rojas, C. A., Rudas, A., and Rangel, O. (2004). El Chocó biogeográfico: ambiente físico. In: *Colombia diversidad biótica IV, El Chocó biogeográfico/Costa Pacífica* (pp. 1–22).

Páez, V. P., Bock, B. C., Alzate-Estrada, D. A., Barrientos-Muñoz, K. G., Cartagena-Otalvaro, V. M., Echeverry-Alcendra, A., ... and Vallejo-Betancur, M. M. (2022). Turtles of

Colombia: an annotated analysis of their diversity, distribution, and conservation status. *Amphibian and Reptile Conservation*, 16(1), 106–135.

Pérez-Escobar, O. A., Antonelli, A., and Bogarín, D. (2019). Pleistocene refugia and orchid diversification in northwestern South America. *Molecular Ecology*, 28(16), 3774–3790.

Sedano-Cruz, J. M., Pérez-Escobar, O. A., and Vargas-Ramírez, M. (2024). Landscape drivers of diversification in northwestern South America. *Ecology and Evolution*, 14(3), e12345.

Stanford, C. B., Iverson, J. B., Rhodin, A. G. J., van Dijk, P. P., Mittermeier, R. A., Kuchling, G., ... and Walde, A. D. (2020). Turtles and tortoises are in trouble. *Current Biology*, 30(12), R721–R735.

TTWG [Turtle Taxonomy Working Group: Rhodin, A. G. J., Iverson, J. B., Fritz, U., Gallego-García, N., Georges, A., Shaffer, H. B., and van Dijk, P. P.] (2025). *Turtles of the World: Annotated Checklist and Atlas of Taxonomy, Synonymy, Distribution, and Conservation Status (10th Ed.)*. Chelonian Research Monographs, 10, 1–575.

Tailor-Rengifo, M., and Rentería-Moreno, A. (2011). Diversidad de reptiles en el Chocó. In: Rangel-Ch, J. O. (ed.) *Colombia diversidad biótica XI*. Instituto de Ciencias Naturales, Bogotá.

A. Anexo A (Capítulo 1)

Long-range PCR and primer walking. Two long-range PCR reactions were performed using DNA extractions from three *Trachemys medemi* samples to target the mitogenome. The resulting amplicons overlap by more than 1000 bp and have a length of 11824 bp and 6797 bp, respectively. Following the methodology of Fritz et al. (2012, 2023), the following primer pairs were employed for the long-range PCR reactions: Tscripta_tPhe.for + Tscripta_tLeu_a.rev and Tscripta_tArg_b.for + DES2_rev. Each reaction was performed in a 50 µl volume containing 1 unit of TaKaRa LA Taq DNA polymerase, Hot-Start Version (Clontech Laboratories Inc.), 10 mM of dNTPs, 5 pmol of each primer, and volumes according to manufacturer's recommendation. PCR conditions included an initial denaturation at 93°C for 3 min, 35 cycles of 93°C for 20 s, 57°C for 30s, 68°C for 12 min, and a final elongation step at 68°C for 20 min. PCR products were visualized in a 2% agarose gel and purified using the ExoSAP-IT enzymatic clean-up (USB Europe GmbH, Staufen, Germany). Cleaned long-range PCR products were then used for PCR sequencing reactions. Initially, amplification and sequencing were performed using primers specified in the Supplementary Table 1. Subsequently, specific primers designed from the *T. medemi* sequences were used to fill the remaining gaps in the assembly. The first part of the tRNA-Phe and the final of control region's 3'-end were amplified using the primers medemitRNA-Phe (5'-TGGGTAACCAATATACCCCA-3'), medemitRNA-Phe2 (5'-TTAGATTGCTAGGGCGTTTT-3'), and medemiCRend (5'-ACCCACGACAGTAATTTTCA-3') under the following thermocycling conditions: Initial denaturation at 94°C for 5 min, followed by 30 cycles of 94°C for 30s, 60°C for 45s, 72°C for 30 s, and final elongation at 72°C for 10 min. Sequencing was performed on an ABI 3730 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) following manufacturer's instructions.

