UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO

DIVERSIDAD Y ESTRUCTURA GENÉTICA DE LA DANTA CENTROAMERICANA (TAPIRUS BAIRDII) EN EL NOROESTE DE LA CORDILLERA DE TALAMANCA Y EN EL CORREDOR BIOLÓGICO TENORIO-MIRAVALLES, COSTA RICA

Tesis sometida a la consideración de la Comisión del Programa de Estudios de Posgrado en Biología para optar al grado y título de Maestría Académica en Biología con énfasis en Genética y Biología Molecular

JUAN CARLOS DELGADO CARAZO

Ciudad Universitaria Rodrigo Facio, Costa Rica

2023

Dedicatoria

A mis papás, mi hermano y mi abuela, por todo su apoyo y cariño en todo momento.

A Bronia, por sostenerme a lo largo de este proceso y compartirme su apoyo y amor siempre.

A los grupos de monitoreo e investigación de vida silvestre del Cerro de la Muerte, Herradura, San Jerónimo y Pejivalle: por abrirme la puerta de sus casas, defender día a día la vida silvestre y llenarme de motivación por mi trabajo.

Agradecimientos

A l Canje Deuda por Naturaleza E.E.U.U-CR y AAZV Wild Animal Health Fund, por proveer los fondos necesarios para la ejecución de este proyecto.

A Costa Rica Wildlife Foundation, Nai Conservation y ASANA, por facilitar la logística y ejecución de este proyecto.

Al SINAC, en especial a ACC, ACLAP y ACAT, por respaldar la ejecución de este proyecto.

A Kerstin Wilhelm y Prof. Dr. Simone Sommer, por recibirme en dos ocasiones en su laboratorio y por haber compartido conmigo su conocimiento y experiencia.

A Doña Hannia, por todo su apoyo y compresión durante mi proceso de maestría.

A Dr. Eric Fuchs, por las herramientas facilitadas para trabajar con genética de poblaciones.

A Prof. Dra. Rebeca Mora-Castro, por su apoyo a mi carrera profesional y por todos sus consejos.

A Dra. Andrea Chaves R. y Valeria Gallardo, por todo su apoyo en LABGENCON y en mi carrera profesional.

A Andrea Chaves H., por creer en mi trabajo y apoyar el proyecto “Jardineras del Bosque”.

A la Universidad de Georgia, en especial al DVM Jorge Rojas Jiménez y la Prof. Dra. Sonia Hernandez-Divers, por su apoyo y respaldo en Tenorio-Miravalles.

A José Ramírez y Cristina Aguilar, por su compañía en las giras de campo y su apoyo para la recolección de muestras.

A Prof. Dr. Federico Albertazzi, por su mentoría y asesoramiento en asuntos moleculares desde mis primeros años de carrera, así como por sus consejos.

A Dr. Gustavo Gutiérrez, por su confianza al recibirme en LABGENCON y por apoyarme de múltiples formas durante todo mi proceso de maestría.

A M.Sc. Esteban Brenes, por abrirme las puertas a Nai Conservation, por ser un mentor en conservación, por apoyar la ejecución de todas las etapas de este proyecto y por su paciencia y comprensión.

Hoja de Aprobación

"Esta tesis fue aceptada por la Comisión del Programa de Estudios de Posgrado en Biología de la Universidad de Costa Rica, como requisito parcial para optar por el grado y título de Maestría Académica en Biología con énfasis en Genética y Biología Molecular".

Dr. Rafael Murillo Muñoz
Representante de la Decana
Sistema Estudios de Posgrado

Dr. Gustavo Gutiérrez Espeleta
Profesor Guía

Dr. Federico Albertazzi Castro
Lector

M.Sc. Esteban Brenes Mora
Lector

Dr. Eric Fuchs Castillo
Director del Programa de Posgrado en Biología

Juan Carlos Delgado Carazo
Sustentante

Tabla de Contenido

Lista de Cuadros

Lista de Figuras

Introducción

La danta centroamericana y su estado de conservación

La danta centroamericana (*Tapirus bairdii*) es el mamífero terrestre más grande del neotrópico (Brooks et al 1997). Su ámbito de distribución actual se extiende desde sur de México hasta el noroeste de Colombia, con casos de extinciones locales importantes como en El Salvador, donde ya no es posible encontrarla (García et al. 2016). Las dantas son animales estrictamente herbívoros que pueden consumir hasta 40 kg de biomasa vegetal diarios de más de 100 especies de plantas distintas (Foerster & Vaughan 2002; Hibert et al. 2011).

Los tapires desempeñan un papel ecológico relevante como dispersores de semillas (Janzen 1982, Naranjo-Piñera 1995, O'Farril et al. 2012, O'Farrill *et al.* 2013, Paolucci *et al.* 2019) y como moduladores y redistribuidores de los ciclos de nutrientes en el suelo (Villar et al. 2021). Se ha sugerido que las dantas pueden contribuir a mitigar cambio climático porque se ha podido predecir mayor fijación de carbono y menor extinción local de especies de árboles grandes en bosques donde las dantas están presentes (O'Farrill *et al.* 2013, Bello *et al.* 2015). Por sus servicios ecosistémicos dentro de los bosques que habita, se les conoce como las jardineras de los bosques.

La danta centroamericana se encuentra seriamente amenazada. Sus principales amenazas son la fragmentación y pérdida del hábitat, la cacería actual e históricamente persistente y, más recientemente, los atropellos en carretera (García *et al.* 2016). Se estima que en la actualidad quedan menos de 4 500 individuos y se proyecta que el número disminuirá un 50% en los próximos 25 años en consecuencia del continuo cambio en el uso de la tierra, principalmente (Schank *et al.* 2017). Por esto, la danta se encuentra en peligro de extinción según la lista roja de la Unión Internacional para la Conservación de la Naturaleza (UICN), en todo su ámbito de distribución actual.

Tapirus bairdii aparece también en la lista prioritaria de animales Evolutivamente Distintivos y Globalmente Amenazados (EDGE; Isaac *et al.* 2007) y

se encuentra en el apéndice I de la Convención sobre el Tráfico Internacional de Especies de Flora y Fauna Amenazadas (CITES), que es la categoría máxima de protección que concede la convención contra el tráfico internacional de especies. En el caso particular de Costa Rica, la danta se incluye en la lista de especies en peligro y con poblaciones reducidas de la Ley de Conservación de la Vida Silvestre (LCVS) y se contempla dentro de los planes de manejo de distintos parques nacionales como Parque Nacional Chirripó, Parque Nacional Tapantí-Macizo de la Muerte y Parque Nacional Volcán Tenorio.

Genética y su relevancia en la conservación de las dantas

La reducción en el tamaño de las poblaciones de dantas como consecuencia de sus amenazas no solo es relevante porque incrementa las probabilidades de extinciones locales y totales, sino también porque incrementa la vulnerabilidad de las mismas a la pérdida de diversidad genética por diferentes factores como la deriva génica y el incremento de endogamia (Harrison & Hastings 1996, Keller & Waller 2002, O'Grady *et al.* 2006). La diversidad genética es relevante para la biología de la conservación puesto que la pérdida de variabilidad puede incidir negativamente sobre la supervivencia, la resistencia a patógenos y enfermedades, las tasas de crecimiento y los rendimientos reproductivos de los individuos (Lacy 1987, Lande 1998, Reed & Frankham 2003).

Por otra parte, la endogamia -entendida como la generación de descendencia a partir de progenitores consanguíneos-, generalmente disminuye el valor adaptativo de los individuos e incrementa los riesgos de extinción de las poblaciones al afectar la supervivencia, depredación, la resistencia a enfermedades y el rendimiento reproductivo de los individuos, entre otros (Frankham 1998, Keller & Waller 2002). A su vez, esta pérdida de diversidad también puede incidir sobre la capacidad de las poblaciones para adaptarse a cambios ambientales a corto y largo plazo, bajo diferentes escenarios evolutivos (Crandall *et al.* 2000, Sommer 2005).

La genética es una herramienta que ha demostrado ser muy poderosa e informativa para el monitoreo de especies silvestres (Noss 1990, Schwartz *et al.*

2007), la indagación de las dinámicas poblacionales (Allendorf & Luikart 2009, De Barba 2010) y el desarrollo de estrategias de conservación de especies amenazadas (Allendorf & Luikart 2009, De Barba 2010, Frankham 2010). Además, el uso de la genética a partir de muestras no invasivas –principalmente pelos y heces– ha permitido la detección y monitoreo de poblaciones silvestres de animales amenazados, que de otra forma serían muy difíciles de estudiar por sus bajas densidades y hábitos elusivos, entre otros (Waits & Paetkau 2005).

La obtención de resultados genéticos robustos y congruentes a partir de muestras no invasivas ha sido validada para un gran número de taxones, incluyendo la danta sudamericana (Pinho *et al.* 2014) y la danta centroamericana (Zavala-Páramo *et al.* 2017). Además, es una herramienta con alto poder informativo que puede permitir detectar sesgos en estudios de poblaciones silvestres de animales cuando se utilizan otras herramientas de monitoreo. Por ejemplo, en un estudio genético reciente a partir de las heces de leopardos de las nieves fue posible determinar la sobre-estimación de la densidad de individuos en las poblaciones de estos animales, al utilizar únicamente cámaras trampa (Chetri *et al.* 2019).

La principal desventaja con las muestras no invasivas es que luego del proceso de extracción se suele obtener baja concentración de ADN, baja calidad de ADN e incluso presencia de ADN foráneo (proveniente de la dieta y de microorganismos), por lo que el genotipado a partir de este tipo de muestras es susceptible a errores (Broquet *et al.* 2007). Aun así, existen diferentes estrategias para minimizar los posibles errores que se presenten (Pompanon *et al.* 2005). Por ejemplo, con el enfoque multi-tubo (Taberlet *et al.* 1997) se determinan genotipos a partir de la observación reiterada de alelos específicos al realizar múltiples amplificaciones. Otras aproximaciones permiten detectar posibles errores de genotipeo *a posteriori* (e.g. alelos nulos), a partir de las frecuencias alélicas que se obtengan para un marcador (van Oosterhout *et al.* 2004).

El estudio genético de la danta centroamericana fue declarado prioridad en el plan de acción del grupo de especialistas de tapires (TSG) de la IUCN del año 2005 (Medici *et al.* 2005). Pese a esto, solo cinco investigaciones han explorado la genética de *T. bairdii* a la fecha (Norton & Ashley 2004a, Norton & Ashley 2004b,

Ruiz-García *et al.* 2012, McCann 2015, Zavala-Páramo *et al.* 2017). De estas investigaciones, solo dos incluyeron dantas silvestres de Costa Rica (Norton & Ashley 2004a; Ruiz-García *et al.* 2012). Sin embargo cada una tomó en cuenta un número muy reducido de animales: 15 y 9 individuos respectivamente..

Por estructura genética poblacional se entiende el grado de división de las poblaciones en subpoblaciones o unidades reproductivas locales, que se encuentran diferenciadas y separadas unas de otras (Allendorf & Luikart 2009). Comprender si existe estructura genética es relevante en el manejo y conservación de vida silvestre porque cada población -definida por criterio de diferenciación genética- sufre sus propios procesos adaptativos, que no pueden ser inferidos del estudio de otras poblaciones ni de otras especies (Moritz 1994). Asimismo, los análisis de estructura genética de las poblaciones silvestres pueden brindar evidencia de la existencia o carencia de flujo genético entre diferentes grupos (Pritchard *et al.* 2000). Mantener el flujo genético entre poblaciones es especialmente importante en nuestro contexto puesto que Costa Rica es parte de una eco-región particularmente susceptible a la pérdida de biodiversidad funcional, dadas sus especies en peligro de extinción y las amenazas antropológicas (González-Maya *et al.* 2017).

Esta investigación busca atender un vacío importante en la literatura sobre los tapires centroamericanos, puesto que el estudio genético de la especie ha sido desatendido, aun cuando desde el 2005 el grupo de especialistas en tapires (TSG) de la IUCN declaró prioritaria esta línea de estudio. En un primer paso, se establecerá un panel de marcadores moleculares (microsatélites) aptos –es decir, que permitan resultados reproducibles y congruentes– para la amplificación de ADN obtenido a partir de muestras no invasivas de animales silvestres encontrados en condiciones naturales. Este panel, después de su adecuada validación, representará una herramienta importante para el estudio de las poblaciones silvestres de la danta centroamericana.

A partir del análisis de la información genética encontrada, se determinará la diversidad y estructura genética de las poblaciones de estos animales en el noroeste de la Cordillera de Talamanca y en Corredor Biológico Tenorio-Miravalles, regiones

de gran importancia en términos de biodiversidad para el país y para el mantenimiento de las poblaciones de estos animales, pero que han sido poco estudiadas, con el fin de describir los patrones moleculares de las mismas. Asimismo, se indagará la existencia de procesos endogámicos, que puedan tener repercusiones en la remanencia de las poblaciones a corto y largo plazo.

Artículo Científico

GENETIC DIVERSITY & STRUCTURE OF THE CENTRAL AMERICAN TAPIR (TAPIRUS BAIRDII) IN THE NORTHWEST OF THE TALAMANCA MOUNTAIN RANGE AND THE TENORIO-MIRAVALLES BIOLOGICAL CORRIDOR, COSTA RICA

Abstract

Tapirus bairdii is the largest terrestrial herbivore living in Mesoamerica, distributed from southern Mexico to northwestern Colombia. Tapirs play relevant ecological roles as seed dispersers and nutrient cycling redistributors and modulators, among other ecosystem services provided. However, *T. bairdii* is endangered throughout its current range. The genetic study of the Central American tapir was declared a priority in the 2005 Tapir Specialist Group action plan, but few investigations exploring the genetics of *T. bairdii* have been published to date. We used invasive (blood, tissue, pulled out hairs) and non-invasive (feces, fallen hairs) samples and a set of seven microsatellite loci to describe the genetic diversity and structure of three a priori defined populations: Cerro de la Muerte (CM), Herradura-San Jerónimo (H-SJ) and Tenorio-Miravalles (TM) of wild tapirs from two high-biodiversity regions in Costa Rica: the Northwest of Talamanca Mountain Range (NWT) and Tenorio-Miravalles Biological Corridor (TMBC). We confirmed the presence of tapir-specific DNA in the collected samples by sequencing a fragment of 114-148 bp of *cytb* gene. We found moderate levels of genetic diversity (Ho=0.52) when analyzing the whole Costa Rican metapopulation. Our estimates of diversity by population ranged in the middle of previous reports from other populations of the species (Ho= 0.46-0.58). We found significant genetic structure among NWT and TMBC region (G'_{ST} Nei=0.0587; $p<0.001$) and among the three populations using F_{ST} and G-Statistics. AMOVA analysis found significant genetic structure among NWT and TMBC regions ($\phi CT=0.0602$; $p<0.001$) and marginal genetic structure among populations ($\phi SC=0.0479$; $p<0.05$). Bayesian analysis of genetic clusters (K) grouped individuals into K=2, suggesting a similar configuration of individuals as the regional stratification did. The weak genetic structure between members of H-SJ and TM, plus the clear differentiation of CM and TM, suggest a strong putative pathway of gene flow from TM to H-SJ through the Atlantic slope, its protected areas and connectivity routes, whilst a weak putative pathway of gene flow from TM to CM through the Pacific slope is suggested. We recommend the inclusion of non-invasive samples in more population genetics studies, especially through the collaboration with local communities coexisting with wildlife as a strategy to offset the generalized lack of information about neotropical wildlife species. Using this approach, we achieved to collect, process, and analyze the highest number of samples reported to date for any American tapir species using microsatellite markers.

Keywords: Mesoamerican tapir, genetic connectivity, local community, population genetics, cytb, molecular identification, non-invasive samples, invasive samples.

Introduction

Tapirs are an ancient group of mammals that diverged from their closest living relatives (rhinoceroses and equids) during the early Eocene approximately 55 Mya (Ruiz-Garcia et al. 2012). The fossils of more than nine different genera of the

Tapiridae family are recognized from North America, Europe, and Asia, evidencing a wider distribution of the group in the past (Hulbert 2005; Deng 2006; Colbert 2007; Ruiz-García et al. 2012). The current Tapiridae family is monotypic: *Tapirus* is the only surviving genus comprised of four extant species: *T. bairdii*, *T. terrestris*, *T. pinchaque*, and *T. indicus* (Medici 2011). Three species (*T. bairdii*, *T. pinchaque* and *T. terrestris*) are found in the Neotropics of Central and South America, while *T. indicus* inhabits the tropical forests of Southeast Asia (García et al. 2016; Lizcano et al. 2016; Varela et al. 2019; Traeholt et al. 2016). Compared to other Neotropical mammals, tapirs have been little studied in the wild, largely because of their low densities and their predominantly crepuscular/nocturnal behavior (Naranjo 2009; García et al. 2012).

Diverse ecological studies demonstrate that tapirs play a relevant ecological roles as highly efficient seed dispersers (Janzen 1982; Naranjo-Piñera 1995; O'Farril et al. 2012; O'Farrill et al. 2013; Paolucci et al. 2019) and nutrient cycling redistributors and modulators (Villar et al. 2021). They may also contribute to mitigating climate change because the habitats in which they live have been observed to have higher carbon fixation rates and lower extinction of large tree species (O'Farrill et al. 2013; Bello et al. 2015). Because of their ecosystem services within the forests they inhabit, they are known as the gardeners of the forests. Despite their ecological relevance, all tapir species are seriously threated (García et al. 2016; Lizcano et al. 2016; Varela et al. 2019; Traeholt et al. 2016).

The Central American tapir or Baird's tapir (*T. bairdii)* is the largest terrestrial herbivore living in Mesoamerica (Brooks et al 1997). Its current range of distribution extends from southern Mexico to northwestern Colombia, with cases of significant local extinctions such as in El Salvador, where it is no longer possible to find it (Schank et al. 2020). They can consume up to 40 kg of plant biomass per day from more than 100 different plant species (Foerster & Vaughan 2002; Hibert et al. 2011). *T. bairdii* is highly threatened by habitat fragmentation and loss of habitat, current and historically persistent hunting and, more recently, roadkills (García et al. 2016). It is estimated that less than 4 500 individuals remain, and the number is projected to decrease by 50% in the next 25 years because of continued land use change, among other factors (Schank et al. 2017). Because of this, the tapir is endangered throughout its current range, according to the red list of the International Union for Conservation of Nature IUCN (García et al. 2016).

T*. bairdii* is also included in the priority list of Evolutionarily Distinct and Globally Threatened Animals (EDGE; Isaac et al. 2007) and is included in Appendix I of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), which is the highest category of protection granted by the convention against international species trafficking. In the case of Costa Rica, the tapir is included in the list of endangered species and species with reduced populations of the Wildlife Conservation Law (LCVS) and is included in the management plans of

different national parks such as Chirripó National Park, Tapantí-Macizo de la Muerte National Park, Volcán Tenorio National Park, Miravalles Jorge Manuel Dengo National Park and Corcovado National Park.

Genetics and genomics have proven to be very powerful and informative tools for monitoring wild species, investigating population dynamics, and developing conservation strategies for threatened species (Noss 1990; Schwartz et al. 2007; ; Frankham 2010; De Barba 2010; Allendorf et al. 2022). In addition, the use of genetics from non-invasive samples -mainly hairs and feces- has allowed the detection and monitoring of wild populations of threatened animals, which would otherwise be very difficult to study due to their low densities and elusive habits (Waits & Paetkau 2005).

The genetic study of the Central American tapir was declared a priority in the 2005 IUCN Tapir Specialist Group (TSG) action plan (Medici et al. 2005). At the same time, the implementation of recovery strategies like relocation, reintroduction and *ex situ* management requires data on tapir's genetic variability (Ruiz-García et al. 2012). However, few investigations exploring the genetics of *T. bairdii* have been published to date (Norton & Ashley 2004a; Norton & Ashley 2004b; Ruiz-García et al. 2012; McCann 2015; Zavala-Páramo et al. 2017). Two studies used mitochondrial DNA to assess comparative phylogenetics, demographic history, and genetic variability of tapirs (Ruiz-García et al. 2012; McCann 2015). Two studies included wild tapirs from Costa Rica (Norton & Ashley 2004a; Ruiz-García et al. 2012). However, each considered one Costa Rican population comprised of a small sample size: 15 and 9, respectively.

