GENETIC CHARACTERIZATION OF *Caiman crocodilus fuscus* POPULATIONS IN THE TOLIMA AND CASANARE DEPARTMENTS BASED ON MITOCHONDRIAL MARKERS

Diana M. Díaz Moreno

University of Tolima

ABSTRACT

Over-exploitation of species of the Order Crocodylia led to the creation of regulatory and management measures for their conservation. Colombia is the main producer of *Caiman crocodilus* skins. The Colombian normative established a restocking quota by which farms must release a percentage of their production. However, the lack of population assessments and knowledge regarding the genetic composition of natural populations have made this task impossible. Given that to date there are no genetic studies for the species in the Tolima Department, we characterized genetically three populations present across Tolima department. Mitochondrial genes *COI* and *CytB* were analyzed to quantify their genetic variability, population differentiation, demographic history, and genetic structure. Additionally, sequences deposited in databases were used as well as new sequences for the country (Casanare) to define haplotypic relationships at a mitochondrial scale. We identified 12 haplotypes across the Tolima department with high values of genetic diversity and a strong population differentiation. No evidence of population expansion or reduction was found, detectable with the markers used. The phylogeographic analysis showed 5 large lineages for the species and located the Tolima populations within the same clade distributed from southern Costa Rica to Tolima (*Caiman crocodilus fuscus*) with certain structure within the department. Our results show that the distribution of genetic diversity should be taken into account when carrying out releases or any type of translocation, since it could affect the haplotypic integrity and therefore the natural structure of the populations.

Keywords: Mitochondrial DNA, Crocodylia, Genetic Diversity, Phylogeography

INTRODUCTION

Caiman crocodilus fuscus is one of the most used crocodile on the planet. In countries such as Colombia, there are insufficient data to define its status at the local level (Balaguera-Reina & Velasco, 2019). Historically, there has been a lack of information on the status of the order Crocodylia in the Tolima department and the scarce information that exists is not easily accessible or is very unrepresentative (M. A. Morales-Betancourt, Fajardo-Patiño, Lasso, & De la Ossa, 2013). From the available population analyses, it is evident that it is urgent to implement different conservation and repopulation strategies in the Alto Magdalena, since there are areas where the species has practically disappeared (Ardila-Robayo, Moreno-Arias, Zúñiga, Martinez-Barreto, & Suarez-Daza, 2009). Therefore, we characterized the genetic diversity, population differentiation, phylogenetic relationships, and phylogeographic structure of *C. c. fuscus* populations in the Tolima Department through the analysis of the variability of the mitochondrial genes COI and CytB focused to generate the necessary base information to propose management and conservation recommendations for the species as well as preserve the haplotypic integrity of the subspecies. We also used 7 sequences of *Caiman crocodilus crocodilus* from Casanare department to complement the phylogenetic analysis. The knowledge generated will not only serve as a basis for future studies on the species, but will also allow the adequate design of repopulation strategies for the species in the department.

MATERIALS AND METHODS

Locations and sampling: The populations of La Mariposa lagoon (Municipality of Ibagué), the Río Viejo wetland (Municipality of San Luis), the Hidroprado Dam (Municipalities of Prado, Cunday, and Purificación), and Las Cabras lagoon (Municipality of Doima) were assessed at night to capture a

collect samples of *Caiman crocodilus fuscus*. From each individual, 1 to 2 mL of blood was collected from the caudal or nuchal vein. The collection data was uploaded to the Colombian Biodiversity Information System under report certificate No. 16C7330414C. Rafael Antelo, from the Palmarito Foundation, donated tissue samples of *Caiman crocodilus crocodilus*, from the localities of Palmarito and Wisirare in the department of Casanare. These samples were used to complement the phylogenetic analysis

Extraction and Sequencing: All DNA extractions were performed using a modified Phenol-Chloroform-Isoamyl Alcohol extraction protocol (Sambrook, Fritsch & Maniatis, 1989). The mitochondrial genes *Cytochrome Oxidase 1 (COI)* and *Cytochrome B (CytB)* were amplified using the primers of Palumbi, (1991) and Jimenez-Alonso, (2016a) respectively. The amplification products were sent to SSigMol at the National University of Colombia for sequencing via Applied Biosystems 3730XL Sanger Sequencer.