Phylogenetic analyses. Phylogenetic analyses were conducted using two datasets. The first dataset included *Trachemys* mitogenomes obtained from Russel and Beckenbach (2008), Yu et al. (2014), Park et al. (2021), Ryu et al. (2021), and Fritz et al. (2023, 2024), alongside with mitogenomes from *Pseudemys concinna* (OM935747; Park et al. 2022) and *Graptemys ouachitensis* (OP115973) as outgroups. Protein-coding genes were screened for internal stop codons in Geneious R7 (<http://geneious.com>). Mitogenome alignment was performed with MUSCLE (Edgar 2004) and manually curated to correct alignment errors. Problematic sequence features, such as internal stop codons in coding genes, genes or tRNAs overlapping regions, frame shift inducing in coding region, and non-coding spacer DNA were identified and 232 positions were removed from the alignment. For the second dataset, our dataset was concatenated with the mtDNA alignment used by Fritz et al. (2024) covering almost all continental *Trachemys* taxa, except for *T. hartwegi*. Their dataset comprised 3221 bp and included sequences of the partial 12S gene and the complete ND4L, ND4, and *cyt b* genes plus part of the adjacent tRNA-Thr gene. Missing data was coded as Ns. The best partitioning schemes and evolutionary models (Supplementary Tables S4, S5) were determined by using a greedy search and the Bayesian Information Criterion in PartitionFinder 2 (Lanfear et al. 2017). For the extended mtDNA dataset, PartitionFinder 2 detected a single partitioning scheme with HKY+G+I model as the best fit. This partition scheme was used for the complete mitogenome dataset.

Phylogeny was analyzed using maximum likelihood (ML) and Bayesian inference (BA). ML analyses were performed in IQ-Tree 1.6.12 (Nguyen et al. 2015) with 1000 non-parametric thorough bootstrap replicates to assess node support, plotted against the best tree obtained. Bayesian analyses were conducted in MrBayes 3.2.6 (Ronquist et al. 2012), with two simultaneous runs and four chains each, sampling every 1000th generation for 50 million generations. After a burn-in of 25%, convergence was assessed by verifying that the average standard deviation of split frequencies were below 0.01 and the effective sample size exceeded 200 for all parameters using Tracer 1.7.1 (Rambaut et al. 2018).

Supplementary Table A1. Primers used for PCR sequencing reactions. Position according to the mapping of each primer to the mitochondrial genome of *Trachemys scripta elegans* (MW019443).