This research seeks to address an important gap in the knowledge on Central American tapirs, since little information about the wild population genetic diversity and structure of the species exists. In a first step, a panel of molecular markers (microsatellites) was established and validated for the amplification of DNA obtained from invasive (blood, tissue) and non-invasive (feces, hairs) samples from wild animals found in four different populations (defined *a priori*) of Costa Rica. Based on the analysis of the genetic information found, the genetic diversity and structure of the populations of tapirs was assessed and described. Likewise, the existence of endogamic processes that may have repercussions on the remanence of the populations in the short and long term will be investigated.

Materials & Methods

Study Region: The Northwest of the Talamanca Mountain Range (NWT)

The Talamanca Mountain Range is the largest continuous forested area in Costa Rica, comprised of protected areas as well as inhabited areas such as forest reserves, productive zones, and urban areas, all within two Biospheres Reserves:

La Amistad (UNESCO n.d.a) and Savegre (UNESCO n.d.b). The Talamanca Mountain Range comprise almost 25% of the country and exhibits a pronounced altitudinal gradient, starting at 80 meters above sea level and reaching the highest point in Costa Rica at 3820 meters above sea level. Furthermore, this area is influenced by both Caribbean and Pacific slopes, has eight life zones (Janzen 1991), and its geological origin gives it a high rate of endemism, making it one of the most biodiverse hotspots in the world (Kapelle et al. 2000, Barrantes 2009, UNESCO n.d.a).

The Talamanca Mountain Range represents one of the most important sites for the conservation of the Central American tapir in its distribution range (González-Maya et al. 2012, Schank et al. 2017, 2020). In the northwestern sector of the Talamanca Mountain Range (NWT) there are protected areas and forest remnants that are part of several of the most important and extensive biological connectivity networks in Costa Rica (Arias et al. 2008). These sites are key to guarantee the movement of fauna, to ensure the genetic connectivity of the populations between slopes and to maintain viable populations in the long term that can ensure the regeneration and resilience of the forests (Arias et al. 2008; Schank et al. 2017).

Sample collection was carried out during 2019-2021 within protected areas (Tapantí-Macizo de la Muerte National Park, Los Quetzales National Park and Río Macho Forest Reserve) and private properties outside the protected areas, where evidence of activity of tapirs has been recorded (direct sightings, traces, feces, or camera trap photographs). Three regions of influence were considered inside NWT region, namely: i) Cerro de la Muerte (CM), ii) Herradura de Rivas, Pérez Zeledón (H) and iii) San Jerónimo, Pérez Zeledón (SJ), with different sample collection sites in each one (Fig. 1). CM, H and SJ land comprise forests, pastures for livestock production, agroforestry systems and human settlings, mostly rural communities, and touristic infrastructure.

These three groups were considered as populations *a priori*, an hypothesis later tested, because of two main factors. First, a considerable geographic distance exists among the center of this regions of interest (min. 15 km among them; Fig. 1). Second, we identified several communities inhabiting each region of influence. Community members interested to participate in this project attended several workshops on environmental education, camera-trapping and biological sample collection before establishing wildlife monitoring and research groups (WMRGs) for each CM, H and SJ regions. Each WMRG participated actively in the selection of sites for the collection of biological samples in their own communities. Thus, each population defined *a priori* is associated with a different human community and a unique WMRG that helped to locate and collect the biological samples. WMRGs members included natural guides, students, reformed hunters and other backgrounds not necessarily directly related to natural resources.

Study Region: Tenorio-Miravalles Biological Corridor (TMBC)

The Tenorio-Miravalles Biological Corridor (TMBC) is part of the Guanacaste Mountain Range in northwestern Costa Rica. A biological corridor is defined by Costa Rican environmental authorities (MINAE) as a delimited continental, marine-coastal and insular territory, whose primary purpose is to provide connectivity between protected wild areas, as well as between natural or modified landscapes, ecosystems and habitats; to ensure the maintenance of biodiversity and ecological and evolutionary processes (Regulación del Programa Nacional de Corredores Biológicos 2016).

TMBC covers 12.696,1662 ha of extension and is located between Miravalles Jorge Manuel Dengo National Park and Volcán Tenorio National Park, including part of the cantons of Upala (Alajuela), Bagaces and Cañas (Guanacaste). TMBC is part of the Biosphere Reserve Agua y Paz (UNESCO n.d.c) and 91% of its territory is found between 400-700 meters above sea level, with an annual rainfall between 2500-4500 mm and an average annual temperature between 22-26°C (Pastor-Parajeles 2022).

Most of the corridor is covered by forests (52%), pastures for livestock production (20%), agroforestry systems with low tree density (11%), and rural or urban settlements (5%) (Baltodano & Zamora 2009; Bautista-Solís et al. 2012). CBTM was founded in 2001 as a link area between protected areas in the north and center of Costa Rica and between the Pacific and Atlantic slopes (Bautista-Solís et al. 2012). In addition, the ecosystems found at TMBC are suitable for large species in danger of extinction such as tapirs. The existence of primary & young secondary forests (tacotales), the availability of endemic flora on which tapirs feed (e.g. *Parmentiera valerii*) and the proximity to abundant bodies of water of good quality have been identified as significant variables explaining the habitat occupation of tapirs in TMBC (Carbonell & González 2000).

Sample collection was carried out during 2022-2023 within protected areas (Volcán Tenorio National Park & Miravalles Jorge Manuel Dengo National Park) and private properties outside the protected areas, where evidence of activity of these animals has been recorded (direct sightings, traces, feces, or camera trap photographs). Only one region of influence was considered in TMBC because of the proximity of sample collection sites (Fig. 2). This population was named Tenorio-Miravalles (TM).

Collection of Biological Samples

A dual invasive and non-invasive sampling collection approach was developed to collect a total of 104 samples. High quality samples obtained correspond first to tissue samples (N=4) and hairs (N=2) retrieved from deceased animals in road kills

at the Interamerican Highway in NWT region during 2017-2022. These samples were preserved in 90% ethanol at -20 C until DNA extraction. Second, blood samples (N=5) and hairs (N=4) were collected from immobilized individuals (Fig. S2D) during 2022-2023 as part of an ongoing project describing the movement patterns and resource selection of tapirs in the TMBC region by capturing and outfitting radio collars to wandering individuals. Approximately 1 mL of full blood was drawn into a tube containing either EDTA or RNAlater (Thermo Fisher Scientific) as preservatives. Samples were preserved at -80 C prior to DNA extraction.

Most of non-invasive samples collected correspond to feces (N=88, Fig. S2A), but fallen hairs were also retrieved once (N=1; Fig. S2B). Inspections of latrines (sites of excrement depositions), transit areas and bodies of water were carried out in search of fecal samples. The collection of feces in the NWT region involved the active participation of WMRGs composed of local communities' members coexisting with wildlife (Fig. S2C). The groups received theoretical and practical training to adequately collect non-invasive samples and were supervised by researchers during the field trips. Approximately 70% of samples from NWT were collected by researchers with previous experience and the rest 30% by WMRGs members. In TMBC, the collection of feces was performed by field technicians and researchers with experience in sample collection.

Tapir stools were visually identified and subsampled in several tubes. A small section (approx. ~2.5 gram) from the outside layer of the stool (a section enriched by mammal epithelial cells from intestinal walls) was placed in a 4 mL sterile cryovial previously filled with 2.5 mL of RNAlater. 2-4 replicates were collected from each identified independent deposition and immediately shacked to homogenize the fecal material with the preservative. The separation and collection of fecal subsamples in the field is recommended since a better performance extracting host DNA is obtained and less PCR inhibition results when subsampling *in situ* (Ramón-Laca *et al.* 2015).

The organoleptic characteristics (*e.g.* color, scent, presence of flies, presence of fungi) of the samples were observed and recorded in a field book as indirect indicators of the deposition time of feces in the environment. Most of the samples were georeferenced in the field using DMS, DM or UTM coordinate systems and then homogenized to DM to project maps of samples collected. The samples were transported approximately one week after collection to the laboratory, where stored at -80C until further processing.

Extraction & Quantification of DNA

Scrapings from the outer layer of the stool (subsamples) stored in RNAlater were processed for DNA isolation (N=88). Approx. 220 mg of starting material were taken to obtain DNA using the QIAamp DNA Stool Mini Kit (Qiagen). DNA was

extracted from feces according to the manufacturer's instructions initially (protocol A), but two major modifications were applied to optimize yield performance (both implemented in protocol B & C). First, the reaction mixtures tubes were incubated for digestion with lysis buffer at 55°C and low agitation for 24-48 h. Secondly, a two-step elution was performed (1st round: 80 ul of Elution Buffer, 2nd round: 50 ul E.B.) for each sample, with both elution including a 20 min incubation at 40°C to maximize DNA recovery. The samples were extracted in batches of 6-10 tubes at a time, most frequently 6 samples per batch.

200 ul of blood (N=5) were used as starting material to obtain DNA using the PureLink Genomic DNA Mini Kit (Thermo Fisher Scientific). The extraction process was identical to the manufacturer's instructions, but prior to the final elution the reaction mixtures were incubated at 40°C for 20 min to maximize DNA recovery. Approx. 25 mg of connective tissue (ear, N=4) was used to isolate DNA. The protocol followed manufacturer instructions with incubation of sample for digestion at 55°C for 16-24 hours. 20-30 hairs (including follicle) per sample (N=7) were pooled together to isolate DNA. The PureLink Genomic DNA Mini kit (Thermo Fisher Scientific) was used with three relevant modifications of the tissue protocol: 1) the amount of genomic digestion buffer and proteinase K was doubled to 360 and 40 ul respectively, 2) the reaction mixtures tubes were incubated for digestion at 55°C with low agitation for 24 h and 3) prior to the final elution the reaction mixtures were incubated at 40°C for 20 min to maximize DNA recovery. DNA quantity and quality was estimated from 2 ul aliquots analyzed on a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific). DNA integrity was checked running 1.5% agarose gel electrophoresis.

Validation of the Dataset

To confirm the presence of tapir-specific and viable DNA in the extracted samples, a partial fragment (approx. 150 bp) of mitochondrial *cytb* gene was amplified and sequenced in a random subset of samples (N=31) using the primers L15601 5'-TACGCAATCCTACGATCAATTCC-3' & H15748 5'-GGTTGTCCTCCAATTCA TGTTAG-3' (Lopez-Oceja et al. 2016). 12.50 ul of AmpliTaq 360 Master Mix (Thermo Fisher Scientific), 7.65-9.90 ul of nuclease-free water, 0.80 ul of each primer (5uM) and 6.17-126.60 ng (1-3.25 ul) of total DNA were mixed in a final volume of 25 ul. PCR was performed in a Veriti 96-Well Thermal Cycler (Applied Biosystems), following an initial 10 min denaturation at 94°C, 12 cycles of 30 s at 94°C, 30 s at 49-47°C (ΔT -0.2°C per cycle after 2nd round) and 45 s at 72°C, followed by 28 cycles of 30 s at 94 °C, 30 s at 46.5 °C, 45 s at 72°C and a final extension of 5 min at 72 °C. PCR products and their corresponding negative control were visualized in UV light after electrophoresis in 1.5% agarose gel. PCR-positive samples were purified and Sanger method

sequenced in both directions by Macrogen Inc, South Korea.

Sequences obtained were automatically trimmed at both ends (-25 bp each) using Geneious Prime 2022.2.2. Trimmed ends were extended (up to 60 bp) if poor quality values (HQ%< 70) were observed. Pairwise alignments -Global alignment with free end gaps, 93% similarity, gap open penalty of 13 and gap extension penalty of 3- of both directions were generated for each sample. A consensus sequences was extracted from the alignment and used as query to assess the taxonomic assignment of each. When a sample produced only one high-quality strand (%HQ>90), the alignment and consensus steps were omitted, and further analysis continued using that sequence as the query.

The taxonomic assignment of the queries was predicted using BLAST. BLAST calculates alignment scores (S) based on the similarity between the query and the sequences deposited at NCBI dataset. BLAST employs local alignments, allowing to assign taxonomic identity based on a probabilistic criterion (Altschul et al. 1990; Johnson et al. 2008). E value, % pairwise identity, % query cover and % grade was calculated for all samples examined.

E value: the number of different alignments with scores equivalent to or better than S that is expected to occur in a database search by chance. The lower the E value, the more significant the score and the alignment. % Pairwise identity: the percentage of pairwise residues that are identical in the alignment, including gap vs. non-gap residues, but excluding gap vs. gap residues. % query cover: the percentage of the query sequence covered by the hit (internal gaps are counted as coverage). % grade: a weighted score for the hit comprised of E value, pairwise identity and the coverage.

Validation of the Panel of Molecular Markers

Previous studies validated the use of fecal samples for population genetics studies in tapirs, including the focal species *T. bairdii* (Pinho *et al.* 2014; Zavala-Páramo *et al.* 2017). A subset of DNA samples (N=64) including high-quality DNA samples were used in standardizing and optimizing the amplification and genotyping of 15 microsatellite loci used in this study (Table I). Six of the microsatellite loci were selected from a recent validation with captive *T. bairdii* individuals (Zavala-Páramo et al. 2017). Another six selected loci come from microsatellite design studies for *T. bairdii* & *T. terrestris*, without recent validation (Norton & Ashley 2004; da Silva et al. 2009; Sanches *et al.* 2009). Three additional polymorphic markers designed for other perissodactyls were tested for cross-species amplification and validation (Breen et al. 1997; Nielsen et al. 2008).

PCR mixtures included 7.50 ul of AmpliTaq 360 Master Mix (Thermo Fisher Scientific), 4.50-5.00 ul of nuclease-free water, 0.75 ul of each primer (5uM) and 4.00-186.00 ng of total DNA (1-1.5 ul) for a final volume of 15 ul. PCR was performed

in a Veriti 96-Well Thermal Cycler (Applied Biosystems), starting with an initial 8 min denaturation at 94°C; 40 cycles comprised of i) 30 s at 94°C, ii) 30 s at Annealing temperature (Ta; Table I, IV) and iii) 45 s at 72°C; followed by a final extension of 5 min at 72 °C. 640 PCR products (N=64, 10 loci) were visualized in an ABI-3500 Sanger sequencer (Applied Biosystems) at Escuela de Biología, Universidad de Costa Rica and genotyped using GENEMARKER version 2.4.2 (Hulce et al. 2011). The genotypes obtained from high quality samples allowed the validation of alleles and product size ranges observed in the non-invasive samples.

After the validation essay, two independent genotyping runs were conducted for all samples (N=98) using the partially validated markers at the time (N=8) and following a modification of the multi-tube approach (Taberlet et al. 1996, Pompanon *et al.* 2005). PCR products were genotyped using capillary electrophoresis services of Marcogen Inc. (Seoul, South Korea). The presence of genotyping errors, null alleles and its frequency were evaluated for each marker (N=8) using MICRO-CHECKER (van Oosterhout & Shipley 2004). *Pegas* package v.1.2. (Paradis 2010) for R v.4.2.2 (R Core Team 2017) was used to determine if the markers behave according to Hardy-Weinberg Equilibrium (HWE) model. Probability of HWE was estimated with a Monte Carlo procedure of 10^4 replicates at metapopulation and population levels.

Linkage Disequilibrium (LD) was assessed through the calculation of the standardized index of association *rbarD* of the validated markers (N=7) at the metapopulation level and for each population using *poppr* package v.2.9.3 (Kamvar et al. 2014) in R v.4.2.2. The null hypothesis for rbarD assumes that there is recombination among all the loci and between population. Statistical significance of LD was obtained by comparing LD values observed in the data against 10^4 simulated data sets (one-sided permutation test).

The power of the validated set of molecular markers (N=7) to discriminate genotypes was evaluated by calculating the probability of identity $P_{(ID)}$. $P_{(ID)}$ is defined as the probability of two unrelated individuals having the same genotype in a population (Waits et al. 2001). $P_{(ID)}$unbiased (after sample size corrections) values were obtained using GIMLET v.1.3.3 (Valière 2002). The R v.4.2.2 package *poppr* v.2.9.3 (Kamvar et al., 2014) was used to plot a genotype accumulation curve, which assesses the power of a random set of markers in discriminating between multi-loci genotypes (MLGs) and unique individuals. This was done by random sampling one to N-1 markers without replacement for 10^3 iterations and counting the number of MLGs observed.

Estimates of Genetic Diversity & Inbreeding

Estimates of genetic diversity were obtained from the genotypes of the validated microsatellites (N=7). Number of alleles, allelic richness, observed

heterozygosity (Ho) and expected heterozygosity (Hs) were calculated using the *hierfstat* v.05-11 (Goudet & Jombart 2015) and *adegenet* v.2.1.10 (Jombart 2008) packages in R 4.2.2. The F_{IS} index is defined as the coefficient of inbreeding of an individual with respect to its local subpopulation (Allendorf & Luikart 2009). The F_{IS} inbreeding index was calculated for each population, to estimate the existing inbreeding degree. The estimation was performed with the *poppr* package v.2.9.3 (Kamvar et al. 2014) in R v.4.2.2. All these estimates allow determining whether the levels of diversity exhibited in the studied populations are high or low, compared to other reported populations of tapirs.

Gene Flow & Genetic Structure Analysis

Pairwise F_{ST}, G'_{ST} Nei, G'_{ST} Hedrick & D_{EST} (and their respective probability after 10^3 iterations) were obtained to inspect genetic structure at both regional and population levels. The estimates were computed with GenAlEx v.6.503 (Peakall & Smouse 2012). To test the hypothesis that both regional and population levels have an important role in shaping population structure, an analysis of molecular variance (AMOVA) was performed using *poppr* package v.2.9.3 (Kamvar et al. 2014) in R 4.2.2. was used to compute the analysis, using 10^4 permutations to test the significance of phi-statistics.

A Bayesian analysis of genetic clusters (K) was conducted to confirm the results of the estimates of genetic flow and to determine whether there are differentiated populations due to population isolation. Sets of 10 independent runs per K (K= 1 to 11) were computed using an admixture model with correlated allele frequencies and $3x10^4$ MCMC iterations after a burn-in of $7.5x10^4$ replicates. The analysis was carried out with STRUCTURE v.2.3.4 (Pritchard et al. 2000). The number of most probable genetic clusters was identified according to the Evanno method using STRUCTURE HARVESTER (Earl 2012). Assignment bar graphs for individuals and populations were generated and customized using STRUCTURE PLOT v.2.0 (Ramasamy et al. 2014).

To complement the STRUCTURE analysis, a Discriminant Analysis of Principal Components (DAPC) was used (Jombart et al. 2010). DAPC follows a multivariate approach (free from Hardy–Weinberg and LD assumptions) to evaluate the population structure and grouping of the data. DAPC combines Principal Components Analysis (PCA), K-means clustering and discriminant analysis (DA) to detect and visualize population structure (Jombart et al. 2010). DAPC was computed using the *adegenet* package v.2.1.10 (Jombart 2008) in R v.4.2.2 and cluster assignments were pre-defined corresponding with assumed populations.

After identifying a genetic structure signal, a Mantel test that correlated a matrix of geographic distances between sites of collection (transformed into

Euclidean distance) and a matrix of genetic distances based on the Gst fixation index was computed using *vegan package 2.6.4* (Oksanen et al. 2007) for R v.4.2.2. The Mantel test allowed to evaluate if the levels of genetic differentiation observed respond to the geographical distance between the sites of collection of samples, ie to evaluate isolation by distance (IBD; Bohonak 2002). The gene flow and genetic structure analysis results, together with the results of the analysis of genetic diversity & inbreeding, will make possible to determine if there is evidence of priority subpopulations for management and conservation, in order to ensure long-term population prevalence.

Results

Extraction & Quantification of DNA

Total DNA was successfully isolated and purified from 104 biological samples processed including feces (N=88), blood (N=5), tissue (N=4) and hairs (N=7; Table SI). DNA concentrations found ranged from 1.30 ng/ul up to 186.40 ng/ul. All fecal samples (N=88) yielded quantifiable concentrations of total DNA after extraction (N=88; 56.67 ± 33.79 ng/ul). Protocol A yielded the lowest amount of DNA (N=9; 9.57 ± 6.40 ng/ul), whilst Protocol B yielded the highest concentration (N=59; 63.38 ± 36.78 ng/ul) followed by Protocol C (N=20; 53.96 ± 16.68 ng/ul).