Data analysis

Genetic diversity, population differentiation and demographic history of *C. c. fuscus* in Tolima: Polymorphic sites (S), nucleotide diversity (π), haplotype diversity (h) and number of haplotypes (#H) were calculated for each population in each gene separately and concatenated, using the DnaSP 6.12 program (Rozas et al., 2019). In addition, we assessed the representativeness of the haplotypes found in the EstimateS 9.1 program (Colwell, 2014). The paired Fst statistic (R. R. Hudson, Slatkin, & Maddison, 1992) was estimated in the ARLEQUIN 3.5 program (Excoffier & Lischer, 2010), to estimate the degree of population differentiation (Arenas-Calle & García-Merchán, 2011; Caldera, Jaramillo, Cochero, Pérez-Doria, & Bejarano, 2013; Richard R. Hudson, 2000; Rozas, 2009; Rozas et al., 2019; Wright, 1951). Finally, We look for evidence of whether population expansion/contraction based on D of Tajima and D and F of Fu and Li in the DnaSP 6.12 program (Caldera et al., 2013; Fu & Li, 1993; Hartl & Clark, 1997; Rozas, 2009; Rozas et al., 2019; Tajima, 1984).

Intra and interpopulation genealogical relationships of *C. c. fuscus* in Tolima: A MUSCLE alignment was performed in the Geneious Prime program (Edgar, 2004; Kearse et al., 2012) with the concatenated sequences of the *COI* and *CytB* genes of 21 individuals from Tolima, 7 from Casanare, 1 from *Paleosuchus trigonatus* and 1 from *Alligator mississippiensis*. This matrix was divided into six partitions via DnaSP 6.12 (Rozas et al., 2019) and analyzed via JModelTest 2 (Darriba, Taboada, Doallo, & Posada, 2012). The Akaike Information Criterion (AIC) was used to establish the best model for nucleotide substitution. With this information, molecular phylogenies were performed with the Bayesian Inference (IB), Maximum Parsimony (MP), and Maximum Likelihood (MV) algorithms. IB was performed in the MrBayes 3.2 program (Ronquist et al., 2012). An additional alignment, which contained only the 28 sequences of *Caiman crocodilus* from Tolima and Casanare, was used to build a haplotype network in the PopArt program (Leigh & Bryant, 2015).

Phylogenetic and phylogeographic relationships of *C. c. fuscus* populations in Tolima: An alignment was made with the concatenated sequences of the *COI* and *CytB* genes of 45 individuals of *Caiman crocodilus*, 1 of *Paleosuchus trigonatus* and 1 of *Alligator mississippiensis* from Venegas-Anaya et al., (2008) plus the 21 *C. c. fuscus* from Tolima and the 7 *C. c. crocodilus* from Casanare for a final alignment of 75 sequences. Phylogenies were constructed by IB, MP and MV methods as before. A haplotype network was created using the TCS algorithm at PopArt (Clement, Posada, & Crandall, 2000; Leigh & Bryant, 2015). To establish phylogeographic relationships, a nested clade analysis was performed in the ANeCA 1.2 program (Panchal, 2006). Files containing clade distances (Dc) and nested clade distance (Dn) were generated in GEODIS 2.2 (D. Posada, Crandall, & Templeton, 2000). Finally, the significance of the distances and the distribution values of each locality were calculated in

ANeCA 1.2, (Panchal, 2006). Those clades with significant values were passed by the inference key of (Templeton, 2004).

RESULTS AND DISCUSSION

Genetic diversity, population differentiation and demographic history of *C. c. fuscus* in Tolima: A total of 21 sequences (7 for each locality) of high-quality COI and CytB genes were obtained for Tolima and 7 for Casanare. For both *COI* and *CytB* genes, most of the polymorphic sites were found in the third codon position, as reported by Venegas-Anaya et al., (2008). Twelve unique haplotypes were found in collected individuals of *C. c. fuscus* in Tolima and 4 unique haplotypes in Casanare of *C. c. crocodilus* (Annex 1), the first ones for this subspecies in the country. The most frequent haplotypes in Tolima were HD Ccf 08 (n = 4) the only one shared between localities (Ibague and Prado), followed by the haplotypes HD Ccf 03 (n = 3) and HD Ccf 12 (n = 4). In Casanare the most frequent haplotype was HD Ccc 15 (n = 4).