Primer	Sequence (5'-3')	LR-PCR	Position
Tscripta_tPhe.for	AGCACGGCACTGAAGTTGCC	1	22-41
L25195_for	AAACTGGGATTAGATACCCCACTAT	1	506-530
sulcata_cytb-12S_Rev	AGGCAAGTCGTAACAAGG	1	1001-1018
16SA_For	CGCCTGTTTATCAAAAACAT	1	1985-2004
graeca_cytb_16S_Rev	AACCGTGCAAAGGTAGCGTA	1	2078-2097
16SB_Rev	ACGTGATCTGAGTTCAGACCGG	1	2570-2591
medemi1F*	AAATGCAAAAGGCCTAAACC	1	2760-2779
medemi5F*	ACTCTCAGCAGGATTCCTAT	1	3643-3662
medemi5R*	ACCAATTACCAACCCTAC	1	5022-5041
L-turtCOI_For	ACTCAGCCATCTTACCTGTGATT	1	5420-5443
medemi1R*	TTTGGTCGTCTCCTAAAAGG	1	5552-5571
Jerry_For	CAACAYTTATTTTGATTTTTGG	1	6130-6152
medemi2F*	TTTTAGGCTTCATCGTGTGA	1	6259-6278
H-turtCOIb_Rev	CGAGCCTATTTTACATCTGCAAC	1	6340-6362
medemi6F*	CATGTACAAACCCAAGAAAG	1	6966-6985
medemi2R*	TCTGAGCATTGTCCGTAATA	1	7671-7690
medemi6R*	TTTTACGGACAATGCTCAGA	1	7695-7714
nigraATP6-cytb_For	AAGCCCACAAAYCCTWGGAA	1	8070-8090
graeca_12S_ATP6_Rev	ACAGCCAATTTAACAGCCGG	1	8525-8545
medemi3F*	CAGCAGCAGCACTACTTATA	1	8764-8783
nebulosa_long_F_For	CCTAGAATGAGCAGAATCGA	1	9911-9931
Tscripta_tArg_b.for	CTTAGTTAATCGTGATTAAACTCCACGGC	1, 2	9963-9992
medemi3R*	TAGGTGAGTTTGGTGAACG	1, 2	10035-10054
L-ND4_For	GTAGAAGCCCAATCGCAG	1, 2	10951-10971
Tscripta_tHis_a.for	CATTAGACTGTGGCTCTAAAAATAGGAGTT	1, 2	11692-11721
Tscripta_tHis_b.rev	GTGGCTCTAAAAATAGGAGTTCAAACCTC	1, 2	11701-11730
medemi4F*	GCTCCCTCACTTTTAAAGGA	1, 2	11777-11796
LR1-Rev_Tscripta_tLeu_a.rev	CCATTGGTTTTAGAAACCATCCACCC	1, 2	11828-11854
H-Leu_Rev	TGGTGCAAATCCAAGTAAAAGTAAT	2	11855-11880
medemi7F*	CTGCCCTATCCCAAAATGAT	2	12747-12766
medemi7R*	TTAACAAACAACCCTCAACCA	2	13931-13950
nigra_ND6-12S_For	AACYAACATCCCRCCCAAAT	2	14199-14219
medemi4R*	TAGTCTTAGGAGCAGCTTCT	2	14227-14246
sulcata_cytb-12S_For	CCTACAAATCACCACCGGAA	2	14599-14618
sulcata_ATP6-cytb_Rev	AGCCCATATCACCCGAGATG	2	14671-14690
Tscripta_tThr_a.rev	AAAGCATTGTCTTGTAACCAAAGACTGA	2	15635-15664

*Primers newly designed for this study.

Supplementary Table A2. Annotation of the *Trachemys medemi* mitogenome (BTBC13207).

Name	Start	Stop	Length	Strand	Intergenic/overlapping nucleotides
tRNA-Phe	1	70	70	+	0
12S rRNA	71	1.039	969	+	0
tRNA-Val	1.040	1.110	71	+	0
16S rRNA	1.111	2.726	1.616	+	1
tRNA-Leu	2.728	2.802	75	+	0
ND1	2.803	3.773	971	+	0
tRNA-Ile	3.774	3.843	70	+	-1
tRNA-Gln	3.843	3.914	72	-	-2
tRNA-Met	3.914	3.982	69	+	-1
ND2	3.983	5.021	1.039	+	0
tRNA-Trp	5.022	5.096	75	+	1
tRNA-Ala	5.098	5.166	69	-	1
tRNA-Asn	5.168	5.240	73	-	1
tRNA-Cys	5.267	5.332	66	-	-5
tRNA-Tyr	5.333	5.403	71	-	1
COX1	5.405	6.948	1.544	+	-5
tRNA-Ser	6.944	7.015	72	-	2
tRNA-Asp	7.018	7.087	70	+	0
COX2	7.088	7.774	687	+	1
tRNA-Lys	7.776	7.847	72	+	-2
ATP8	7.850	8.038	189	+	-31
ATP6	8.008	8.691	684	+	-1
COX3	8.691	9.474	784	+	-1
tRNA-Gly	9.475	9.543	69	+	0
ND3	9.544	9.893	350	+	0
tRNA-Arg	9.894	9.963	70	+	0
ND4L	9.964	10.261	297	+	-7
ND4	10.255	11.635	1.381	+	0
tRNA-His	11.636	11.705	70	+	0
tRNA-Ser	11.706	11.771	66	+	0
tRNA-Leu	11.772	11.843	72	+	-2
ND5	11.842	13.671	1.830	+	128
ND6	13.800	14.324	525	-	0
tRNA-Glu	14.325	14.392	68	-	4
CYTB	14.397	15.536	1.140	+	3
tRNA-Thr	15.540	15.613	74	+	1
tRNA-Pro	15.615	15.684	70	-	2
D-Loop	15.687	>16711	>955	+	0

Supplementary Table A3. Composition and skewness of PCGs, tRNAs, rRNAs, and control region of several *Trachemys* species and the two outgroups.