Blood yielded the lowest concentration of DNA (N=5; 15.04 ± 4.12 ng/ul) and tissue the highest (N=4; 109.67 ± 45.52 ng/ul). Hairs yielded intermediate amounts of DNA (N=7; 38.71 ± 35.87 ng/ul). Values of 260/280 ratio found were closer to theoretical optimal, but 260/230 ratio differed considerably, especially in samples extracted from feces (Table SI). Fecal samples also displayed considerable degradation but still presented high molecular weight when visualized in 1.6% agarose gels (Fig. S4).

Validation of the Dataset

29 of 31 samples were successfully sequenced. *Cytb* consensus sequences (N=27) lengths obtained after trimming of ends and pairwise alignment ranged from 122 to 148 bp (Table II). Two samples generated a single direction fragment of 114 and 117 bp that was used as query. All 29 queries were examined independently and returned the same accession JF718880 (*Tapirus bairdii cytochrome b (CYTB) gene, complete cds; mitochondrial*) as the best match of the BLAST analysis (Table II). All queries reached 100% of coverage, whilst pairwise identity ranged from 99.3 to 100%. The weighted score grade ranged from 99.5 to 100%.

Validation of the Panel of Molecular Markers

Five microsatellite loci (ASB2, AF129732, AY138542, Tter05 and Tter13) were rapidly discarded because of high rates of unspecific products identified (*e.g.* Fig S4). Tba15 and Tte01 were also discarded due to low amplification success (31.3 and 29.7% respectively) after the validation essay. A partially validated panel of 8 loci (Table I) were selected to amplify the whole data set: Tte12, Tter09, Tter14, Tba23, Tter18, TtGt137, Tter04, Tba20.

98 samples from five populations defined *a priori* (Cerro de la Muerte: CM, N=53; Herradura: H, N=10; Inconclusive: INC, N=3; San Jerónimo: SJ, N=8 and Tenorio-Miravalles: TM, N=24) were genotyped at 8 microsatellite loci in a multi-tube approach of two independent replicates. After the construction of consensus genotypes for all samples, a total of 7.65% of missing data was identified. Samples with more than 5% missing data (N=36) were removed from further analysis to avoid bias in ordination analysis (Putman & Carbone 2014). 32 of 36 samples removed (approx. 90%) were form fecal origin, whilst 4 came from blood (N=1), tissue (N=1) and hairs (N=2).

The complete INC population was removed from analysis because of its low population number (N=3; allele frequencies are not accurately estimated with only three individuals) and artificial origin. H and SJ populations were merged into a new population (H-SJ) supported in geographical proximity and prior examination of genetic distances (Fig. S5). One additional sample was excluded from the data set because of its exclusive geographical origin outside from the study region. The final metapopulation consisted of N=58 samples distributed in three predefined populations (Cerro de la Muerte: CM, N=27; Herradura-San Jerónimo: H-SJ, N=12 and Tenorio-Miravalles: TM, N=19).

Null allele frequencies were calculated for eight loci using four different methods (Table S2). Five loci showed evidence of null alleles (Tba20, Tter09, Tter14, Tba23, TtGt137) at the metapopulation level. Only one locus (Tba20) presented a null allele frequency estimation consistently above 0.20 using all four estimators, possibly introducing significant bias in the results (Chapuis & Estoup 2007). The rest of the loci showed null allele frequencies between 0.0725-0.1597 (Brookfield estimator; Table S2). HWE analysis indicated seven of eight loci departed from HWE at metapopulation level (Table III). However, analysis at population level showed most of the markers behaved as expected under HWE (Table III). Only Tba20 departed from HWE at metapopulation level and all three population levels. Because of departures from HWE and high null allele frequencies observed, Tba20 was excluded from the molecular marker panel.

Evidence of LD was found for eight of the total 28 paired loci comparisons (Table S3) in the metapopulation (N=58, rbarD=0.0656, 10^4 iterations, p=0.0001). Additional values of association index (I_a) and their respective p-value are presented

in table S3. LD analysis at the population level revealed that H-SJ (N=12, rbarD=0.1472, 10^4 iterations, p=0.0005) and TM (N=19, rbarD=0.0959, 10^4 iterations, p=0.0005) showed significant evidence of LD. Since deviations from HWE and occurrence of LD may indicate the occurrence of non-random mating in wild populations, such as population structuring, the rest of markers were retained for subsequent genetic analyses (Lim et al. 2022).

The power of discrimination of the final microsatellite panel (N=7) was estimated with the calculation of $P_{(ID)}$unbiased for each marker (Table IV). The combined power of the loci reached $P_{(ID)}$unbiased= 9.77e-25. The genotype accumulation curve shows that more than 54 (95%) of the total number of multi-loci genotypes observed (MLGs, N=58) could be distinguished using six or more random markers from the established panel in this study (Fig. 3). Two genotypes were found to be replicates: a fecal sample (92c) and a hair sample (99c) from TM population.

Estimates of Genetic Diversity & Inbreeding

Genetic diversity statistics differed greatly across all 7 loci analyzed (Table IV). Allelic diversity (Ad) across loci was 8.43 ± 4.65. The observed heterozygosity (Ho) ranged from 0.31 (Tter18) to 0.78 (Tte12.), while inbreeding index F_{is} ranged from -0.36 (Tte12) to 0.36 (Tter09). The estimation of genetic diversity also differed among the three populations (Table V). The highest values of observed heterozygosity (Ho), allelic diversity (Ad) and allelic richness (Ar) were found in CM population and the lowest values in TM. The inbreeding coefficients estimated were similar across the three populations CM (Fis=0.0890), H-SJ (Fis=0.1218) and TM (Fis=0.0863).

Gene Flow & Genetic Structure Analysis

We found significant genetic structure among regions (G'_{ST} Nei=0.0587; p<0.001) and among the three populations (Table VI) using F_{ST} and G-Statistics. The AMOVA analysis found significant genetic structure among NWT and TMBC regions (ϕCT=0.0602; p<0.001, Table VII) and marginal genetic structure among populations (ϕSC=0.0479; p<0.05; Table VII). Bayesian analysis of genetic clusters (K) grouped individuals into K=2, suggesting a similar configuration of individuals as the regional stratification did (Fig. 4A, B). DAPC at regional level showed a clear overlapping among NWT and TMBC, supporting moderate levels of differentiation found between them (Fig 4C). K=3 had poor support, suggesting that *a priori* designated populations were artificial. (Fig. 5A, B). The DAPC plot at population level is consequent with G-statistics and STRUCTURE results, since a weak differentiation between the three groups is observed (Fig 5C). Genetic distance estimated among collection sites did not correlate to the geographic distance

between collection sites in the whole metapopulation, rejecting IBD hypothesis (rMantel= 0.0489, 10^4 permutation, p= 0.2076; Fig.6).

Discussion

In the present study we: 1) validated a panel of molecular markers (*cytb* + seven microsatellite loci) suitable for the taxonomic assignment of samples (*cytb*) and the description of genetic diversity & structure (microsatellite loci) of *T. bairdii* using tissue, blood, hairs and feces, 2) estimated the genetic diversity & structure of three wild tapir populations from Costa Rica for the first time, 3) evaluated inbreeding levels in the Costa Rican populations and 4) demonstrated that genetic distances found between populations are no caused by IBD.

Extraction & Quantification of DNA

DNA extracted from blood samples obtained the lowest concentrations of all biological material types used in this study (N=5; 15.04 ± 4.12 ng/ul). Values obtained for 260/280 and 260/230 ratios (Table S1) also displayed displacement from the theoretical optimum (1.80 and 2.00-2.20, respectively) suggesting impurities remained after DNA extraction (Table S1). Optimization of blood DNA extraction may be achievable since we have identified a better-performing protocol in terms of yield and quality of isolates. Early tests on blood extractions from other vertebrates show a better performance of Qiagen DNeasy Blood and Tissue kit compared to PureLink Genomic DNA Mini Kit, but optimization for tapirs' blood is pending. Additionally, several protocols have been developed to deal with degraded blood (Mardan-Nik & Maryam 2019; Carrasco et al. 2020). Yet only one of five DNA samples from blood were discarded due to missing alleles in only one microsatellite loci, evidencing high quality DNA.

When extracting DNA from fecal samples, Protocol B and C performed dramatically better than Protocol A. Protocol A followed manufacturer's instructions without any modification, while protocol B and C shared all our proposed modifications but differ in the year of extraction (Protocol B: 2019-2021; Protol C: 2022-2023). The differences found between protocols A and B & C were expected because QIAamp DNA Stool Mini kit is optimized to extract DNA from human feces, while tapirs are exclusive herbivores and their stool may contain plant remains and a wide variety of secondary metabolites derived from their diet (Monteiro et al. 1997; Wehausen et al. 2004).

Validation of the Dataset

Fecal DNA is predominantly from exogenous origin. Non-target DNA derived from microorganisms (endogenous from gut microbiota and exogenous from the environment after deposition of the feces in the wild), diet remains and fungi constitute the major proportion of DNA isolated from stool, compared to a minor proportion of endogenous DNA or target DNA (Perry et al. 2010; Snyder-Mackler et al. 2016; Chiou & Bergey 2018). Therefore, DNA concentration observed after DNA extraction from fecal samples does not represent the amount of the target DNA (tapir specific DNA). We developed a dataset validation essay aimed to confirm the presence of target DNA in our isolates, with emphasis on fecal samples.

Around 92.5% (24/26) of the fecal samples analyzed in the essay amplified and produced high quality fragments of *cytb* gene (120-148 bp; Table II) that matched to *T. bairdii* sequences in NCBI dataset with high confidence values (Table II). This evidence the presence of specific target DNA above the threshold of detection. The high success of amplification obtained can be explained because of several factors.: First, a successful DNA extraction protocol, improved with the modifications, yielded high amounts of total DNA, including enough target DNA to be detected. Secondly, the primer pair selected (L15601 & H15748; Lopez-Oceja et al. 2016) amplified a small fragment (150 pb) of *cytb*, which was compatible with the degraded DNA observed in non-invasive samples. Thirdly, the short length of target amplicon allowed us to set a high number of cycles of amplification (N=40) increasing the chance to amplify target DNA without the risk of depleting the PCR reaction and, finally, the DNA polymerase used, AmpliTaq Gold (Thermo Fisher Scientific), has a very high sensitivity (1-3 copies; Zimmermann et al. 1998), ideal to capture specific target DNA in low copy number. After we validated the presence of target DNA in DNA isolates, we continued to validate the panel of microsatellite loci.

All types of starting biological material (blood, tissue, hairs, feces) yielded high-quality amplicons of *cytb* fragment, allowing the taxonomical assignment of the samples with confidence (Table II). Because of this, we recommend primers L15601 & H15748 for any molecular identification needed for *T. bairdii*, especially when starting with poor quality samples.

Validation of the Panel of Molecular Markers

Five microsatellite loci (*ASB2, AF129732, AY138542, Tter05* and *Tter13*) were rapidly discarded after unspecific products were detected. All these markers proceeded from primer pairs designed to work with other species and their cross-amplification was not possible. Another two markers (*Tba15* and *Tte01*) were discarded after low amplification success -but no unspecific products- in a sample subset, suggesting their optimization could be achieved. Both markers were

originally designed and validated for *T. baidii*, but they have not been used since their publication (Norton & Ashley 2004). Marker *Tba20* was discarded after returning the strongest signals of HW disequilibrium and null alleles frequencies. The presence of null alleles translates into a higher proportion of homozygotes and thus in reduced estimates of genetic diversity, it can also lead to the overestimation of the genetic differentiation (Latorre-Cárdenas et al. 2020). In fact, *Tba20* returns the highest value of Fis observed: 0.88, an estimate very far from the other markers (F_{IS}: -0.36-0.36).

The validated set of markers (N=7) employed in this study represent the highest number of microsatellite loci used to date to study the genetics of *T. bairdii* (Table VIII). The combined power of discrimination of the set was estimated in $P_{(ID)}$unbiased= 9.77e-25 (Table IV). These results suggest that using seven loci is sufficient to distinguish different individuals ($P_{(ID)}$unbiased < 0.01; Waits et al. 2001). The genotype accumulation curve supports the $P_{(ID)}$ analysis after reaching the 95% of all multi-loci genotypes observed using 6 markers (Fig. 3).

Estimates of Genetic Diversity & Inbreeding

We found moderate levels of microsatellite genetic diversity (Ho=0.52) in the tapir *T. bairdii* when analyzing the whole Costa Rican metapopulation. Our estimates by population (table V) ranged in the middle of previous reports from different populations of the species and the genus (Table VIII). All three populations analyzed in this study exhibit higher estimates of genetic diversity (Ho=0.46-0.58) than previously reported values from wild populations of Central America, but the estimates are close to overlap. Norton & Ashley (2004a) reported modest levels of observed heterozygosity (N= 15 and 15; Ho=0.39 and 0.41) in the populations of *T. bairdii* in Corcovado, Costa Rica and Darien, Panamá. Zavala-Páramo et al. (2017) reported the highest levels of Ho found in any tapir species to date (Ho=0.82), but the population examined is the smallest of all and was comprised exclusively by captive Baird's tapirs from several zoos in Mexico and USA.

CM population exhibit the highest genetic diversity of all three populations examined in this study, which is partially expected because CM also has the highest population size (N=27), the wider area of influence (Fig. 1) and a high abundance and occupancy of tapirs (González-Maya 2012, Brenes-Mora 2018). H-SJ displayed intermediate levels of diversity, even when it had the smallest population size (N=12). Both CM and H-SJ are part of the same region NWT, which is recognized as part of the core habitats recognized for *T. bairdii* (Schank 2017, Schank 2020). NWT region is also part of La Amistad Forest, the fourth largest wilderness in Central America covering around 14991 km2 (WCS n.d.a).

Several important protected areas like La Amistad International Park, Chirripó National Park, Tapantí-Macizo de la Muerte National Park and Los Quetzales

National Park have provided the maximum degree of protection by environmental authorities to the wild populations of tapirs and other mammals inhabiting this region for at least 48 years (González-Maya et al. 2012, González-Maya et al. 2015, Brenes-Mora 2018). Because of the close proximity of this protected areas to the collection sites, we expected both populations CM and H-SJ to display high levels of genetic diversity. It is also important to recall that H-SJ was constituted by the individuals of Herradura and San Jerónimo original populations, possibly introducing a bias towards higher diversity estimates (Fig. 1).

TM (N=19) showed the lowest levels of genetic diversity observed, but it has the smallest area of influence too (Fig. 2). Nevertheless, its estimates are moderately higher than other healthy wild population of tapirs (Corcovado, Darién) reported before (Norton & Ashley 2004a). TMBC is identified as a core habitat patch of the connectivity network modeled for *T. bairdii* (Schank et al. 2020). Protected areas like Volcán Tenorio National Park, Miravalles-Jorge Manuel Dengo National Park and Rincón de la Vieja National Park, all around TMBC, represent important sites for the conservation of wild tapir populations in the northern part of Costa Rica (Carbonell & González 2000; Amit et al. 2009). Wild tapirs are also frequently reported outside protected areas in TMBC (Pastor-Parajeles 2022). TMBC is close to Indio Maíz-Tortugero Forest, an extensive wilderness part of the five great forests of Mesoamerica with presence of wild tapir populations (WCS n.d.b; García et al. 2016). Hence, we also expected relatively high levels of genetic diversity in this population.

Private alleles found in the three populations may support local differentiation (Table VI). The inbreeding levels assessed in CM, H-SJ and TM populations ranged from F_{IS}= 0.086 to 0.122. These values are similar to the estimates reported in the wild metapopulation of tapirs in Central America (N=33; F_{IS}=0.11; Norton & Ashley 2004) but higher than estimates from Corcovado, Costa Rica (N=15; F_{IS}; -0.058; Norton & Ashley 2004) and Darién, Panamá (N=15; F_{IS}; 0.032; Norton & Ashley 2004).

Gene Flow & Genetic Structure Analysis

Our findings suggest that the genetic structure of *T. bairdii* found using seven microsatellite loci can be attributed primarily to isolation between regions (Table VI, VII) and secondarily to limited gene flow between populations (Table VI, VII). These results were expected because more than 150 km separate NWT from TMBC, including barriers to genetic flow such as the Gran Area Metropolitana (GAM), the largest urban agglomeration in Costa Rica, and an extensive road network. NWT and TMBC are also part of two independent but proximal key regions for the connectivity and preservation of the Baird's tapir (Schank 2017, Schank 2020).

Marginal levels of genetic structure were found among H-SJ population and TM, despite the closer linear proximity of CM to TM compared to H-SJ. (Table VI; Fig S1). The assignment bar plots of K=3 (Fig. 4A, B) showed that H-SJ population members are redistributed in all different clusters, but predominantly to TM cluster. The H-SJ population is also predominantly redistributed to TMBC cluster when considering K=2 (Fig 3A, B). Both redistribution of H-SJ members suggests: 1) the artificial composition of the population, which is expected because of its *a priori* definition and dual (H and SJ) origin and 2) a weak genetic structure between H-SJ and TM populations.

The marginal genetic structure found between H-SJ and TM, plus the significant differentiation of CM and TM (F_{ST}= 0.0583, higher than the differentiation between NWT and TMBC), suggest a stronger putative pathway of gene flow between TM and H-SJ through the Atlantic slope and its protected areas (e.g. La Amistad International Park, Chirripó National Park, Tapantí-Macizo de la Muerte National Park, Braulio Carrillo National Park, Arenal National Park, etc.) and connectivity routes; and a weaker putative pathway of gene flow between TM and CM through the Pacific slope. The suggested putative gene flow pathway through the Atlantic slope has been identified as part of three of the Jaguar Conservation Units or JCUs in Costa Rica, that is, expert-defined areas that are believed to have resident jaguar populations, an adequate prey base and high habitat quality for this species (Salom-Pérez et al. 2021). At least two habitat patches in the Atlantic slope are also recognized as key components of the connectivity network modeled for *T. bairdii* (Schank et al. 2020). Additionally, a recent proposal for practical and effective biological corridors to connect protected areas in northwest Costa Rica identified and promote four sub-corridors to link five major protected areas in northwest Costa Rica through the Atlantic slope, with all of them linking to larger protected areas in the central portion of the country (Moran et al. 2019).

In the other hand, a recent evaluation of forest cover and fragmentation in Costa Rica demonstrates a higher deforested, fragmented and urbanized landscape through the Pacific slope of the country, potentially hindering gene flow (Cunningham et al. 2020). Urban centers like Esparza, San Mateo, Orotina, Atenas, Palmares and San Ramón together with extensive road infrastructure (*e.g.* Pan-American Highway, Route 3, Route 27, Route 34, etc.) may present resistance to gene flow between NWT and TMBC regions through the Pacific Slope of the country.

We recommend sampling additional populations from both Atlantic and Pacific slopes of Costa Rica and perform additional population and landscape genetics analysis to validate the asymmetrical gene flow observed in this study and to better understand the patterns of genetic structure in the wild tapir metapopulation of Costa Rica. Finally, isolation by distance was not found to be a main cause of genetic structure (Fig. 6), but finer spatial analysis is recommended because of the dichotomous distribution of geographical distances in our study design.

Conservation & Management Implications

Latin America is the region of the world with the highest observed loss in biodiversity in current years (WWF 2020). The available information related to wildlife species in the region (*eg.* ecological data, morphometric data, genetic data) is scarce compared to other regions in the world (Collen et al. 2008; Feeley & Silman 2010). The amount of information available on wildlife has been found to increase further from the Tropics (Collen et al. 2008; Feeley & Silman 2010). The unavailability of basic information related to wildlife species in the Neotropics is a challenge for the application of techniques such as modern modeling, predictive tools and the enforcement of conservation and management strategies such as *ex situ* management (Feeley & Silman 2010; Ruiz-García et al. 2012; Zavalo-Páramo et al. 2017). This bias could be potentially worst for tapirs and *T. bairdii* if we consider that: i) big herbivores are highly susceptible to threats related to the change of use of land (Ripple et al. 2015; Atwood et al. 2020), ii) in larger species, the extinction risk is driven by a combination of environmental factors and intrinsic traits (Cardillo et al. 2005) and 3) *T. bairdii* is already considered an endangered species with declining populations by IUCN and Costa Rican Environmental Institutions.