Intra- and inter-population genealogical relationships of *C. c. fuscus* in Tolima: Within Tolima, Maximum Likelihood (ML) (Appendix 2) was able to resolve all clades. This phylogeny recognizes as separate lineages the individuals of both subspecies and within these, it groups the sequences coming from the same populations showing a population differentiation. In the Tolima samples, three groups of individuals with high support values can be distinguished: the first with samples only from the Prado locality, a second group with samples from both Ibagué and Prado, and a third with sequences from San Luis and Ibagué. The presence of these groups is consistent with what was found previously, which showed high levels of population differentiation. However, this grouping makes sense if we consider that Ibagué and Prado share the most frequent haplotype.

As illustrated in the phylogenetic trees (Annex 3), San Luis has a higher haplotypic integrity than the other localities, which share the HD Ccf 8 haplotype. The Ibague haplotypes are in an intermediate position in relation to the Tolima localities, so the association between these haplotypes does not seem to be given by their geographical proximity. The populations of Ibague and Prado are the most distant from each other within our sample. Furthermore, the Prado dam represents a strong physical barrier that, in theory, should isolate the populations present there. However, the haplotypic analysis shows two possible scenarios: the HD Ccf 8 haplotype is shared by these populations since a time before the construction of the dam or the entry/exit of this haplotype occurred after this separation. The first scenario implies the existence of a spatial connectivity between these populations that is not supported by the high population structure found, however, the markers used in this study may not be the most appropriate to resolve this uncertainty. However, the short generation time that has passed since the construction of the dam (a little more than 60 years), is too short to observe a marked genetic differentiation between the two populations. With respect to the second scenario, the anthropogenic translocation is contemplated since there is evidence from multiple reports on the CORTOLIMA web page, where the release of this species in different parts of the department is exposed, as a result of its entry into the entity's Center for Wildlife Care and the positive evaluation of the animals for their release. Finally, haplotypes from Casanare are 45 mutational steps away from those of Tolima, other authors have hypothesized that the current distribution of the species' populations could be explained not only by the separation generated by the rise of the eastern mountain range but by other events that would extinguish the ancestral populations of trans-Andean Colombia, including those of Tolima, for a subsequent repopulation process from Central America (Jimenez-Alonso, 2016a; Venegas-Anaya et al., 2008).

Phylogeographic relationships of the *C. c. fuscus* populations in Tolima: By analyzing the sequences obtained from GenBank (Venegas-Anaya et al., 2008) together with those that were the product of

this study, we were able to identify 5 independent lineages in the species (Annex 4). The fuscus 2 lineage corresponds to individuals from northern Costa Rica and fuscus 1 to those from southern Costa Rica and Panama, Colombian Caribbean and Tolima, the latter being the last divergence within the sample. Each one of the divergences between the different lineages found presents support values close to 100% with the methodologies used, being the lowest of 0.84 in IB and 64% in MP, corresponding to the second lineage separation. As a complement to the above, the network of haplotypes (Appendix 5) was constructed with one representative per haplotype identified both in this study and by Venegas-Anaya et al., (2008) found that the lineage that groups individuals from Tolima, Barranquilla, Panama and Southern Costa Rica, is 17 mutational steps away from the lineage that groups the haplotypes from Northern Costa Rica, 21 from the lineage of the haplotypes from Mexico and Salvador, 41 from the group of animals from Casanare and Peru and finally 49 from the lineage located in the most ancestral position of the phylogeny reconstructed so that the relationships established in the trees are preserved in the network.

The analysis shows that geographical associations within the clade 3-4 (fuscus 1 lineage) could be explained by ancestral gene flow followed by extinction of intermediate populations or by restricted gene flow, but with some long-distance dispersal in intermediate areas not occupied by the species, however, the inference shows that there are too few clades within the haplogroup to discriminate between range expansion, colonization with restricted dispersal, or gene flow. Additionally, the inference key indicates that the geographic relationships within clades 3-5 (Crocodilus 2 lineage) are consistently explained by allopathic fragmentation. For the other clades, the null hypothesis that there is no relationship between genetic and geographic distances could not be rejected, but this could be due to the low number of clades and haplotypes per lineage. Our results agree with the phylogenetic relationships reported by Jimenez-Alonso, (2016) and Venegas-Anaya et al., (2008), supporting the existence of 5 clades within *Caiman crocodilus*. Additionally, it was possible to identify that the sequences from Tolima are grouped together with the largest haplogroup, which ranges from southern Costa Rica to the foothills of the western slope of the eastern cordillera. The network of haplotypes made with the sequences available in GenBank plus those from Tolima and Casanare, obtained a topology that suggests a greater relationship between the animals of northern Mesoamerica and those of southwestern South America (separated by more than 3000 km) and therefore a greater genetic distance between the sequences that are geographically closer (300 km approx.). In other words, the animals most distant from each other in genetic terms, based on what was found in this study, are those from Tolima and Casanare.