Species	Accession no.	A (%)	C (%)	G (%)	T (%)	GC (%)	AT skew	GC skew
Complete mitogenome								
<i>Graptemys ouachitensis</i>	OP115973	34.1	26.0	13.1	26.8	38.6	0.11987	-0.32992
<i>Pseudemys concinna</i>	OM935747	34.4	26.1	13.0	26.6	38.6	0.12787	-0.33504
<i>Trachemys medemi</i>	#####	34.3	26.1	12.8	26.6	39.0	0.12274	-0.34190
<i>Trachemys scripta elegans</i>	KM216748	34.1	26.0	12.9	26.9	38.7	0.11803	-0.33676
<i>Trachemys scripta scripta</i>	KM216749	34.2	26.0	12.9	26.9	38.7	0.11948	-0.33676
<i>Trachemys scripta troostii</i>	MW122292	34.2	26.0	12.9	26.9	38.7	0.11948	-0.33676
<i>Trachemys decorata</i>	OX453475	34.1	26.2	12.9	26.7	38.6	0.12171	-0.34015
<i>Trachemys decussata angusta</i>	OX453476	34.3	26.2	12.7	26.8	38.5	0.12275	-0.34704
<i>Trachemys decussata decussata</i>	OX453477	34.3	26.2	12.8	26.7	38.5	0.12459	-0.34359
<i>Trachemys grayi emolli</i>	OX453490	34.3	26.1	12.8	26.8	38.9	0.12275	-0.34190
<i>Trachemys grayi panamensis</i>	OX453491	34.3	26.1	12.8	26.8	38.9	0.12275	-0.34190
<i>Trachemys stejnegeri vicina</i>	OX453495	34.2	26.2	12.9	26.8	38.6	0.12131	-0.34015
<i>Trachemys terrapen</i>	OX453498	34.3	26.2	12.8	26.6	38.5	0.12644	-0.34359
<i>Trachemys venusta venusta</i>	OX453499	34.0	26.0	12.9	26.8	38.5	0.11842	-0.33676
<i>Trachemys venusta taylori</i>	OZ038157	34.0	26.0	12.9	26.8	38.5	0.11842	-0.33676
<i>Trachemys venusta cataspila</i>	OZ038159	34.8	26.2	13.0	26.8	38.9	0.12987	-0.33673
<i>Trachemys venusta uhrigi</i>	OZ038162	34.1	26.1	12.9	26.9	38.8	0.11803	-0.33846
PCGs								
<i>Graptemys ouachitensis</i>	OP115973	33.2	27.6	11.8	27.3	39.4	0.09752	-0.40102
<i>Pseudemys concinna</i>	OM935747	33.6	27.7	11.7	27.0	39.4	0.10891	-0.40609
<i>Trachemys medemi</i>	#####	33.4	27.9	11.7	27.1	39.5	0.10413	-0.40909
<i>Trachemys scripta elegans</i>	KM216748	33.4	27.5	11.6	27.4	39.1	0.09868	-0.40665
<i>Trachemys scripta scripta</i>	KM216749	33.4	27.5	11.6	27.4	39.1	0.09868	-0.40665
<i>Trachemys scripta troostii</i>	MW122292	33.5	27.5	11.6	27.4	39.1	0.10016	-0.40665
<i>Trachemys decorata</i>	OX453475	33.4	27.8	11.6	27.3	39.3	0.10049	-0.41117
<i>Trachemys decussata angusta</i>	OX453476	33.4	27.9	11.5	27.2	39.3	0.10231	-0.41624
<i>Trachemys decussata decussata</i>	OX453477	33.4	27.8	11.6	27.2	39.3	0.10231	-0.41117
<i>Trachemys grayi emolli</i>	OX453490	33.6	27.8	11.5	27.2	39.2	0.10526	-0.41476
<i>Trachemys grayi panamensis</i>	OX453491	33.6	27.7	11.5	27.2	39.2	0.10526	-0.41327
<i>Trachemys stejnegeri vicina</i>	OX453495	33.3	27.7	11.6	27.3	39.3	0.09901	-0.40967
<i>Trachemys terrapen</i>	OX453498	33.4	27.8	11.6	27.3	39.3	0.10049	-0.41117
<i>Trachemys venusta venusta</i>	OX453499	33.2	27.7	11.8	27.3	39.4	0.09752	-0.40253
<i>Trachemys venusta taylori</i>	OZ038157	33.1	27.7	11.9	27.3	39.5	0.09603	-0.39899
<i>Trachemys venusta cataspila</i>	OZ038159	33.2	27.7	11.9	27.3	39.5	0.09752	-0.39899
<i>Trachemys venusta uhrigi</i>	OZ038162	33.2	27.7	11.8	27.3	39.4	0.09752	-0.40253
tRNAs								