Community-based initiatives supported by Academia, NGO and Public Institutions represent a great opportunity to offset the lack of information of wildlife in the Tropics. In our present study, we achieved to collect, process, and analyze the highest number of samples reported to date for any American tapir species using microsatellite markers (Table VIII). This collective achievement was supported by the field work of WMRGs composed of local communities' members coexisting with wildlife and interested in its preservation. Understanding and promoting the positive interactions of local communities with nature is an important component in achieving effective conservation, as top-down initiatives that have worked well in temperate and developed regions have often been applied with limited success in the Tropics (Sodhi et al. 2011; Büscher & Fletcher 2019).

It is estimated that only 22% of wildlife genetics studies include non-invasive samples (Zemanova 2019). A high rate of genotyping errors, the high cost of sample replication to achieve confidence levels and the failure to obtain reproducible genotypes could reflect in a poor implementation of non-invasive samples in more wildlife genetics studies (Smith & Wang 2014; Zemanova 2019). Our final validated dataset included approx. 88% of non-invasive samples (N=51/58), supporting their validity as a reliable genetic source to conduct wild population studies. In terms of molecular markers, conservation genetics is shifting from the use of microsatellite DNA markers to single nucleotide polymorphisms or SNPs (Puckett 2017). However, the SNPs implementation is more expensive and microsatellites still remain as a cost-efficient and effective conservation tool that can provide critical information on

population structures of threatened wild species, especially in developing countries (Puckett 2017; Vashistha et al. 2020; Lim et al. 2022; Hauser et al. 2021).

At least two previous studies of tapirs employing fecal samples reported issues of amplification success after DNA isolation. Pinho et al. (2014) reported low microsatellite amplification success after 15 days of extraction, whilst McCann (2015) was unable to reproduce any microsatellite genotype from fecal DNA. Both studies performed similar DNA isolation protocols using commercial kits, following most of manufacturers' instructions as our Protocol A of extraction did. We found that the modification of the original DNA isolation protocol in two key steps: incubation and elution (Protocol B and C), resulted in a significative improvement of DNA yield, but also in high DNA stability since three-years-old DNA isolates have produced reproducible genotypes for both *cytb* and the microsatellite panel. Thus, we highly recommend the use of non-invasive samples as well as the modification of commercial DNA isolation protocols when working with fecal samples from wildlife, considering these kits are generally designed and optimized to work with human samples.

We found moderate levels of microsatellite genetic diversity in the *T. bairdii* population analyzed. All three populations analyzed exhibit higher estimates of genetic diversity (Ho=0.46-0.58) than previously reported values from wild populations of Central America, even when most of the samples come from collection sites outside protected areas. Both NWT and TMBC are recognized as key regions for the connectivity and preservation of wild populations of tapirs in Costa Rica (Schank 2017, Schank 2020). Besides, NWT is part of La Amistad Forest, while TMBC is really close to Indio Maíz-Tortuguero Forest, two of the biggest wilderness areas in Mesoamerica and essential components of connectivity in Central America (WCS n.d.a, n.d.b., García et al. 2016). The relatively high genetic diversity observed in all populations analyzed may reflect the proximity to protected areas, recognized as the most important core population centers for the species in the country (García et al. 2016).

The genetic structure analysis is consistent with the strong influence of both core areas because the regional stratification accounts for most of the structure found. Nevertheless, the AMOVA analysis, the pairwise F statistics comparisons among populations and the presence of private alleles in each group support the population stratification. The marginal genetic structure found between H-SJ and TM and the differentiation between CM and TM (higher than the differentiation between NWT and TMBC), suggest a stronger putative pathway of gene flow between TM and H-SJ through the Atlantic slope and its protected areas and connectivity routes; and a weaker putative pathway of gene flow between TM and CM through the Pacific slope. Sampling additional populations from both Atlantic and Pacific slopes of Costa Rica must be a priority to perform additional population and landscape genetics analysis to validate the asymmetrical gene flow observed in this study and to better

understand the patterns of genetic structure in the wild tapir metapopulation of Costa Rica.

Bibliography

Allendorf, F. W., & Luikart, G. (2009). *Conservation and the genetics of populations*. John Wiley & Sons.

Allendorf, F. W., Funk, W. C., Aitken, S. N., Byrne, M., & Luikart, G. (2022). *Conservation and the Genomics of Populations*. Oxford University Press.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215(3), 403-410.

Amit, R., Alfaro, L. D., & Carrillo, E. (2009). Relación entre la presencia de jaguar (Pantheraonca) y de sus presas en el Área de Conservación Guanacaste, Costa Rica. *Revista de Ciencias Ambientales, 38(2),* 6-10.

Arias, E., Chacón, O., Herrera, B., Induni, G., Acevedo, H., Coto, M., & Barborak, J. R. (2008). Las redes de conectividad como base para la planificación de la conservación de la biodiversidad: propuesta para Costa Rica. *Recursos Naturales y Ambiente*, (54).

Atwood, T. B., Valentine, S. A., Hammill, E., McCauley, D. J., Madin, E. M., Beard, K. H., & Pearse, W. D. (2020). Herbivores at the highest risk of extinction among mammals, birds, and reptiles. *Science advances*, *6*(32), eabb8458.

Baltodano, A., & Zamora, J. C. (2009). Diseño de una red de conectividad en el Corredor Biológico Tenorio-Miravalles. Informe de taller. Bijagua, Costa Rica. *CATIE*

Barrantes, G. (2009). The role of historical and local factors in determining species composition of the highland avifauna of Costa Rica and Western Panamá. *Revista de Biología Tropical 57*:333–349.

Bautista-Solís, P., Gutiérrez-Montes, I. A., Aguilar, J., Cotto, E., Gómez, C., González, M., Guillén, D., Mendoza, J., Morales, I., Pinoth, R., Posada, K., Quiñónez, G., Salazar, Á., Salagado, M., Steinvorth, K., & Zambrano, M. (2012). *Capitales de la comunidad y la conservación de los recursos naturales: El caso del Corredor Biológico Tenorio-Miravalles*. CATIE

Bello, C., Galetti, M., Pizo, M. A., Magnago, L. F. S., Rocha, M. F., Lima, R. A., ... & Jordano, P. (2015). Defaunation affects carbon storage in tropical forests. *Science advances*, *1*(11), e1501105.

Bohonak, A. J. (2002). IBD (isolation by distance): a program for analyses of isolation by distance. *Journal of Heredity*, *93*(2), 153-154.

Breen, M., Lindgren, G., Binns, M. M., Norman, J., Irvin, Z., Bell, K., ... & Ellegren, H. (1997). Genetical and physical assignments of equine microsatellites—first integration of anchored markers in horse genome mapping. *Mammalian genome, 8(4),* 267-273.

Brenes-Mora, E. (2018). Patrones de actividad, selección de hábitat y atropellos de danta (*Tapirus bairdii*) en bosque atravesado por una carretera en la Cordillera de Talamanca, Costa Rica. *Magister Scientiae thesis*, Universidad Nacional.

Brooks, D. M., Bodmer, R. E., & Matola, S. (Eds.). (1997). *Tapirs: status survey and conservation action plan* (Vol. 38). IUCN.

Büscher, B., & Fletcher, R. (2019). Towards convivial conservation. *Conservation & Society, 17(3)*, 283-296.

Carbonell, F. & Gonzalez, J. (2000). Análisis Ecológico para la Determinación del Hábitat Actual y Potencial del Tapir (*Tapirus bairdii)* en el Parque Nacional Volcán Tenorio y Zona Protectora Miravalles. *ACA-SIG-INBIO.* http://www.inbio.eas.ualberta.ca/es/estudios/PDF/Informe_Danta.pdf Accessed on 17 April 2023.

Cardillo, M., Mace, G. M., Jones, K. E., Bielby, J., Bininda-Emonds, O. R., Sechrest, W., ... & Purvis, A. (2005). Multiple causes of high extinction risk in large mammal species. *Science*, *309*(5738), 1239-1241.

Carrasco, P., Inostroza, C., Didier, M., Godoy, M., Holt, C. L., Tabak, J., & Loftus, A. (2020). Optimizing DNA recovery and forensic typing of degraded blood and dental remains using a specialized extraction method, comprehensive qPCR sample characterization, and massively parallel sequencing. *International Journal of Legal Medicine*, *134*, 79-91.

Chapuis, M. P., & Estoup, A. (2007). Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, *24*(3), 621-631.

Chiou, K. L., & Bergey, C. M. (2018). Methylation-based enrichment facilitates low-cost, noninvasive genomic scale sequencing of populations from feces. *Scientific Reports*, *8*(1), 1975.

Colbert, M. (2007). New fossil discoveries and the history of Tapirus. *Tapir Conservation*, *16*, 12-14.

Collen, B., Ram, M., Zamin, T., & McRae, L. (2008). The tropical biodiversity data gap: addressing disparity in global monitoring. *Tropical Conservation Science, 1*(2), 75-88.

Cunningham, D., Cunningham, P., & Fagan, M. E. (2020). Evaluating forest cover and fragmentation in Costa Rica with a corrected global tree cover map. *Remote Sensing, 12(19),* 3226.

da Silva, A. G., Lalonde, D. R., & Russello, M. A. (2009). Isolation and characterization of microsatellite loci in a Neotropical ungulate, the lowland tapir (*Tapirus terrestris*). *Conservation Genetics Resources*, *1*(1), 39.

de Barba, M., Waits, L. P., Garton, E. O., Genovesi, P., Randi, E., Mustoni, A., & Groff, C. (2010). The power of genetic monitoring for studying demography, ecology and genetics of a reintroduced brown bear population. *Molecular Ecology*, *19*(18), 3938-3951.

Deng, T. (2006). Paleoecological comparison between late Miocene localities of China and Greece based on Hipparion faunas. Geodiversitas, 28(3), 499-516.

Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources*, *4*(2), 359-361.

Feeley, K. J., & Silman, M. R. (2011). The data void in modeling current and future distributions of tropical species. *Global Change Biology, 17*(1), 626-630.

Foerster, C. R., & Vaughan, C. (2002). Home Range, Habitat Use, and Activity of Baird's Tapir in Costa Rica1. *Biotropica*, *34*(3), 42

Frankham, R. (2010). Challenges and opportunities of genetic approaches to biological conservation. *Biological Conservation*, *143*(9), 1919-1927.

Garcìa, M., Jordan, C., O'Farril, G., Poot, C., Meyer, N., Estrada, N., Leonardo, R., Naranjo, E., Simons, Á., Herrera, A., Urgilés, C., Schank, C., Boshoff, L. & Ruiz-Galeano, M. (2016). *Tapirus bairdii*. *The IUCN Red List of Threatened Species* 2016:e.T21471A45173340. https://dx.doi.org/10.2305/IUCN.UK.2016-1.RLTS.T21471A45173340.en. Accessed on 17 April 2023.

Garcia, M., Medici, E. P., Naranjo, E. J., Novarino, W., & Leonardo, R. S. (2012). Distribution, habitat and adaptability of the genus Tapirus. Integrative zoology, 7(4), 346-355.

González-Maya, J. F., Schipper, J., Polidoro, B., Hoepker, A., Zárrate-Charry, D., & Belant, J. L. (2012). Baird's tapir density in high elevation forests of the Talamanca region of Costa Rica. *Integrative Zoology, 7(4),* 381-388.

Gonzalez-Maya, J. F., Víquez-R, L. R., Belant, J. L., & Ceballos, G. (2015). Effectiveness of protected areas for representing species and populations of terrestrial mammals in Costa Rica. *PloS one, 10(5*), e0124480.

Goudet, J. & T. Jombart. (2015). hierfstat: Estimation and Tests of Hierarchical F-Statistics. *R package version* 0.04-22.

Hauser, S. S., Athrey, G., & Leberg, P. L. (2021). Waste not, want not: microsatellites remain an economical and informative technology for conservation genetics. *Ecology and Evolution, 11(22),* 15800-15814.

Hibert, F., Sabatier, D., Andrivot, J., Scotti-Saintagne, C., Gonzalez, S., Prévost, M. F., ... & Richard-Hansen, C. (2011). Botany, genetics and ethnobotany: a crossed investigation on the elusive tapir's diet in French Guiana. *PLoS One*, *6*(10), e25850.

Hulbert Jr, R. C. (2005). Late Miocene Tapirus (Mammalia, Perissodactyla) from Florida, with description of a new species, Tapirus webbi. *Bulletin of the Florida Museum of Natural History, 45(4*), 465-494.

Hulce, D., Li, X., Snyder-Leiby, T., & Liu, C. J. (2011). GeneMarker® genotyping software: Tools to increase the statistical power of DNA fragment analysis. *Journal of biomolecular techniques: JBT*, *22*(Suppl), S35.

Isaac, N. J., Turvey, S. T., Collen, B., Waterman, C., & Baillie, J. E. (2007). Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PloS one*, *2*(3), e296.

Janzen, D. H. (1982). Seeds in tapir dung in Santa Rosa National Park, Costa Rica. *Brenesia*.

Janzen, D. H. (1991). *Historia Natural de Costa Rica*. Editorial de la Universidad de Costa Rica, San Jose, Costa Rica.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic acids research, 36(suppl_2)*, W5-W9.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics, 24(11),* 1403-1405.

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 1-15.

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

Kapelle, M., L. van Omme, y M. E. Juarez. (2000). Lista de la flora vascular de la cuenca superior del Rio Savegre, San Gerardo de Dota, Costa Rica. *Acta Botanica Mexicana, 51*:1-38.

Latorre-Cardenas, M. C., Gutiérrez-Rodríguez, C., & Lance, S. L. (2020). Isolation and characterization of 13 microsatellite loci for the Neotropical otter, Lontra longicaudis, by next generation sequencing. *Molecular Biology Reports*, *47*(1), 731-736.

Lizcano, D.J., Amanzo, J., Castellanos, A., Tapia, A. & Lopez-Malaga, C.M. (2016). *Tapirus pinchaque*. *The IUCN Red List of Threatened Species* 2016: e.T21473A45173922. https://dx.doi.org/10.2305/IUCN.UK.2016-1.RLTS.T21473A45173922.en. Accessed on 17 April 2023.

Lopez-Oceja, A., Gamarra, D., Borragan, S., Jiménez-Moreno, S., & De Pancorbo, M. M. (2016). New cyt b gene universal primer set for forensic analysis. *Forensic Science International: Genetics, 23*, 159-165.

Mardan-Nik, M., Saffar Soflaei, S., Biabangard-Zak, A., Asghari, M., Saljoughian, S., Tajbakhsh, A., ... & Ghayour-Mobarhan, M. (2019). A method for improving the efficiency of DNA extraction from clotted blood samples. *Journal of Clinical Laboratory Analysis*, *33*(6), e22892.

McCann, N. P. (2015). *The conservation of Baird's Tapir (*Tapirus bairdii*) in Honduras*. Doctoral thesis, Cardiff University.

Medici, E. P. (2011). Family tapiridae (TAPIRS). *Handbook of the mammals of the world*, *2*, 182-204

Medici, E. P., Carrillo, L., Montenegro, O. L., Miller, P. S., Carbonell, F., Chassot, O., & Mendoza, A. (2005). *Baird's tapir (*Tapirus bairdii*) conservation workshop population and habitat viability assessment (PHVA)*. Final report. IUCN/SSC Conservation Breeding Specialist Group, Belize, Central America.

Monteiro, L., Bonnemaison, D., Vekris, A., Petry, K. G., Bonnet, J., Vidal, R., ... & Mégraud, F. (1997). Complex polysaccharides as PCR inhibitors in feces: Helicobacter pylori model. *Journal of Clinical Microbiology*, *35*(4), 995-998.

Moran, M. D., Monroe, A., & Stallcup, L. (2019). A proposal for practical and effective biological corridors to connect protected areas in northwest Costa Rica. *Nature Conservation, 36,* 113-137.

Naranjo-Piñera, E. (1995). Abundancia y uso de hábitat del tapir (Tapirus bairdii) en un bosque tropical húmedo de Costa Rica. *Vida Silvestre Neotropical*, *4*(1), 20-31.

Naranjo, E. J. (2009). Ecology and conservation of Baird's tapir in Mexico. Tropical Conservation Science, 2(2), 140-158.

Nielsen, L., Meehan-Meola, D., Kilbourn, A., & Alcivar-Warren, A. (2008). Characterization of microsatellite loci in the black rhinoceros (*Diceros bicornis*) and white rhinoceros (*Ceratotherium simum*): their use for cross-species amplification and differentiation between the two species. *Conservation genetics, 9(1),* 239-242.

Norton, J. E., & M. V Ashley. (2004a). Genetic variability and population structure among wild Baird's tapirs. *Animal Conservation*, 7, 211–220.

Norton, J. E., & Ashley, M. V. (2004b). Genetic variability and population differentiation in captive Baird's tapirs (Tapirus bairdii). *Zoo Biology: Published in affiliation with the American Zoo and Aquarium Association, 23*(6), 521-531

Noss, R. F. (1990). Indicators for monitoring biodiversity: a hierarchical approach. *Conservation Biology*, *4*(4), 355-364.

O'Farrill, G., Calmé, S., Sengupta, R., & Gonzalez, A. (2012). Effective dispersal of large seeds by Baird's tapir: a large-scale field experiment. *Journal of Tropical Ecology*, *28*(1), 119-122.

O'Farrill, G., Galetti, M., & Campos-Arceiz, A. (2013). Frugivory and seed dispersal by tapirs: an insight on their ecological role. *Integrative Zoology*, 4-17.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M. H. H., Oksanen, M. J., & Suggests, M. A. S. S. (2007). The vegan package. *Community ecology package*, *10*(631-637), 719.

Paolucci, L. N., Pereira, R. L., Rattis, L., Silvério, D. V., Marques, N. C., Macedo, M. N., & Brando, P. M. (2019). Lowland tapirs facilitate seed dispersal in degraded Amazonian forests. *Biotropica*, *51*(2), 245-252.

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

Pastor-Parajeles, S. (2022). Conectividad del hábitat según la aceptabilidad de las y los propietarios de propiedades privadas hacia el desplazamiento de la danta, Tapirus bairdii (Tapiridae), en el Corredor Biológico Tenorio-Miravalles. *Licentiate thesis*, Universidad Nacional de Costa Rica.

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, *28*(19).

Perry, G. H., Marioni, J. C., Melsted, P., & Gilad, Y. (2010). Genomic-scale capture and sequencing of endogenous DNA from feces. *Molecular Ecology*, *19*(24), 5332-5344.

Pinho, G. M., da Silva, A. G., Hrbek, T., Venticinque, E. M., & Farias, I. P. (2014). Kinship and social behavior of lowland tapirs (Tapirus terrestris) in a Central Amazon landscape*. PloS One, 9*(3), e92507.

Pompanon, F., Bonin, A., Bellemain, E., & Taberlet, P. (2005). Genotyping errors: causes, consequences and solutions. *Nature reviews. Genetics*, *6*(11), 847.

Pritchard, J. K., Stephens, M., & P. Donnelly. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945-959.

Puckett, E. (2017) Variability in total project and per sample genotyping costs under varying study designs including with microsatellites or SNPs to answer conservation genetic questions. *Conservation Genetics Resources 9:*289–304

Putman, A. I., & Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution, 4*(22), 4399-4428.

R Core Team (2017). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org.

Ramasamy, R. K., Ramasamy, S., Bindroo, B. B., & Naik, V. G. (2014). STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. *SpringerPlus*, *3*(1), 1-3.

Ramón-Laca, A., Soriano, L., Gleeson, D., & Godoy, J. A. (2015). A simple and effective method for obtaining mammal DNA from faeces. *Wildlife Biology, 21*(4), 195-204.

Regulación del Programa Nacional de Corredores Biológicos, N° 40043 -MINAE (2016). http://www.pgrweb.go.cr/scij/Busqueda/Normativa/Normas/nrm_texto_completo.aspx?param1=NRTC&nValor1=1&nValor2=83424&nValor3=107128&strTipM=TC. Accessed on 25 May 2023.

Ripple, W. J., Newsome, T. M., Wolf, C., Dirzo, R., Everatt, K. T., Galetti, M., ... & Van Valkenburgh, B. (2015). Collapse of the world's largest herbivores. *Science advances*, *1*(4), e1400103.

Ruiz-García, M., Vásquez, C., Pinedo-Castro, M., Sandoval, S., Castellanos, A., Kaston, F., ... & Shostell, J. (2012). Phylogeography of the Mountain Tapir (Tapirus pinchaque) and the Central American Tapir (Tapirus bairdii) and the origins of the three Latin-American tapirs by means of mtCyt-B sequences. *Current topics in phylogenetics and phylogeography of terrestrial and aquatic systems. InTech, Rijeka*, 83-116.

Sanches, A., de Figueiredo, M. G., Hatanaka, T., de Paula, F. F. P., Silveira, L., Jácomo, A. T. A., & Galetti, P. M. (2009). Microsatellite loci isolated from the lowland tapir (Tapirus terrestris), one of the largest Neotropical mammal. *Conservation Genetics Resources*, *1*, 115-117.

Salom-Pérez, R., Corrales-Gutiérrez, D., Araya-Gamboa, D., Espinoza-Muñoz, D., Finegan, B., & Petracca, L. S. (2021). Forest cover mediates large and medium-sized mammal occurrence in a critical link of the Mesoamerican Biological Corridor. *PLoS one, 16(3),* e0249072.

Schank, C. J., Cove, M. V., Arima, E. Y., Brandt, L. S., Brenes-Mora, E., Carver, A., ... & Miller, J. A. (2020). Population status, connectivity, and conservation action for the endangered Baird's tapir. *Biological Conservation*, *245*, 108501.

Schank, C. J., Cove, M. V., Kelly, M. J., Mendoza, E., O'Farrill, G., Reyna-Hurtado, R., ... & Miller, J. A. (2017). Using a novel model approach to assess the distribution and conservation status of the endangered Baird's tapir. *Diversity and Distributions*, *23*(12), 1459-1471.

Schwartz, M. K., Luikart, G., & Waples, R. S. (2007). Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution*, *22*(1), 25-33.

Smith, O., & Wang, J. (2014). When can noninvasive samples provide sufficient information in conservation genetics studies?. *Molecular Ecology Resources, 14(5),* 1011-1023.

Snyder-Mackler, N., Majoros, W. H., Yuan, M. L., Shaver, A. O., Gordon, J. B., Kopp, G. H., ... & Tung, J. (2016). Efficient genome-wide sequencing and low-coverage pedigree analysis from noninvasively collected samples. *Genetics*, *203*(2), 699-714.

Sodhi, N. S., Butler, R., & Raven, P. H. (2011). Bottom-up conservation. *Biotropica, 43(5*), 521-523.

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., ... & Bouvet, J. (1996). Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic acids research, 24*(16), 3189-3194.

Traeholt, C., Novarino, W., bin Saaban, S., Shwe, N.M., Lynam, A., Zainuddin, Z., Simpson, B. & bin Mohd, S. (2016). *Tapirus indicus*. *The IUCN Red List of Threatened Species* 2016: e.T21472A45173636. https://dx.doi.org/10.2305/IUCN.UK.2016-1.RLTS.T21472A45173636.en. Accessed on 17 April 2023.

UNESCO (n.d.a). *La Amistad Biosphere Reserve, Costa Tica.* https://en.unesco.org/biosphere/lac/la-amistad_costa-rica. Accessed on 17 April 2023.

UNESCO (n.d.b). *Savegre Biosphere Reserve, Costa Rica*. https://en.unesco.org/biosphere/lac/savegre. Accessed on 17 April 2023.

UNESCO (n.d.c). *Agua y Paz Biosphere Reserve, Costa Rica.* https://en.unesco.org/biosphere/lac/agua-y-paz. Accessed on 30 May 2023.

Valière, N. (2002). GIMLET: a computer program for analysing genetic individual identification data. *Molecular Ecology Resources, 2*(3), 377-379.

van Oosterhout, C., Hutchinson, W. F., Wills, D. P., & Shipley, P. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Resources*, *4*(3), 535-538.

Varela, D., Flesher, K., Cartes, J.L., de Bustos, S., Chalukian, S., Ayala, G. & Richard-Hansen, C. (2019). *Tapirus terrestris*. *The IUCN Red List of Threatened Species* 2019:e.T21474A45174127. https://dx.doi.org/10.2305/IUCN.UK.2019-1.RLTS.T21474A45174127.en. Accessed on 17 April 2023.

Vashistha, G., Deepika, S., Dhakate, P. M., Khudsar, F. A., & Kothamasi, D. (2020). The effectiveness of microsatellite DNA as a genetic tool in crocodilian conservation. *Conservation Genetics Resources, 12*, 733-744.

Villar, N., Paz, C., Zipparro, V., Nazareth, S., Bulascoschi, L., Bakker, E. S., & Galetti, M. (2021). Frugivory underpins the nitrogen cycle. *Functional Ecology*, *35*(2), 357-368.

Waits, L. P., & Paetkau, D. (2005). Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. *The Journal of Wildlife Management*, *69*(4), 1419-1433.

Waits, L. P., Luikart, G., & Taberlet, P. (2001). Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Molecular Ecology, 10*(1), 249-256.

WCS. (n.d.a). *La Amistad*. https://programs.wcs.org/gcf-the5greatforests/en-us/-About-The-5-Great-Forests-of-Mesoamerica/La-Amistad. Accessed on 30 May 2023.

WCS. (n.d.b). *Indio Maíz-Tortuguero*. https://programs.wcs.org/gcf-the5greatforests/es-es/-Sobre-Los-5-Grandes-Bosques-de-Mesoamerica/Indio-Ma%C3%ADz-Tortuguero. Accessed on 30 May 2023.

Wehausen, J. D., Ramey, R. 2., & Epps, C. W. (2004). Experiments in DNA extraction and PCR amplification from bighorn sheep feces: the importance of DNA extraction method. *Journal of Heredity*, *95*(6), 503-509.

WWF (2020). *Living Planet Report 2020-Bending the curve of biodiversity loss*. World Wildlife Fund. https://membership.zsl.org/sites/default/files/LPR%202020%20Full%20report.pdf Accessed on 17 April 2023.

Zavala-Páramo, A. S., Oyama, K., Mendoza, E., Zavala-Páramo, M. G., Pollinger, J., & Smith, T. B. (2017). Variabilidad genética en individuos en cautiverio en México de la especie en peligro de extinción *Tapirus bairdii*. *Revista mexicana de biodiversidad, 88(2),* 480-484.

Zemanova, Miriam A. "Poor implementation of non-invasive sampling in wildlife genetics studies." *Rethinking Ecology 4 (2019):* 119-132.

Table I. **List of 15 microsatellite (STR) loci, primer sequences, motifs, size ranges and annealing temperatures used in this study.** Locus: Locus Name; **Bold:** Microsatellites included in analysis of metapopulation (N:58); *Italic:* Markers discarded after unspecificity (ASB2, AF129732, AY138542,Tte01, Tter05), low amplification success (Tba15, Tte01) and HW disequilibrium & high null alleles frequencies (Tba20). Sequence 5´-3´: Sequence of primers used in this study. Dye: fluorescent dye label used, Motif: Motif type. Range: Size range previously reported. Ta(C): Annealing temperature previously reported. Reference: previous report.

Locus	Sequence 5´-3´	Dye	Motif	Range	Ta(°C)	Reference
Tter14	GATCCTCCTGTTTGCAGAT	6FAM	$(CA)_{22}$	174-208	56-58	Sanches et al. 2009
	AGCCAAATGTTTTAGACTGAG				56-58	
Tter04	CGTTAGCATGATCTCTAGACC	VIC	$(TG)_{20}$	230-264	56-58	Sanches et al. 2009
	CCAGATGAGAAGCAGGATAG				56-58	
Tter09	GGACACTCAAGTGGGTCAAG	NED	$(CAGG)_7$	168-192	56-58	Sanches et al. 2009
	AGTGTATGCTTGTCGCGC				56-58	
Tba20	AACCCAAGTTGTCCGTCAACAG	VIC	$(AC)_{15}$	224-246	65	Norton & Ashley 2004
	GCAGTTGTCTCTGACCGTGTGTTAG				65	
Tba23	ACAGTTTGTTCCCTCCAAGGTTG	6FAM	$(AC)_{14}$	198-238	53	Norton & Ashley 2004
	GCAGGTCAAATATACTGTCAGCCTGG				53	
TtGt137	ACCATATGCCAAGGGTTTTG	6FAM	$(GT)_{17}$	253-286	51-59	da Silva et al. 2009
	GCTGCCTTCATAGTGGCTTC				51-59	
Tter05	TGCCCTGATTTAGAGAAAAC	6FAM	$(GT)_{10}$	199-209	56-58	Sanches et al. 2009
	AGGAGAAGTTAGAAGGGGAA				56-58	
Tter13	CCATGCAATTAAGAGAAAGC	VIC	$(CA)_{20}$	252-282	56-58	Sanches et al. 2009
	CAGCTAAGGACAGGAAAATG				56-58	
Tter18	AGAGTGTCAGATGTCCTGCC	NED	$(CA)_7$	98-120	56-58	Sanches et al. 2009
	TGCTTTGTGTTTGAGTGTGC				56-58	
Tba15	TTGACCTTTTCATAAGCAGCC	NED	$(AC)_{19}$	214-238	53	Norton & Ashley 2004
	CCATCTCTTCCATTCCAGTTC				53	
Tte01	ATTAAGCAGATGCCAACCTGAAG	6FAM	$(AC)_{25}$	144-154	60	Norton & Ashley 2004
	CCCTGTGGTGTTTTTTGGATC				60	
Tte12	TTAGGGAAATAACAGGTCTGG	NED	$(AC)_{19}$	214-238	55	Norton & Ashley 2004
	GTTGTTTTGCATCCAAATTGG				55	
AY138542	GGCAAAACTAAGAGAACTTG	VIC	$(AC)_{18}$	171-186	52	Nielsen et al. 2008
	GATACCAAACTGGAAATGG				52	
AF129732	CATGTGAAATGGACCGTCAGG	6FAM	$(CA)_{21}$	215-220	52	Nielsen et al. 2008
	ATTTCTGGGAAGGGGCAGG				52	
ASB2	CCTTCCGTAGTTTAAGCTTC	NED	$(GT)_{24}$	222-256	60	Breen et al. 1997
	CACAACTGAGTTCTCTGATAGG				60	

Table II. **Sample information, sequence length of cytb fragment obtained and results from the BLAST search**. JFT18880 accession refers to *Tapirus bairdii cytochrome b (CYTB) gene, complete cds; mitochondrial*. Pop: Original Population. CM: Cerro de la Muerte, H: Herradura, INC: Inconclusive, P: Pilot, SJ: San Jerónimo, TM: Tenorio-Miravalles. Origin: starting material type. *Not consensus sequence achieved.

Sample ID	Pop.	Origin	Seq. Length (bp)	Match NCBI	E Value	Query coverage (%)	Pairwise identity (%)	Grade (%)
5B	CM	Feces	133	JF718880	5.78 e-61	100	100	100
13B	CM	Feces	139	JF718880	2.81 e-64	100	100	100
7B	CM	Feces	136	JF718880	1.28 e-62	100	100	100
18B	H	Feces	138	JF718880	1.00 e-63	100	100	100
21B	SJ	Feces	137	JF718880	3.60 e-63	100	100	100
34B	CM	Feces	122	JF718880	6.7 e-55	100	100	100
38B	CM	Feces	139	JF718880	2.81 e-64	100	100	100
46B	TM	Feces	148	JF718880	3.03 e-69	100	100	100
T2	P	Tissue	143	JF718880	3.85 e-68	100	100	100
S2	P	Blood	134	JF718880	1.62 e-61	100	100	100
16B	CM	Feces	146	JF718880	3.85 e-68	100	100	100
42B	SJ	Feces	145	JF718880	8.25 e-65	100	99.0	99.5
4A	INC	Tissue	145	JF718880	1.77 e-66	100	99.7	99.8
19B	SJ	Feces	147	JF718880	1.08 e-68	100	100	100
36B	CM	Feces	137	JF718880	3.60 e-63	100	100	100
T1	P	Tissue	140	JF718880	7.88 e-65	100	100	100
86C	CM	Feces	146	JF718880	1.79 e-66	100	99.3	99.7
50B	CM	Feces	146	JF718880	3.85 e-68	100	100	100
53B	CM	Feces	146	JF718880	3.85 e-68	100	100	100
78C	TM	Hairs	147	JF718880	1.08 e-68	100	100	100
88C	CM	Feces	146	JF718880	3.85 e-68	100	100	100
91C	TM	Feces	145	JF718880	1.37 e-67	100	100	100
71C	TM	Feces	147	JF718880	1.08 e-68	100	100	100
72C	TM	Feces	123	JF718880	1.88 e-55	100	100	100
94C	TM	Feces	140	JF718880	7.88 e-65	100	100	100
93C	TM	Feces	146	JF718880	1.79 e-66	100	99.3	99.7
66B	CM	Feces	148	JF718880	3.92 e-68	100	99.7	99.8
95C	CM	Feces	114*	JF718880	1.70 e-50	100	100	100
12B	CM	Feces	117*	JF718880	3.80 e-52	100	100	100

Table III. Chi square values and exact p-values obtained for the HW test at the metapopulation (N=58) and population levels (N=12-27) for 8 microsatellite loci. CM: Cerro de la Muerte. H-SJ: Herradura-San Jerònimo TM: Tenorio Miravalles. Chi: Chi square value obtained. P exact: Probability value obtained after X. * HW desiquilibrium (p<0.05).

	Metapopulation (N=58)		CM (N=27)		H-SJ (N=12)		TM (N=19)	
Marker	**Chi**	**P value**	**Chi**	**P value**	**Chi**	**P value**	**Chi**	**P value**
Tte12	21.9729	0.0000*	13.8005	0.0014*	5.4561	0.1852	6.4653	0.1363
Tter09	183.6372	0.0000*	74.8800	0.0095*	13.9592	0.0682	18.4727	0.0846
Tter14	167.4189	0.0048*	69.0719	0.1439	58.0000	0.0692	54.6256	0.0000*
Tba23	61.0822	0.0046*	48.7972	0.0047*	3.8901	0.7672	1.6605	1.0000
Tter18	2.5173	1.0000	3.3075	1.0000	0.2449	1.0000	0.1396	1.0000
TtGt137	180.1981	0.0001*	56.8764	0.0041*	5.5739	0.1517	19.1328	0.0405*
Tter04	15.2256	0.0020*	11.6759	0.0877	1.7755	0.6008	7.0377	0.0492*
Tba20	85.8649	0.0000*	33.2708	0.0000*	12.0000	0.0011*	38.0000	0.0000*

Table IV. Estimates of genetic diversity of seven loci used for genotyping 58 samples of *T. bairdii*. Locus: microsatellite identity; Motif: motif type; Ta (C): Annealing temperature used in this study. Size range: observed allele size variation; Na: number of alleles observed; Ar: Allelic richness, Ho: observed heterozygosity; Hs: expected heterozygosity; Fis: Inbreeding index. $P_{(ID)}$unibiased: probability of two unrelated individuals having the same genotype in a population after sample size correction.

Locus	Motif	Ta (C)	Size range	Na	Ar	Ho	Hs	Fis	$P_{(ID)}$unbiased
Tte12	$(AC)_{19}$	55-53	154-162	4	2.63	0.78	0.58	-0.36	0.23
Tter09	$(CAGG)_{7}$	58-56	86-196	13	5.91	0.33	0.52	0.36	0.04
Tter14	$(CA)_{22}$	58-56	172-218	16	8.12	0.70	0.89	0.21	7.46 e-04
Tba23	$(AC)_{14}$	55-53	210-230	7	3.97	0.59	0.69	0.15	9.72 e-05
Tter18	$(CA)_{7}$	58-56	88-126	8	3.61	0.31	0.28	-0.12	4.60 e-05
TtGt137	$(GT)_{17}$	59-53	218-280	8	3.64	0.43	0.56	0.23	1.18 e-05
Tter04	$(TG)_{20}$	58-56	229-233	3	2.55	0.47	0.59	0.21	3.00 e-06
Summary			86-280	8.43 (±4.65)	4.35 (±2.00)	0.52 (±0.18)	0.59 (±0.18)	0.10 (±0.25)	9.77 e-25

Table V. Estimates of genetic diversity in three populations of *T. bairdii*. N: Population size. Ad: Allelic diversity (N alleles/ N Loci). Ar: Allelic richness. Pa: Private alleles. Ho: Observed heterozygosity. Hs: Expected heterozygosity. Fis: Inbreeding Index.

Population	N	Ad	Ar	Pa	Ho	Hs	Fis
Cerro de la Muerte (CM)	27	7.29	5.59	19	0.58	0.65	0.09
Herradura-San Jerónimo (H-SJ)	12	4.43	4.43	4	0.51	0.60	0.12
Tenorio-Miravalles (TM)	19	4.00	3.57	4	0.46	0.51	0.09

Table VI. Pairwise comparisons of F statistics between region and population levels. Hierarchichal level: 1st: regional level. 2nd: population level. Fst: fixation index. G'st Nei: Nei's standardized Gst. G'st Hedrick: Hedrick standardized Gst. Dest: Jost's D. NWT: Northwest Talamanca region. TMBC: Tenorio-Miravalles Biological Corridor region. CM: Cerro de la Muerte population. H-SJ: Herradura-San Jerónimo population. TM: Tenorio-Miravalles population. *p<0.05; **p<0.001.

Hierarchical level	Pairs compared	F_{ST}	G'_{ST} Nei	G'_{ST} Hedrick	Dest
1st	NWT & TMBC	0.0413**	0.0587**	0.1122*	0.0845*
2nd	CM & H-SJ	0.0294*	0.0239*	0.0528*	0.0412
2nd	CM & TM	0.0583**	0.0874**	0.1723*	0.1327**
2nd	H-SJ & TM	0.0233	0.0078	0.0135*	0.0096

Table VII. Hierarchical Analysis of Molecular Variance (AMOVA) based on microsatellite data (N=7) for *T.bairdii* (N=58) in Costa Rica. Based on pairwise comparisons. DF: degrees of freedom, SSD sum of squares, VC variance component, and % V percent of variance. Φ Phi: Phi estimates. P value: Probability value associated. *p<0.05; **p<0.001.

		DF	SSD	VC	% V	Φ Phi	P value
Between regions	ϕCT	1	8.6694	0.1563	6.0157	0.0602	0.0001**
Among populations within regions	ϕSC	1	4.2682	0.1170	4.5044	0.0479	0.0486*
Within populations	ϕST	55	127.8297	2.3242	89.4799	0.1052	0.3386
Total		57	140.7672	2.5974	100.00		

Table VIII. Studies that estimated tapir genetic diversity from microsatellite markers. Ad: Allelic diversity (N alleles/ N Loci). Ho: Observed Heterozygosity. L: Number of Loci. N: Sample size.

Species	Ad	Ho	L	N	Material Collected	Study area	Reference
T. bairdii	8.43	0.52	7	58	Feces, hairs, blood, tissue.	NWT & TMBC regions of Costa Rica	*Present study*
T. bairdii	8.	0.82	6	9	Feces.	Captive animals	Závala-Páramo et al. (2017).
T. bairdii	3.8	0.39	6	33	Hairs, tissue, blood	Belize, Costa Rica & Panama	Norton & Ashley (2004a)
T. bairdii	3.2	0.48	6	20	Hairs, tissue, blood	North America (captive)	Norton & Ashley (2004b)
	3.3	0.39		17		Central America (captive)	
T. indicus	3.1	0.31	9	67	Blood, tissue, hairs	Peninsular Malaysia	Lim et al. (2022).
T. terrestris	6.6	0.77	5	32	Feces	Central Amazon	Pinho et al. (2014).
T. terrestris	5.0	0.67	10	41	Blood	Argentina	da Silva et al. (2010).
T. terrestris	7.4	0.46	9	24	Blood, hairs	Brazil	Sanches et al. (2009).
T. terrestris	8.0	0.76	5	37	Tissue	French Guiana	De Thoisy et al. (2010).