<i>Graptemys ouachitensis</i>	OP115973	34.3	22.5	16.0	27.2	38.2	0.11545	-0.16883
<i>Pseudemys concinna</i>	OM935747	34.6	22.6	15.7	27.1	38.1	0.12156	-0.18016
<i>Trachemys medemi</i>	#####	34.8	22.4	15.8	26.9	38.1	0.12804	-0.17277
<i>Trachemys scripta elegans</i>	KM216748	34.9	22.5	15.6	27.0	37.9	0.12763	-0.1811
<i>Trachemys scripta scripta</i>	KM216749	34.9	22.5	15.6	27.0	37.9	0.12763	-0.1811
<i>Trachemys scripta troostii</i>	MW122292	34.7	22.6	15.7	27.1	38.1	0.12298	-0.18016
<i>Trachemys decorata</i>	OX453475	34.7	22.6	15.7	27.1	38.1	0.12298	-0.18016
<i>Trachemys decussata angusta</i>	OX453476	34.7	22.6	15.6	27.1	38.0	0.12298	-0.18325
<i>Trachemys decussata decussata</i>	OX453477	34.7	22.8	15.5	27.0	38.1	0.12480	-0.1906
<i>Trachemys grayi emolli</i>	OX453490	34.4	22.2	15.5	27.9	37.7	0.10433	-0.17772
<i>Trachemys grayi panamensis</i>	OX453491	34.3	22.3	15.6	27.9	37.8	0.10289	-0.17678
<i>Trachemys stejnegeri vicina</i>	OX453495	34.7	22.8	15.6	26.9	38.2	0.12662	-0.1875
<i>Trachemys terrapen</i>	OX453498	34.9	22.8	15.4	26.9	38.0	0.12945	-0.19372
<i>Trachemys venusta venusta</i>	OX453499	34.6	22.7	15.7	27.0	38.2	0.12338	-0.18229
<i>Trachemys venusta taylori</i>	OZ038157	34.6	22.7	15.7	27.0	38.2	0.12338	-0.18229
<i>Trachemys venusta cataspila</i>	OZ038159	34.7	22.7	15.7	27.0	38.1	0.12480	-0.18229
<i>Trachemys venusta uhrigi</i>	OZ038162	34.6	22.6	15.7	27	38.1	0.12338	-0.18016
rRNAs								
<i>Graptemys ouachitensis</i>	OP115973	38.7	22.5	16.7	22.1	38.8	0.27303	-0.14796
<i>Pseudemys concinna</i>	OM935747	38.8	22.7	16.5	21.9	39.0	0.27842	-0.15816
<i>Trachemys medemi</i>	#####	38.6	23.1	16.7	21.6	39.4	0.28239	-0.16080
<i>Trachemys scripta elegans</i>	KM216748	38.5	22.9	16.7	21.9	39.3	0.27483	-0.15657
<i>Trachemys scripta scripta</i>	KM216749	38.5	22.9	16.8	21.8	39.2	0.27695	-0.15365
<i>Trachemys scripta troostii</i>	MW122292	38.5	22.9	16.7	21.9	39.3	0.27483	-0.15657
<i>Trachemys decorata</i>	OX453475	38.7	23.3	16.7	21.3	39.7	0.29000	-0.16500
<i>Trachemys decussata angusta</i>	OX453476	38.7	23.3	16.5	21.5	39.5	0.28571	-0.17085
<i>Trachemys decussata decussata</i>	OX453477	38.7	23.1	16.6	21.6	39.4	0.28358	-0.16373
<i>Trachemys grayi emolli</i>	OX453490	38.9	22.9	16.3	21.9	39.3	0.27961	-0.16837
<i>Trachemys grayi panamensis</i>	OX453491	38.9	23	16.4	21.7	39.4	0.28383	-0.16751
<i>Trachemys stejnegeri vicina</i>	OX453495	38.8	23	16.7	21.5	39.4	0.28690	-0.15869
<i>Trachemys terrapen</i>	OX453498	38.8	23.2	16.6	21.3	39.5	0.29118	-0.16583
<i>Trachemys venusta venusta</i>	OX453499	38.6	22.9	16.5	22.0	39.0	0.27393	-0.16244
<i>Trachemys venusta taylori</i>	OZ038157	38.8	22.9	16.4	21.9	38.9	0.27842	-0.16539
<i>Trachemys venusta cataspila</i>	OZ038159	38.7	22.9	16.4	22.0	38.8	0.27512	-0.16539
<i>Trachemys venusta uhrigi</i>	OZ038162	38.7	22.9	16.4	22.0	38.9	0.27512	-0.16539
Control Region								
<i>Graptemys ouachitensis</i>	OP115973	31.5	21.9	13.3	33.4	31.2	-0.02930	-0.24432
<i>Pseudemys concinna</i>	OM935747	33.2	20.9	12.8	33.1	29.9	0.00151	-0.24036
<i>Trachemys medemi</i>	#####	33.7	20.7	12.0	33.5	32.7	0.00193	-0.26605
<i>Trachemys scripta elegans</i>	KM216748	30.2	22.1	14.0	33.7	34.9	-0.05480	-0.22438
<i>Trachemys scripta scripta</i>	KM216749	30.4	22.1	13.9	33.5	34.7	-0.04850	-0.22778

<i>Trachemys scripta troostii</i>	MW122292	30.4	22.2	14	33.4	34.9	-0.04700	-0.22652
<i>Trachemys decorata</i>	OX453475	30.3	21.1	13.8	34.8	30.5	-0.06910	-0.20917
<i>Trachemys decussata angusta</i>	OX453476	31.5	20.6	12.9	35.0	29.5	-0.05260	-0.22985
<i>Trachemys decussata decussata</i>	OX453477	31.5	21.2	13.1	34.2	30.0	-0.04110	-0.23615
<i>Trachemys grayi emolli</i>	OX453490	30.4	21.3	14.0	34.3	30.9	-0.06030	-0.2068
<i>Trachemys grayi panamensis</i>	OX453491	30.1	20.9	14.2	34.7	30.8	-0.07100	-0.19088
<i>Trachemys stejnegeri vicina</i>	OX453495	30.2	21.2	13.7	35.0	30.5	-0.07360	-0.21490
<i>Trachemys terrapen</i>	OX453498	32.5	21.1	12.5	33.8	33.7	-0.01960	-0.25595
<i>Trachemys venusta venusta</i>	OX453499	32.2	21.2	12.8	33.8	30.3	-0.02420	-0.24706
<i>Trachemys venusta taylori</i>	OZ038157	32.4	21.4	12.6	33.6	30.3	-0.01820	-0.25882
<i>Trachemys venusta cataspila</i>	OZ038159	31.6	22.2	12.9	33.3	34.2	-0.02620	-0.26496
<i>Trachemys venusta uhrigi</i>	OZ038162	31.8	22.2	12.7	33.3	34.1	-0.02300	-0.27221