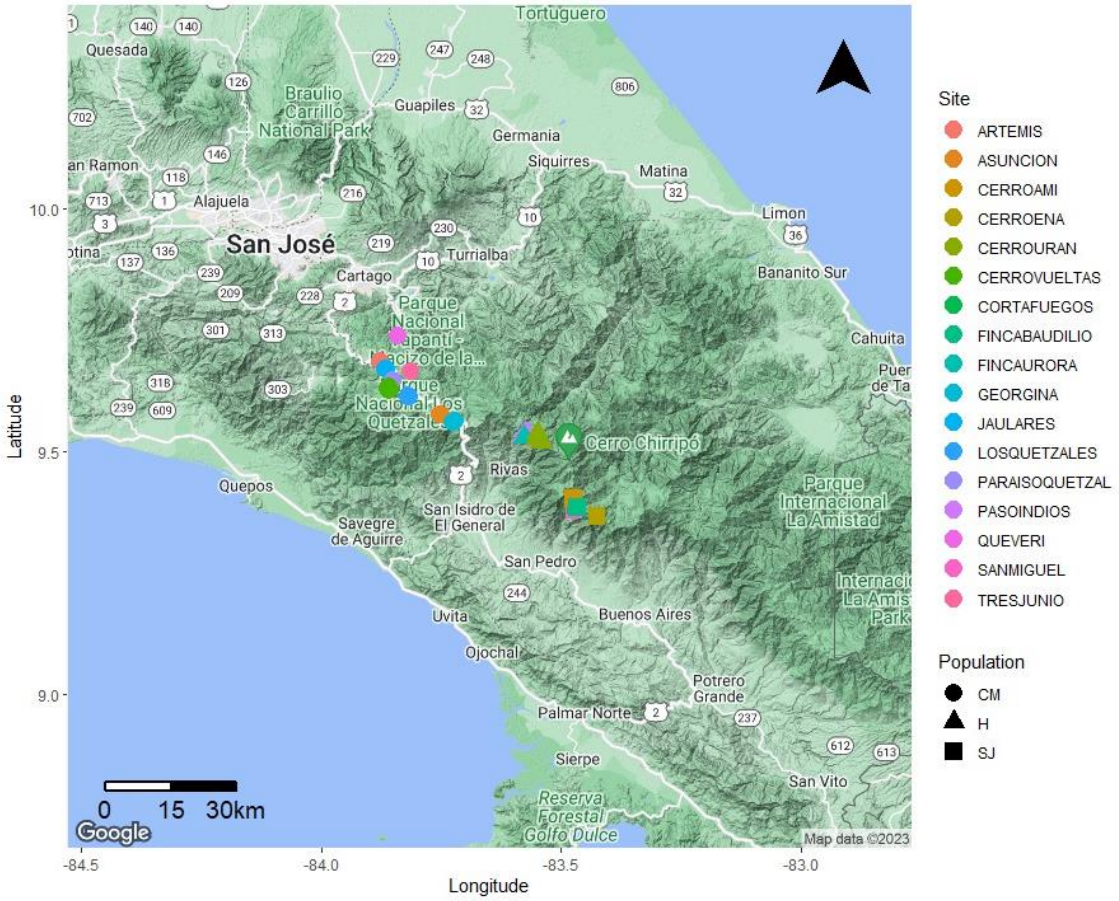

Fig.1. **Location of sampling collection points in NWT region**. 17 different sites of collection were examined in NWT region. Population: Cerro de la Muerte (CM), Herradura (H) and San Jerónimo (SJ).

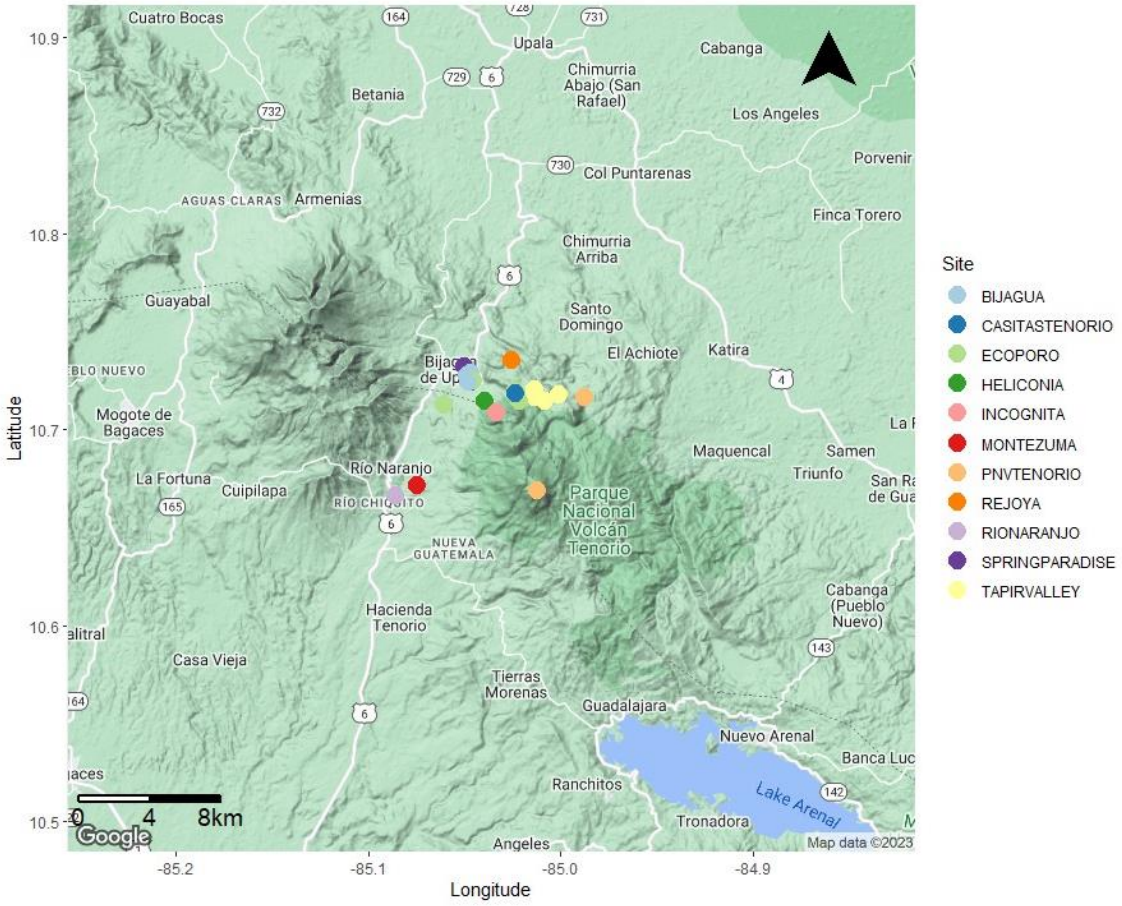

Fig.2. **Location of sampling collection points in TMBC region**. 11 different sites of collection were examined in TMBC region.

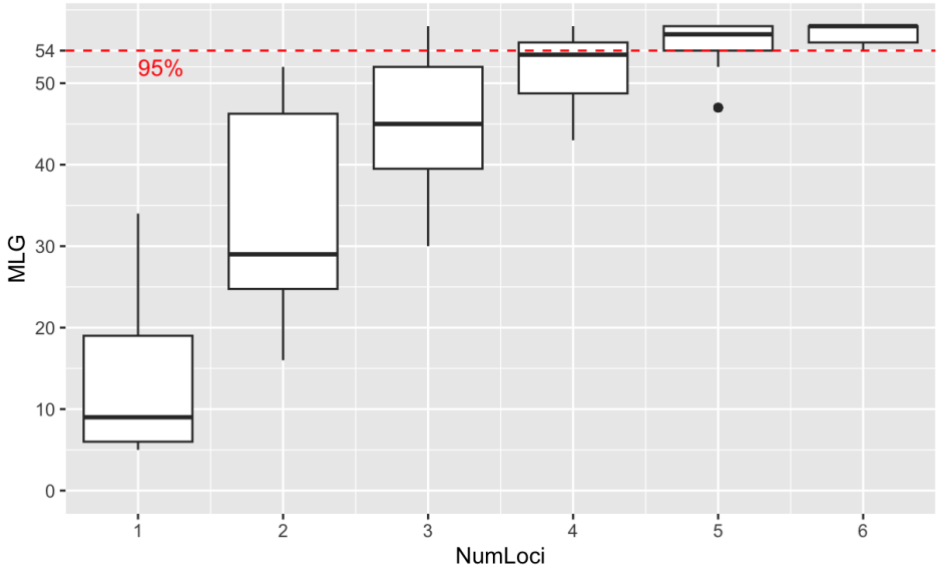

Fig.3. **Genotype cumulative curve for seven microsatellite markers characterized in 58 Central American tapir samples**. More than 54 (95%) of the total number of multiloci genotypes (MLGs, N=58) could be distinguished using six or more random markers from the set (NumLoci).

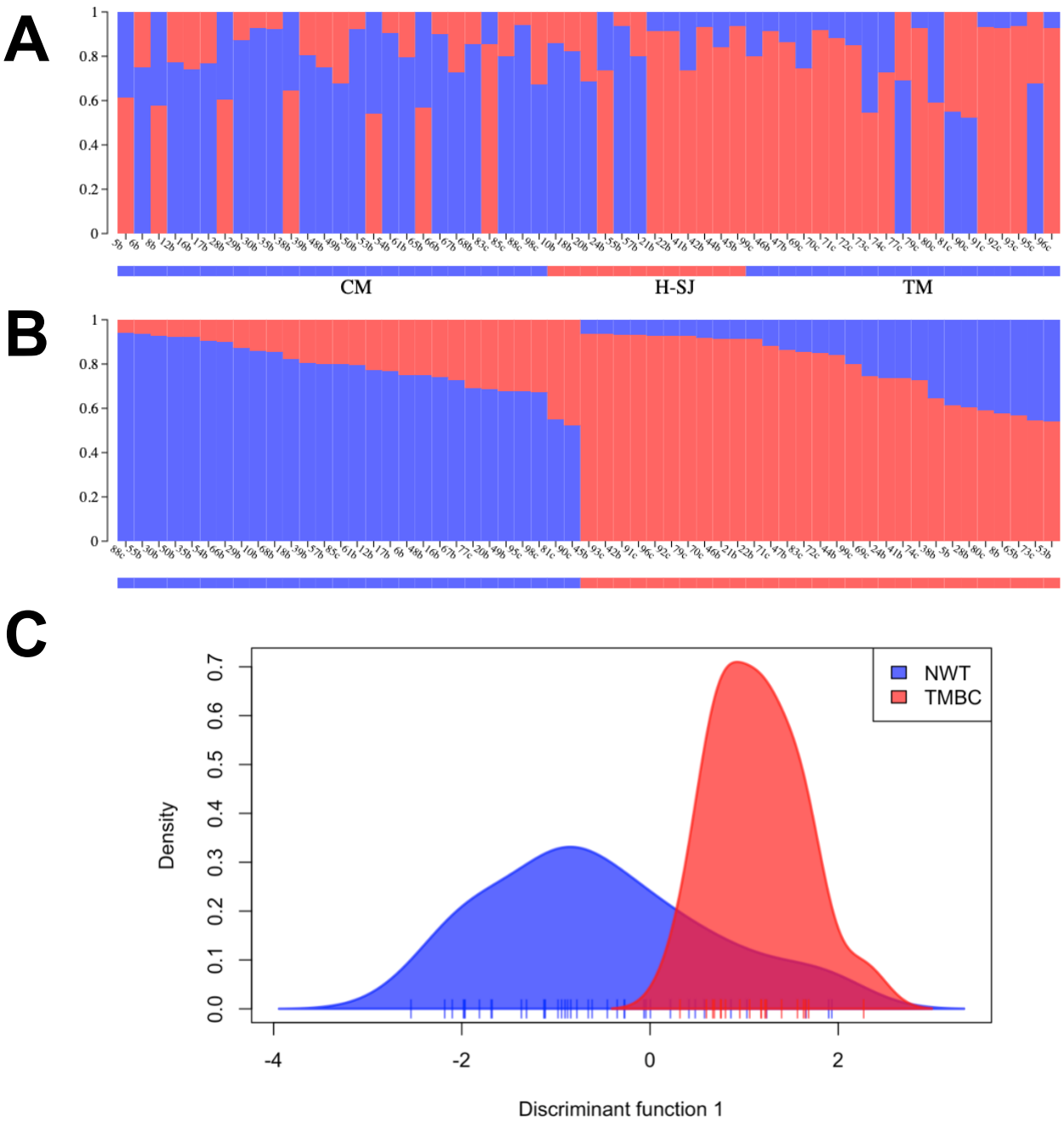

Fig. 4. **A)** Bayesian assignment performed by STRUCTURE for 3 populations in Costa Rica for K=2. Each vertical bar representes and individual and is divided proportionally to the probability of assignment of each individual to each genetic cluster (Q values). Sorted by original order of individuals in predefined populations CM, H-SJ and TM **B)** Bayesian assignment performed by STRUCTURE for 3 populations in Costa Rica for K=2. Each vertical bar represents an individual and is divided proportionally to the probability of assignment of each individual to each genetic cluster (Q values). Sorted by Q **C)** DAPC plot of NWT and TMBC regions, from 7 microsatellite loci data generated in this study.

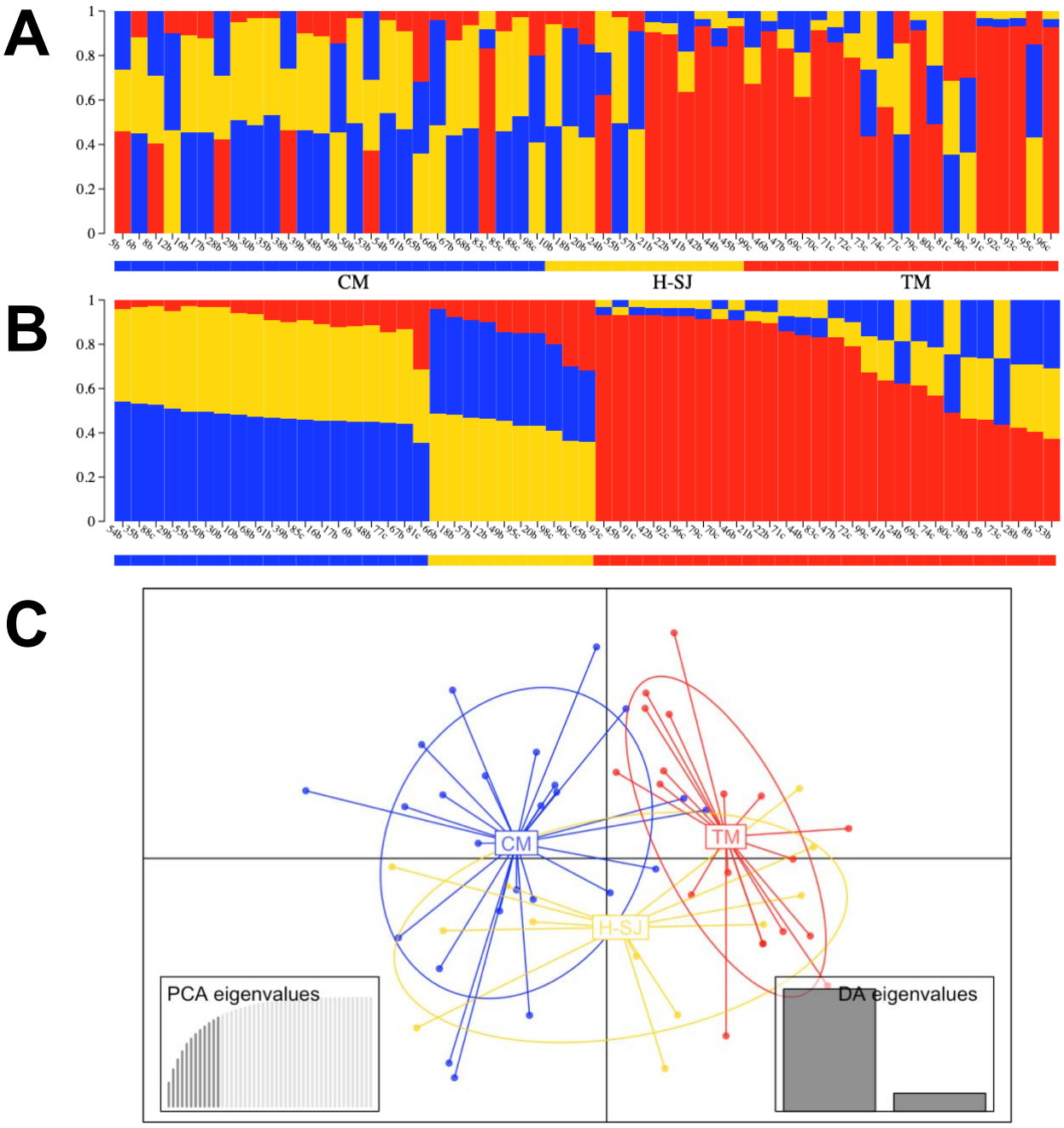

Fig. 5, A) Bayesian assignment performed by STRUCTURE for 3 populations in Costa Rica for K=3. Each vertical bar represents an individual and is divided proportionally to the probability of assignment of each individual to each genetic cluster (Q values). Sorted by original order of individuals in predefined populations CM, H-SJ and TM **B)** Bayesian assignment performed by STRUCTURE for 3 populations in Costa Rica for K=2. Each vertical bar representes and individual and is divided proportionally to the probability of assignment of each individual to each genetic cluster (Q values). Sorted by Q **C)** DAPC plot of CM, H-SJ and TM populations, from 7 microsatellite loci data generated in this study.

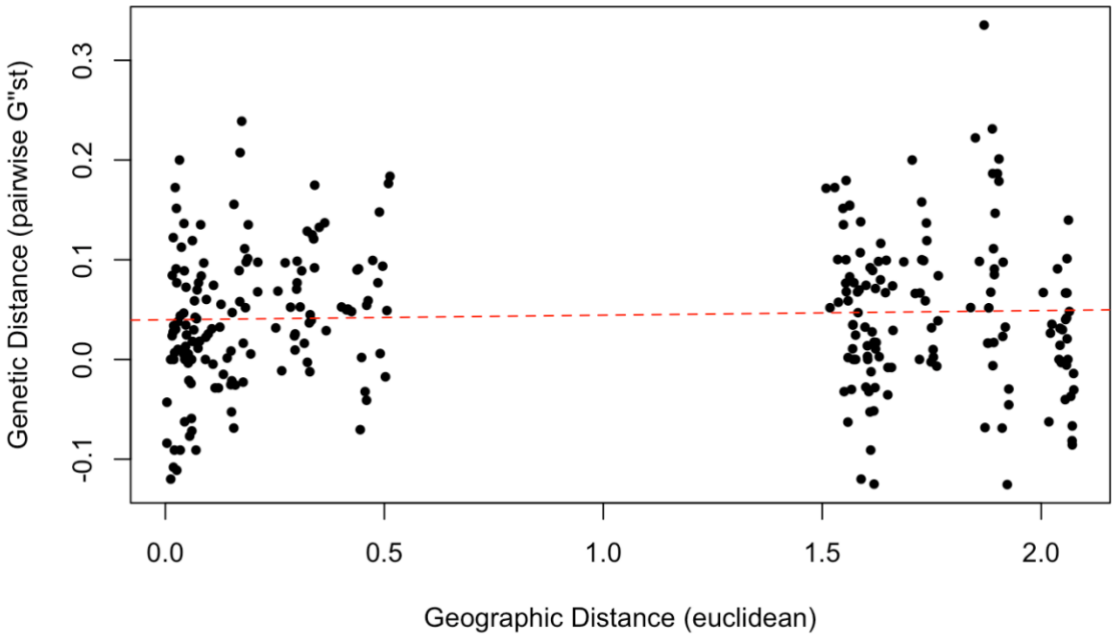

Fig. 6. **Genetic distance between samples did not correlate to the geographic distance between collection sites**. When NWT and TMBC are analyzed together, the IBD hypothesis is rejected (rMantel= 0.0489, 10^4 permutation, p= 0.2076).

Supplementary Material

Table SI. **List of 104 DNA samples isolated and purified with their respective identity**. Origin: type of starting material; Conc DNA: DNA concentration obtained via spectrophotometer (ng/ul). 260/280: absorbance spectra ratio observed between 260/280 nm. 260/230; absorbance spectra ratio observed between 260/230 nm; Population: Predefined group where the sample was collected.