Supplementary Table A4. Data blocks of the mtDNA alignment used for phylogenetic analyses.

tRNA-Phe	1-72
12S	73-1049
tRNA-Val	1050-1121
16S	1122-2757
tRNA-Leu	2758-2833
ND1_pos1	2834-3799\3
ND1_pos2	2835-3799\3
ND1_pos3	2836-3799\3
tRNA-Ile-Gln	3800-4006
ND2_pos1	4007-5044\3
ND2_pos2	4008-5044\3
ND2_pos3	4009-5044\3
tRNA-Trp-Tyr	5045-5400
CO1_pos1	5401-6942\3
CO1_pos2	5402-6942\3
CO1_pos3	5403-6942\3
tRNA-Ser-Asp	6943-7078
CO2_pos1	7079-7762\3
CO2_pos2	7080-7762\3
CO2_pos3	7081-7762\3
tRNA-Lys	7763-7836
ATP8_pos1	7837-7992\3
ATP8_pos2	7838-7992\3
ATP8_pos3	7839-7992\3
ATP6_pos1	7993-8640\3
ATP6_pos2	7994-8640\3
ATP6_pos3	7995-8640\3
CO3_pos1	8641-9420\3
CO3_pos2	8642-9420\3
CO3_pos3	8643-9420\3
tRNA-Gly	9421-9489
ND3_pos1	9490-9838\3
ND3_pos2	9491-9838\3
ND3_pos3	9492-9838\3
tRNA-Arg	9839-9908
ND4L_pos1	9909-10197\3
ND4L_pos2	9910-10197\3
ND4L_pos3	9911-10197\3
ND4_pos1	10198-11568\3
ND4_pos2	10199-11568\3
ND4_pos3	10200-11568\3
tRNA-His-Leu	11569-11776
ND5_pos1	11777-13615\3
ND5_pos2	11778-13615\3
ND5_pos3	11779-13615\3
ND6_pos1	13616-14137\3
ND6_pos2	13617-14137\3
ND6_pos3	13618-14137\3
tRNA-Glu	14138-14205
CYTB-pos1	14206-15342\3
CYTB-pos2	14207-15342\3
CYTB-pos3	14208-15342\3
tRNA-Thr-Pro	15343-15486
Control	15487-16562

Supplementary Table A5. Best evolutionary models and partitioning schemes determined by PartitionFinder 2 for the complete mitogenome dataset. Estimated with the greedy search scheme and the Bayesian Information Criterion.

Subset		
1	HKY+I+G	Control, tRNA-Glu, ND5_pos1, ND5_pos2, ND4L_pos1, ND2_pos2, ND4_pos1, ATP6_pos2, CYTB_pos2, CO1_pos2, ND1_pos2, CO3_pos1, tRNA-Lys, CO3_pos2, tRNA-Val, tRNA-Phe, tRNA-Leu, ND3_pos2, CO1_pos1, ND1_pos1, ND6_pos1, tRNA-Thr-Pro, ND4L_pos2, ATP8_pos1, ND3_pos1, CYTB_pos1, 16S, ATP6_pos1, ND4_pos2, ND2_pos1, ATP8_pos2, ATP8_pos3, CO2_pos2, tRNA-Arg, tRNA-Gly, tRNA-His-Leu, tRNA-Trp-Tyr, tRNA-Arg, tRNA-Ile-Gln, CO2_pos1, tRNA-Thr-Pro, 12S, ND6_pos2
2	GTR+G	CYTB_pos3, ND2_pos3, ND5_pos3, ND6_pos3, ND4L_pos3, CO1_pos3, ND3_pos3, ND1_pos3, ATP6_pos3, ND4_pos3, CO2_pos3, CO3_pos3