# Sample	Sample ID	Origin	Conc DNA	260/280	260/230	Population
1	1a	Hairs	9,5	1,58	1,48	INC
2	2a	Hairs	6,3	1,48	0,58	CM
3	3a	Tissue	144,3	1,82	1,81	INC
4	4a	Tissue	126,6	1,79	1,81	INC
5	5a	Feces	12,6	2,28	1,63	CM
6	6a	Feces	9,2	2,55	2,25	CM
7	7a	Feces	18,1	2,12	2,72	CM
8	8a	Feces	2,1	43,00	1,75	CM
9	9a	Feces	12,7	2,35	3,04	H
10	10a	Feces	2	2,86	1,15	H
11	5b	Feces	15,5	1,94	3,70	CM
12	6b	Feces	16,9	1,95	2,23	CM
13	7b	Feces	28,7	1,87	2,15	CM
14	8b	Feces	1,3	2,18	0,79	CM
15	9b	Feces	7,7	2,04	4,84	H
16	10b	Feces	21,3	0,89	0,40	H
17	11b	Feces	67	1,95	1,41	CM
18	12b	Feces	110,6	2,03	1,75	CM
19	13b	Feces	42,4	1,96	1,27	CM
20	14b	Feces	79,8	1,92	1,48	CM
21	15b	Feces	92,5	1,96	1,65	CM
22	16b	Feces	66,9	1,98	1,50	CM
23	17b	Feces	74,9	1,99	1,61	CM
24	18b	Feces	56,6	1,89	1,46	H
25	19b	Feces	59,7	2,01	1,43	SJ
26	20b	Feces	87	1,88	1,49	H
27	21b	Feces	87,6	1,95	1,66	SJ
28	22b	Feces	28,5	1,87	1,13	SJ
29	23b	Feces	54,4	1,98	1,45	H
30	24b	Feces	72,7	1,99	1,48	H
31	25b	Feces	26,1	1,81	1,08	CM
32	26b	Feces	61,9	1,98	1,57	CM

33	27b	Feces	61	1,96	1,44	CM
34	28b	Feces	53,3	1,84	1,4	CM
35	29b	Feces	20,3	1,82	0,94	CM
36	30b	Feces	115,4	1,79	1,16	CM
37	31b	Feces	63,2	1,96	1,45	CM
38	32b	Feces	60,6	1,97	1,47	CM
39	33b	Feces	37	1,99	1,25	CM
40	34b	Feces	26,7	1,9	1,04	CM
41	35b	Feces	37,8	1,99	1,35	CM
42	36b	Feces	30,5	1,9	1,14	CM
43	38b	Feces	62,2	1,94	1,52	CM
44	39b	Feces	47,6	2	1,5	CM
45	40b	Feces	111,9	1,98	1,33	CM
46	41b	Feces	102,9	1,93	1,7	SJ
47	42b	Feces	49,4	1,89	1,67	SJ
48	43b	Feces	136,2	1,87	1,42	SJ
49	44b	Feces	140,1	1,94	1,86	SJ
50	45b	Feces	36,6	1,97	1,91	SJ
51	46b	Feces	39,3	1,84	1,22	TM
52	47b	Feces	139,1	1,98	1,74	TM
53	48b	Feces	39,3	1,94	1,18	CM
54	49b	Feces	88	1,96	1,75	CM
55	50b	Feces	112,3	1,90	1,78	CM
56	51b	Feces	29	1,81	1,06	CM
57	52b	Feces	40,1	1,85	0,84	CM
58	53b	Feces	28,9	1,77	1,14	CM
59	54b	Feces	44,2	1,80	1,18	CM
60	55b	Feces	55	1,81	1,37	H
61	56b	Feces	48,9	1,82	1,48	H
62	57b	Feces	38,6	1,78	0,97	H
63	58b	Feces	107,6	1,80	1,58	H
64	59b	Feces	133,6	1,82	1,52	CM
65	60b	Tissue	29,8	1,75	1,03	CM
66	61b	Feces	73,2	1,88	1,37	CM
67	62b	Feces	66,7	1,81	1,34	CM
68	63b	Feces	49,1	1,87	1,32	CM
69	64b	Feces	186,4	1,93	1,71	CM
70	65b	Feces	68	1,82	1,52	CM
71	66b	Feces	80,5	1,84	1,53	CM
72	67b	Feces	55,7	1,8	1,36	CM

73	68b	Feces	56,4	1,86	1,53	CM
74	69c	Feces	38	1,95	2,3	TM
75	70c	Feces	50,7	1,86	2,38	TM
76	71c	Feces	70,5	1,9	2,6	TM
77	72c	Feces	50,6	1,9	2,48	TM
78	73c	Feces	27,6	2,07	3,54	TM
79	74c	Feces	30	2,08	2,77	TM
80	75c	Blood	16,1	1,92	1,74	TM
81	76c	Blood	14,2	1,84	0,97	TM
82	77c	Blood	14,8	1,54	0,49	TM
83	78c	Hairs	30,5	1,68	1,19	TM
84	79c	Hairs	23,5	1,78	0,67	TM
85	80c	Hairs	98,9	1,87	1,39	TM
86	81c	Blood	9,3	1,9	0,45	TM
87	82c	Hairs	23	1,8	0,87	CM
88	83c	Feces	64,6	1,9	1,65	CM
89	84c	Feces	48,5	1,85	1,3	CM
90	85c	Feces	59,5	1,83	1,48	CM
91	86c	Feces	63,2	1,89	1,37	CM
92	87c	Feces	58,5	1,84	1,2	CM
93	88c	Feces	72,5	1,81	1,41	CM
94	89c	Feces	54,5	1,79	1,47	TM
95	90c	Feces	57,9	1,82	1,42	TM
96	91c	Feces	39,5	1,86	1,44	TM
97	92c	Feces	37,6	1,84	1,5	TM
98	93c	Feces	59,6	1,7	1,06	TM
99	94c	Feces	98	1,8	1,68	TM
100	95c	Feces	43,9	1,83	1,22	TM
101	96c	Hairs	79,3	1,65	1,17	TM
102	97c	Feces	72,5	1,82	1,38	CM
103	98c	Tissue	58,1	1,78	1,29	CM
104	99c	Blood	20,8	1,46	0,37	TM

Table S2. Loci name, motif, presence of null alleles and four different estimations of null alleles frequencies in the metapopulation (N=58).
*Null allele frequency > 0.20.

Locus	Motif	Null Alleles	Oosterhout	Brookfield1	Brookfield2	Chakraborty
Tte12	Dinucleotidic	No	-0.2607	-0.1291	0	-0.1484
Tter09	Tetranucleotidic	Yes	0.2095*	0.1597	0.1597	0.2633*
Tter14	Dinucleotidic	Yes	0.0858	0.0823	0.0823	0.0964
Tba23	Dinucleotidic	Yes	0.1067	0.0923	0.0923	0.1251
Tter18	Dinucleotidic	No	-0.1767	-0.0289	0	-0.0579
TtGt137	Dinucleotidic	Yes	0.1188	0.0936	0.0936	0.1463
Tter04	Dinucleotidic	No	0.0842	0.0725	0.0725	0.1096
Tba20	Dinucleotidic	Yes	0.3214*	0.2451*	0.2451*	0.7156*

Table S3. Index of association (I_a) & rBarD values with their respective probability values (p-value) after Monte Carlo test with 10^4 iterations calculated for all pairs of loci (N=7) in the metapopulation (N=58) of tapirs. * LD signal ($p<0.05$).

Microsatellite pair	I_a	p-value	rbarD	p-value
TTE12:TTER09	0.0385	0.2100	0.0410	0.1990
TTE12:TTER14	0.0016	0.4670	0.0016	0.4670
TTE12:TBA23	0.0665	0.0890	0.0675	0.0880
TTE12:TTER18	-0.1627	1.0000	-0.1640	1.0000
TTE12:TGT137	-0.0122	0.5690	-0.0126	0.5730
TTE12:TTER04	0.0150	0.3780	0.0154	0.3730
TTER09:TTER14	0.0658	0.0450	0.0722	0.0320*
TTER09:TBA23	0.1624	0.0050	0.1654	0.0040*
TTER09:TTER18	0.1031	0.1200	0.1156	0.0920
TTER09:TGT137	0.1859	0.0030	0.1868	0.0030*
TTER09:TTER04	0.0174	0.3820	0.0176	0.3820
TTER14:TBA23	0.0247	0.2270	0.0255	0.2220
TTER14:TTER18	0.0707	0.0730	0.0707	0.0760
TTER14:TGT137	-0.0046	0.5510	-0.0049	0.5560
TTER14:TTER04	0.0783	0.0150	0.0819	0.0120*
TBA23:TTER18	0.1756	0.0040	0.1833	0.0030*
TBA23:TGT137	0.1950	0.0020	0.1958	0.0020*
TBA23:TTER04	-0.0463	0.8660	-0.0464	0.8660
TTER18:TGT137	0.0971	0.0970	0.1045	0.0840
TTER18:TTER04	0.1080	0.0540	0.1147	0.0480*
TGT137:TTER04	0.0211	0.3380	0.0211	0.3370
TTE12:TTER09	0.0385	0.2100	0.0410	0.1990
TTE12:TTER14	0.0016	0.4670	0.0016	0.4670
TTE12:TBA23	0.0665	0.0890	0.0675	0.0880
TTE12:TTER18	-0.1627	1.0000	-0.1640	1.0000
TTE12:TGT137	-0.0122	0.5690	-0.0126	0.5730
TTE12:TTER04	0.0150	0.3780	0.0154	0.3730
TTER09:TTER14	0.0658	0.0450	0.0722	0.0320*

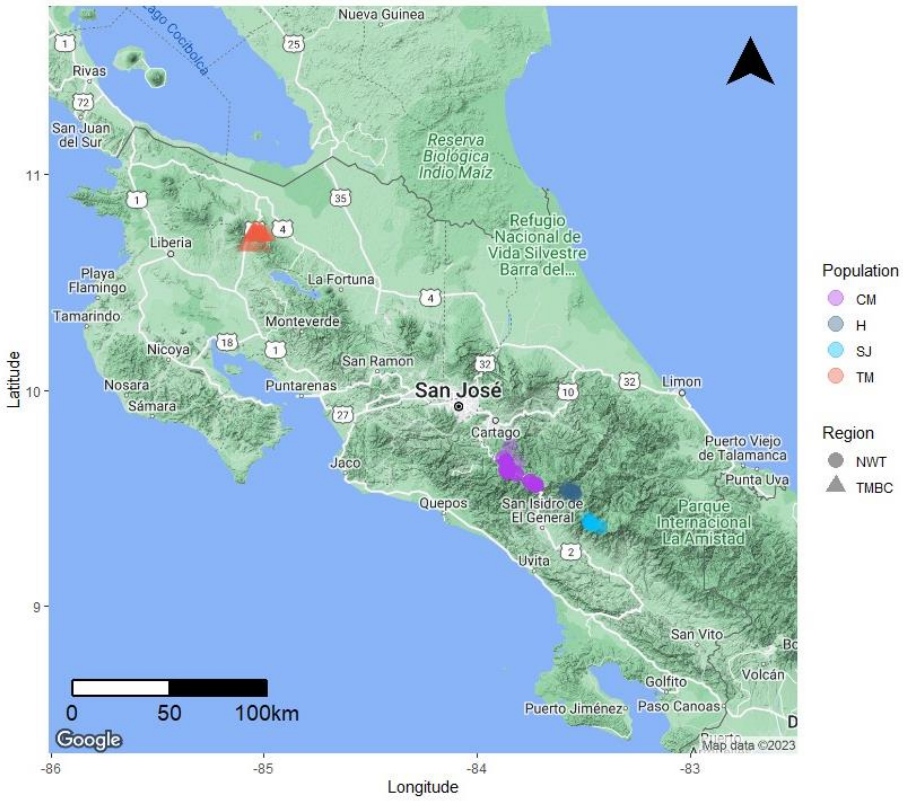

Fig.S1. **Location of sampling collection points in Costa Rica**. Population: Cerro de la Muerte (CM), Herradura (H), San Jerónimo (SJ) and Tenorio-Miravalles (TM). Region: Northwest of Talamanca Mountain Range (NWT) and Tenorio-Miravalles Biological Corridor (TMBC).

Fig.S2. **A) A** fecal sample found in the Cerro de la Muerte population area, next to a collection kit. Subsampling in the field is highly recommended. **B)** Hairs from a wandering tapir found during a field trip. **C)** Collection of samples in the wild was supported by wildlife monitoring and research groups composed of local communities' members. **D)** Immobilization of tapirs allowed us to retrieve blood and hairs from wandering tapirs in TMBC. Photo: Galdric Mossoll.

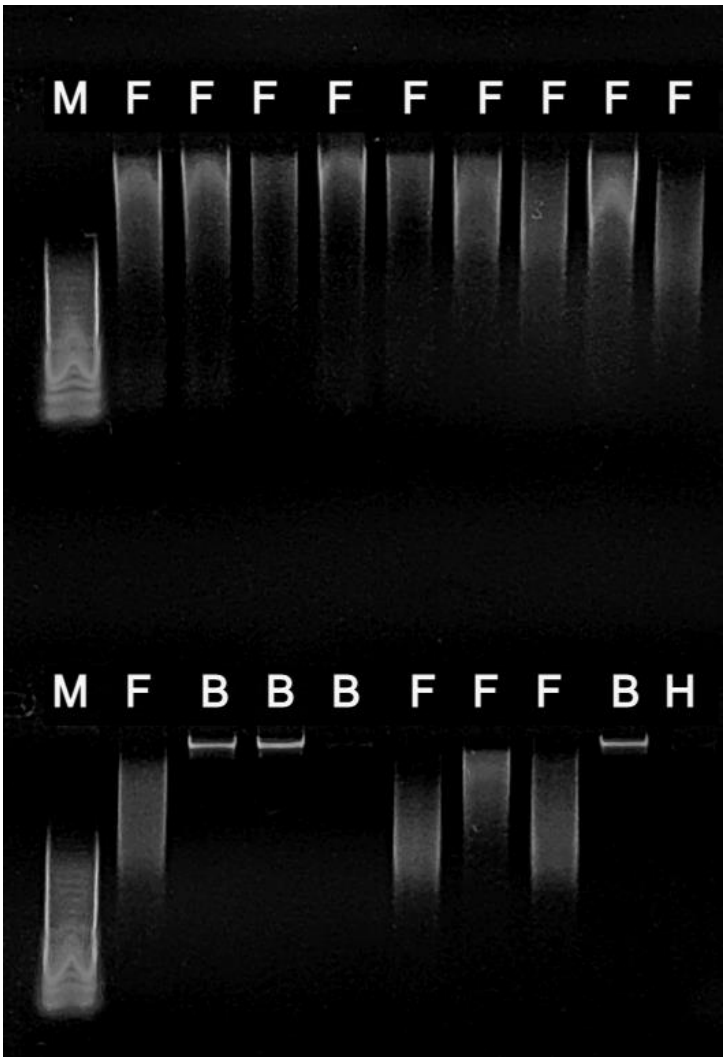

Fig. S3. **1.5% agarose gel of DNA samples extracted from feces (F), blood (B) and hairs (H).** All fecal samples displayed considerable levels of degradation compared to blood samples. M=Molecular ladder.

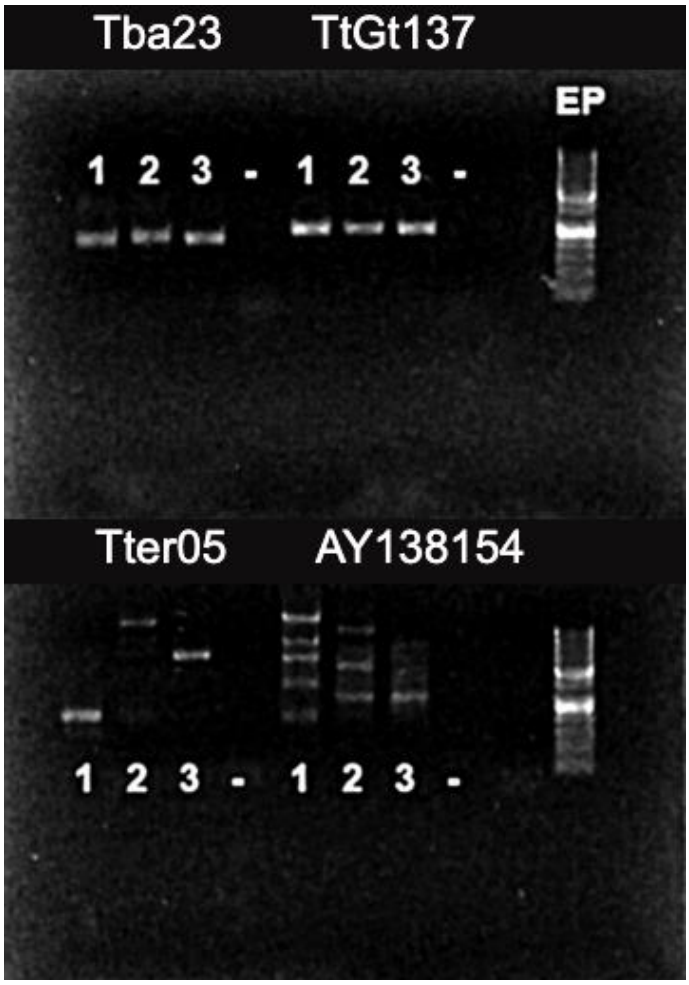

Fig. S4. **1.5% agarose gel of PCR products from three samples (1, 2, 3**). Tba23 and TtGt137 displayed specific amplification, whilst Tter05 and AY138154 presented several unspecific bands. EP: DNA ladder. (-): Negative control.

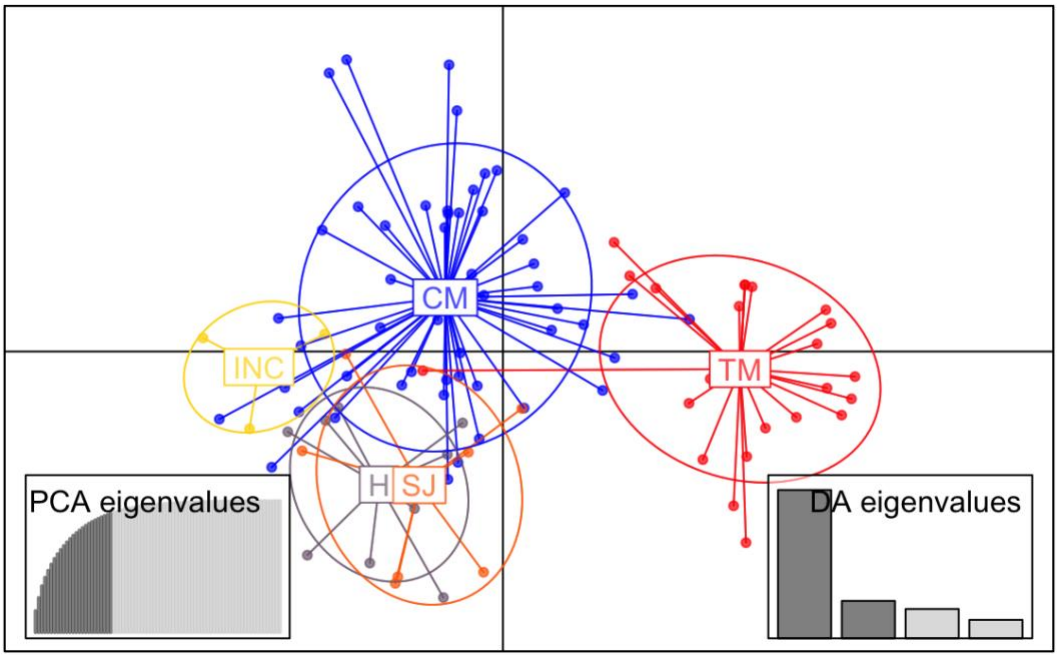

Fig. S5. **DAPC (PCA=25, DA=3) plot of five populations defined *a priori*** (Cerro de la Muerte: CM, N=53; Herradura: H, N=10; Inconclusive: INC, N=3; San Jerónimo: SJ, N=8; Tenorio-Miravalles: TM, N=24) from 8 microsatellite loci data generated in this study.

Conclusiones

América Latina es la región del mundo con la mayor pérdida de biodiversidad (alrededor del 90%) desde 1970 (WWF 2020). La información disponible relacionada a las especies de vida silvestre en la región latinoamericana (por ejemplo: datos ecológicos, datos morfométricos, datos genéticos) es escasa en comparación con otros regiones del mundo (Collen et al. 2008; Feeley & Silman 2010). Se ha encontrado que la cantidad de información disponible sobre especies silvestres aumenta conforme incrementa la distancia a los Trópicos (Collen et al. 2008; Feeley & Silman 2010). La falta de información relacionada con la vida silvestre en nuestra región es un desafío para la aplicación de técnicas modernas de modelaje, herramientas predictivas y para la aplicación de estrategias de manejo y conservación, como lo es el manejo *ex situ* (Feeley & Silman 2010; Ruiz-García et al. 2012; Zavalo-Páramo et al. 2017). Este sesgo podría ser potencialmente peor para las dantas en general y para *Tapirus bairdii* si se considera que: i) los grandes herbívoros son altamente susceptibles a las amenazas relacionadas con el cambio de uso de la tierra (Ripple et al. 2015; Atwood et al. al. 2020), ii) en especies grandes de vertebrados, el riesgo de extinción es impulsado por una combinación de factores ambientales y rasgos intrínsecos (Cardillo et al. 2005) y 3) *T. bairdii* ya se considera una especie en peligro de extinción con poblaciones en declive por la UICN y las instituciones ambientales de Costa Rica.

Las iniciativas de base comunitaria, apoyadas por la Academia, ONGs e Instituciones Públicas, representan una gran oportunidad para compensar la falta de información sobre la vida silvestre en los Trópicos. En el presente estudio logramos recolectar, procesar y analizar la mayor cantidad de muestras reportadas hasta la fecha para cualquier especie de tapir de América (*T. bairdii, T. pinchaque* y *T. terrestris*) analizada utilizando microsatélites. Este logro colectivo fue respaldado por el trabajo en el campo en conjunto a Grupos de Monitoreo e Investigación de Vida Silvestre, compuestos por miembros de las comunidades locales que coexisten con la vida silvestre y están interesados en su preservación. Entender y promover las interacciones positivas de las comunidades locales con la naturaleza es un componente importante para lograr una conservación efectiva, ya que las iniciativas

de conservación descendentes, que han funcionado bien en las regiones templadas y desarrolladas, a menudo se han aplicado con éxito limitado en los Trópicos (Sodhi et al. 2011; Büscher & Fletcher 2019).

Se estima que solo el 22% de los estudios de genética y vida silvestre incluyen muestras no invasivas (Zemanova 2019). La alta tasa de errores de genotipado, el alto costo de la replicación de muestras para lograr altos niveles de confianza y la dificultades técnicas para obtener genotipos reproducibles, podrían estarse reflejando en la baja implementación de muestras no invasivas en los estudios genéticos con vida silvestre (Smith & Wang 2014; Zemanova 2019). Nuestro conjunto de datos final, validado y empleado en los análisis, incluyó aproximadamente 88% de muestras no invasivas (N=51/58), lo que respalda la viabilidad de este tipo de muestras como fuente confiable de información para llevar a cabo estudios de poblacionales. Por otra parte, la genética de la conservación está migrando del uso de microsatélites al empleo de polimorfismos de un solo nucleótido o SNPs (Puckett 2017). No obstante, la implementación de los SNPs continúa siendo más costosa, por lo que los microsatélites prevalecen como una herramienta de conservación rentable y eficaz, capaz de proporcionar información crítica sobre la estructura poblacional de especies silvestres amenazadas, especialmente en países en desarrollo (Puckett 2017; Vashistha et al. 2020; Lim et al. 2022; Hauser et al. 2021).

Al menos dos estudios previos de genética poblacional de tapires a partir de muestras fecales informaron problemas en el éxito de amplificación después del aislamiento de ADN. Pinho et al. (2014) informaron un bajo éxito de amplificación de microsatélites después de solo 15 días de extracción, mientras que McCann (2015) no pudo reproducir ningún genotipo de microsatélite a partir de ADN fecal. Ambos estudios realizaron protocolos de aislamiento de ADN similares usando kits comerciales, siguiendo la mayoría de las instrucciones de los fabricantes como lo hizo nuestro Protocolo A de extracción. Encontramos que la modificación del protocolo original de aislamiento de ADN en dos pasos clave: incubación y elución (Protocolo B y C), resultó en una mejora significativa del rendimiento de ADN, pero también en la alta estabilidad del ADN, ya que los extractos de ADN con tres años

de procesamiento han producido genotipos reproducibles tanto para el gen mitocondrial *cytb* como para el panel de microsatélites validado. Por lo tanto, recomendamos el uso de muestras no invasivas en más estudios poblacionales con animales silvestres y especialmente dantas, así como la modificación de los protocolos comerciales de aislamiento de ADN cuando se trabaja con muestras fecales de vida silvestre, ya que los protocolos originales facilitados en los kits han sido generalmente diseñados y optimizados para el procesamiento de muestras humanas.

Encontramos niveles moderados de diversidad genética de microsatélites en la población de *T. bairdii* estudiada. Las tres poblaciones analizadas exhiben estimaciones de diversidad genética más altos (Ho=0.46-0.58) que los valores informados previamente para otras de dantas silvestres de América Central, incluso cuando la mayoría de las muestras recolectadas provinieron de sitios de recolección fuera de las áreas protegidas. Tanto el Noroeste de la Cordillera de Talamanca (NWT) como el Corredor Biológico Tenorio Miravalles (TMBC) son reconocidas como regiones clave para la conectividad y preservación de las poblaciones silvestres de tapires en Costa Rica (Schank 2017, Schank 2020). Además, NWT es parte del Bosque La Amistad, mientras que TMBC está muy cerca del Bosque Indio Maíz-Tortuguero, dos de las áreas silvestres más grandes de Mesoamérica y componentes esenciales de la conectividad en Centroamérica (WCS n.d.a, n.d.b., García et al. 2016) . La diversidad genética relativamente alta observada en todas las poblaciones analizadas puede reflejar la proximidad a áreas protegidas, reconocidas como los centros de población núcleo más importantes para la especie en el país (García et al. 2016).

El análisis de la estructura genética es consistente con la fuerte influencia de ambas áreas núcleo, porque la estratificación regional explica la mayor parte de la estructura poblacional encontrada. Sin embargo, el análisis de AMOVA, las comparaciones pareadas de estadísticas F entre poblaciones y la presencia de alelos privados en cada grupo apoyan la estratificación a nivel de población La estructura genética marginal encontrada entre las poblaciones de Herradura-San Jerónimo (SJ) y Tenorio-Miravalles (TM), junto la moderada diferenciación entre

Cerro de la Muerte (CM) y TM (más alta que la diferenciación entre NWT y TMBC), sugieren una vía candidata de alto flujo génico entre TM y H-SJ a través de la vertiente Atlántica y sus zonas protegidas y rutas de conectividad; mientras que se sugiere una vía de flujo génico más débil entre TM y CM a través de la vertiente del Pacífico. El muestreo de poblaciones adicionales de las vertientes del Atlántico y el Pacífico de Costa Rica debe ser una prioridad para poder realizar análisis genéticos poblacionales y del paisaje adicionales que permitan validar la hipótesis de flujo genético asimétrico entre ambas vertientes del país, así como para comprender mejor los patrones de estructura genética en la metapoblación de tapires silvestres de Costa. Rica.

En el presente estudio: 1) validamos un panel de marcadores moleculares (cytb + siete loci de microsatélites) adecuados para la identificación molecular de muestras (cytb) y para la descripción de la diversidad y estructura genética (loci de microsatélites) de *T. bairdii* usando muestras de tejido, sangre, pelos y heces, 2) estimamos la diversidad y estructura genética de tres poblaciones silvestres de tapires de Costa Rica por primera vez, 3) evaluamos los niveles de endogamia en las poblaciones costarricenses analizadas y 4) demostramos que las distancias genéticas encontradas entre las poblaciones no son causadas por aislamiento por distancia (IBD).

Bibliografía

Allendorf, F. W., & Luikart, G. (2009). *Conservation and the genetics of populations*. John Wiley & Sons.

Atwood, T. B., Valentine, S. A., Hammill, E., McCauley, D. J., Madin, E. M., Beard, K. H., & Pearse, W. D. (2020). Herbivores at the highest risk of extinction among mammals, birds, and reptiles. *Science advances*, *6*(32), eabb8458.

Bello, C., Galetti, M., Pizo, M. A., Magnago, L. F. S., Rocha, M. F., Lima, R. A., ... & Jordano, P. (2015). Defaunation affects carbon storage in tropical forests. *Science advances*, *1*(11), e1501105.

Brooks, D. M., Bodmer, R. E., & Matola, S. (Eds.). (1997). *Tapirs: status survey and conservation action plan* (Vol. 38). IUCN.

Broquet, T., Ménard, N., & Petit, E. (2007). Noninvasive population genetics: a review of sample source, diet, fragment length and microsatellite motif effects on amplification success and genotyping error rates. *Conservation Genetics, 8(1),* 249-260.

Büscher, B., & Fletcher, R. (2019). Towards convivial conservation. *Conservation & Society, 17(3)*, 283-296.

Cardillo, M., Mace, G. M., Jones, K. E., Bielby, J., Bininda-Emonds, O. R., Sechrest, W., ... & Purvis, A. (2005). Multiple causes of high extinction risk in large mammal species. *Science*, *309*(5738), 1239-1241.

Chetri, M., Odden, M., Sharma, K., Flagstad, Ø., & Wegge, P. (2019). Estimating snow leopard density using fecal DNA in a large landscape in north-central Nepal. *Global Ecology and Conservation, 17*, e00548.

Collen, B., Ram, M., Zamin, T., & McRae, L. (2008). The tropical biodiversity data gap: addressing disparity in global monitoring. *Tropical Conservation Science, 1*(2), 75-88.

Crandall, K. A., Bininda-Emonds, O. R., Mace, G. M., & Wayne, R. K. (2000). Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution,* 15(7), 290-295.

de Barba, M., Waits, L. P., Garton, E. O., Genovesi, P., Randi, E., Mustoni, A., & Groff, C. (2010). The power of genetic monitoring for studying demography, ecology and genetics of a reintroduced brown bear population. *Molecular Ecology, 19(18),* 3938-3951.

Feeley, K. J., & Silman, M. R. (2011). The data void in modeling current and future distributions of tropical species. *Global Change Biology, 17*(1), 626-630.

Foerster, C. R., & Vaughan, C. (2002). Home Range, Habitat Use, and Activity of Baird's Tapir in Costa Rica1. *Biotropica*, *34*(3), 42.

Frankham, R. (2010). Where are we in conservation genetics and where do we need to go? Conservation Genetics, 11(2), 661-663.

Garcìa, M., Jordan, C., O'Farril, G., Poot, C., Meyer, N., Estrada, N., Leonardo, R., Naranjo, E., Simons, Á., Herrera, A., Urgilés, C., Schank, C., Boshoff, L. & Ruiz-Galeano, M. (2016). *Tapirus bairdii*. *The IUCN Red List of Threatened Species* 2016:e.T21471A45173340. https://dx.doi.org/10.2305/IUCN.UK.2016-1.RLTS.T21471A45173340.en Accesado el 17 de abril 2023.

González-Maya, J. F., Martínez-Meyer, E., Medellín, R., & Ceballos, G. (2017). Distribution of mammal functional diversity in the Neotropical realm: Influence of land-use and extinction risk. *PloS one, 12(4),* e0175931.

Harrison, S., & Hastings, A. (1996). Genetic and evolutionary consequences of metapopulation structure. *Trends in Ecology & Evolution, 11(4*), 180-183.

Hauser, S. S., Athrey, G., & Leberg, P. L. (2021). Waste not, want not: microsatellites remain an economical and informative technology for conservation genetics. *Ecology and Evolution, 11(22),* 15800-15814.

Hibert, F., Sabatier, D., Andrivot, J., Scotti-Saintagne, C., Gonzalez, S., Prévost, M. F., ... & Richard-Hansen, C. (2011). Botany, genetics and ethnobotany: a crossed investigation on the elusive tapir's diet in French Guiana. *PLoS One*, *6*(10), e25850.

Isaac, N. J., Turvey, S. T., Collen, B., Waterman, C., & Baillie, J. E. (2007). Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PloS one*, *2*(3), e296.

Janzen, D. H. (1982). Seeds in tapir dung in Santa Rosa National Park, Costa Rica. *Brenesia*.

Keller, L. F., & Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in Ecology & Evolution, 17(5)*, 230-241.

Lacy, R. C. (1987). Loss of genetic diversity from managed populations: interacting effects of drift, mutation, immigration, selection, and population subdivision. *Conservation Biology, 1(2),* 143-158.

Lande, R. (1998). Anthropogenic, ecological and genetic factors in extinction and conservation. *Researches on Population Ecology, 40(3),* 259-269.

Lim, Q. L., Yong, C. S. Y., Ng, W. L., Ismail, A., Rovie-Ryan, J. J., Rosli, N., ... & Annavi, G. (2022). Population genetic structure of wild Malayan tapirs (Tapirus indicus) in Peninsular Malaysia revealed by nine cross-species microsatellite markers. Global Ecology and Conservation, 40, e02321.

McCann, N. P. (2015). *The conservation of Baird's Tapir (*Tapirus bairdii*) in Honduras*. Tesis doctoral, Universidad de Cardiff.

Medici, E. P., Carrillo, L., Montenegro, O. L., Miller, P. S., Carbonell, F., Chassot, O., & Mendoza, A. (2005). *Baird's tapir (Tapirus bairdii) conservation workshop population and habitat viability assessment (PHVA*). Final report. IUCN/SSC Conservation Breeding Specialist Group, Belize, Central America.

Moritz, C. (1994). Defining 'evolutionarily significant units' for conservation. *Trends in Ecology & Evolution, 9(10),* 373-375.

Naranjo-Piñera, E. (1995). Abundancia y uso de hábitat del tapir (Tapirus bairdii) en un bosque tropical húmedo de Costa Rica. *Vida Silvestre Neotropical*, *4*(1), 20-31.

Norton, J. E., & M. V Ashley. (2004a). Genetic variability and population structure among wild Baird's tapirs. *Animal Conservation, 7,* 211–220.

Norton, J. E., & Ashley, M. V. (2004b). Genetic variability and population differentiation in captive Baird's tapirs (Tapirus bairdii). *Zoo Biology: Published in affiliation with the American Zoo and Aquarium Association, 23(6),* 521-531.

Noss, R. F. (1990). Indicators for monitoring biodiversity: a hierarchical approach. *Conservation biology, 4(4),* 355-364.

O'Farrill, G., Calmé, S., Sengupta, R., & Gonzalez, A. (2012). Effective dispersal of large seeds by Baird's tapir: a large-scale field experiment. *Journal of Tropical Ecology*, *28*(1), 119-122.

O'Farrill, G., Galetti, M., & Campos-Arceiz, A. (2013). Frugivory and seed dispersal by tapirs: an insight on their ecological role. *Integrative Zoology*, 4-17.

O'Grady, J. J., Brook, B. W., Reed, D. H., Ballou, J. D., Tonkyn, D. W., & Frankham, R. (2006). Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. *Biological Conservation, 133*(1), 42-51.

Paolucci, L. N., Pereira, R. L., Rattis, L., Silvério, D. V., Marques, N. C., Macedo, M. N., & Brando, P. M. (2019). Lowland tapirs facilitate seed dispersal in degraded Amazonian forests. *Biotropica*, *51*(2), 245-252.

Pinho, G. M., da Silva, A. G., Hrbek, T., Venticinque, E. M., & Farias, I. P. (2014). Kinship and social behavior of lowland tapirs (Tapirus terrestris) in a Central Amazon landscape*. PloS One, 9*(3), e92507.

Pompanon, F., Bonin, A., Bellemain, E., & Taberlet, P. (2005). Genotyping errors: causes, consequences and solutions. *Nature reviews Genetics, 6(11),* 847.

Pritchard, J. K., Stephens, M., & P. Donnelly. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155, 945-959.*

Puckett, E. (2017) Variability in total project and per sample genotyping costs under varying study designs including with microsatellites or SNPs to answer conservation genetic questions. *Conservation Genetics Resources 9:* 289–304.

Reed, D. H., & Frankham, R. (2003). Correlation between fitness and genetic diversity. *Conservation Biology, 17(1),* 230-237.

Ripple, W. J., Newsome, T. M., Wolf, C., Dirzo, R., Everatt, K. T., Galetti, M., ... & Van Valkenburgh, B. (2015). Collapse of the world's largest herbivores. *Science advances*, *1*(4), e1400103.

Ruiz-García, M., Vásquez, C., Pinedo-Castro, M., Sandoval, S., Castellanos, A., Kaston, F., ... & Shostell, J. (2012). Phylogeography of the Mountain Tapir (Tapirus pinchaque) and the Central American Tapir (Tapirus bairdii) and the origins of the three Latin-American tapirs by means of mtCyt-B sequences. *Current topics in phylogenetics and phylogeography of terrestrial and aquatic systems. InTech, Rijeka*, 83-116.

Schank, C. J., Cove, M. V., Arima, E. Y., Brandt, L. S., Brenes-Mora, E., Carver, A., ... & Miller, J. A. (2020). Population status, connectivity, and conservation action for the endangered Baird's tapir. *Biological Conservation*, *245*, 108501.

Schank, C. J., Cove, M. V., Kelly, M. J., Mendoza, E., O'Farrill, G., Reyna-Hurtado, R., ... & Miller, J. A. (2017). Using a novel model approach to assess the distribution and conservation status of the endangered Baird's tapir. *Diversity and Distributions*, *23*(12), 1459-1471.

Smith, O., & Wang, J. (2014). When can noninvasive samples provide sufficient information in conservation genetics studies?. *Molecular Ecology Resources, 14(5),* 1011-1023.

Sodhi, N. S., Butler, R., & Raven, P. H. (2011). Bottom-up conservation. *Biotropica, 43(5*), 521-523.

Sommer, S. (2005). The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology, 2(1),* 16.

Schwartz, M. K., Luikart, G., & Waples, R. S. (2007). Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution, 22(1)*, 25-33.

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., ... & Bouvet, J. (1996). Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic acids research, 24(16),* 3189-3194.

van Oosterhout, C., Hutchinson, W. F., Wills, D. P., & Shipley, P. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Resources, 4(3),* 535-538.

Vashistha, G., Deepika, S., Dhakate, P. M., Khudsar, F. A., & Kothamasi, D. (2020). The effectiveness of microsatellite DNA as a genetic tool in crocodilian conservation. *Conservation Genetics Resources, 12*, 733-744.

Villar, N., Paz, C., Zipparro, V., Nazareth, S., Bulascoschi, L., Bakker, E. S., & Galetti, M. (2021). Frugivory underpins the nitrogen cycle. *Functional Ecology*, *35*(2), 357-368.

Waits, L. P., & Paetkau, D. (2005). Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. *The Journal of Wildlife Management, 69(4),* 1419-1433.

WCS. (n.d.a). *La Amistad*. https://programs.wcs.org/gcf-the5greatforests/en-us/-About-The-5-Great-Forests-of-Mesoamerica/La-Amistad. Accesado el 30 de mayo 2023.

WCS. (n.d.b). *Indio Maíz-Tortuguero*. https://programs.wcs.org/gcf-the5greatforests/es-es/-Sobre-Los-5-Grandes-Bosques-de-Mesoamerica/Indio-Ma%C3%ADz-Tortuguero. Accesado el 30 de mayo 2023.

WWF (2020). *Living Planet Report 2020-Bending the curve of biodiversity loss*. World Wildlife Fund. https://membership.zsl.org/sites/default/files/LPR%202020%20Full%20report.pdf Accesado el 17 de abril 2023.

Zavala-Páramo, A. S., Oyama, K., Mendoza, E., Zavala-Páramo, M. G., Pollinger, J., & Smith, T. B. (2017). Variabilidad genética en individuos en cautiverio en México de la especie en peligro de extinción *Tapirus bairdii*. *Revista mexicana de biodiversidad, 88(2),* 480-484.

Zemanova, Miriam A. Poor implementation of non-invasive sampling in wildlife genetics studies. *Rethinking Ecology 4 (2019):* 119-132.