CONCLUSIONS

- High values of haplotypic and nucleotide diversity were found in the department of Tolima, as well as a high number of haplotypes whose representativeness is supported by an accumulation curve with a high percentage of coverage.

- No evidence was found of demographic processes involving events of population expansion or decrease within the Tolima *C. c. fuscus* populations.

- The evidence shows the existence of a high population differentiation within Tolima, with the presence of unique haplotypes for each of the populations and finding only one haplotype shared by two populations (Prado and Ibagué).

- Contrary to what was expected when taking into account the geographical distance, the population of Ibagué was in an intermediate position to those of San Luis and Prado in terms of

genealogical relationships. The extension of the sampling area, the use of new molecular markers and the application of geographical information systems, would allow an increased understanding of the population structure within the department.

- The populations of *C. c. fuscus* present in Tolima are more related to the populations of *C. c. fuscus* in southern Costa Rica and Panama than to the populations of *C. c. crocodilus* in Casanare. Likewise, the latter are related to the populations of the Peruvian Amazon.

- Evidence was found that supports the hypothesis of an extinction of the lineages that occupied the trans-Andean territory of Colombia prior to the marine incursions from the north of South America.

BIBLIOGRAPHY

Ardila-Robayo, M. C., Moreno-Arias, R. A., Zúñiga, H., Martinez-Barreto, W., & Suarez-Daza, R. M. (2009). *Formulación del Plan de Conservación de la Especie Babilla Caiman crocodilus fuscus en Jusrisdicción de la Corportación Autónoma Regional de Cundinamarca*.

Arenas-Calle, W. C., & García-Merchán, V. H. (2011). Estadística filogeográfica Culex annulirostris : Construcción de redes de haplotipos en estudios de genealogías intraespecíficas, usando tres metodologías diferentes. *Revista de La Asociación Colombiana de Ciencias BIológicas*, *23*(1).

Balaguera-Reina, S., & Velasco, A. (2019). *Caiman crocodilus*.

Caldera, S. M., Jaramillo, M. C., Cochero, S., Pérez-Doria, A., & Bejarano, E. E. (2013). Diferencias genéticas entre poblaciones de Aedes aegypti de municipios del norte de Colombia, con baja y alta incidencia de dengue. *Biomedica*, *33*.

Clement, M., Posada, D., & Crandall, K. (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, *9*.

Colwell, R. K. (2014). EstimateS 9.1. 0. Statistical estimation of species richness and shared species form samples.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and high- performance computing. *Nature Methods*, *9*(8), 772.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.

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

Fu, Y. X., & Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics*, *133*(3).

Hartl, D. L., & Clark, A. G. (1997). *Principles of Population Genetics* (3rd ed.). Canada: Sinauer Associates, INC., Publishers.

Hudson, R. R., Slatkin, M., & Maddison, W. P. (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics*, *132*(2), 583–589.

Hudson, Richard R. (2000). A new statistic for detecting genetic differentiation. *Genetics*, *155*(4).

Jimenez-Alonso, G. (2016a). *Análisis filogenético de Caiman crocodylus (Cope, 1868) en la costa atlantica y pacifica colombiana empleando marcadores de ADN mitocondrial*. Universidad Nacional de Colombia.

Jimenez-Alonso, G. (2016b). *Análisis filogenético de Caiman crocodylus (Cope, 1868) en la costa atlantica y pacifica colombiana empleando marcadores de ADN mitocondrial*. Universidad Nacional de Colombia.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649.

Leigh, J. W., & Bryant, D. (2015). popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110–1116.

Morales-Betancourt, M. A., Fajardo-Patiño, A., Lasso, C. A., & De la Ossa, J. (2013). |Los Crocodylia de Colombia: generalidades y estado de conservación. In M. Morales-Betancourt, C. Lasso, J. De La Ossa, & A. Fajardo-Patiño (Eds.), *Biología y conservación de los Crocodylia de Colombia* (pp. 31–37). Bogotá: Serie Editorial Recursos Hidrobiológicos y Pesqueros Continentales de Colombia. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt (IAvH).

Palumbi, S. (1991). *Simple fool's guide to PCR*. Dept. of Zoology and Kewalo Marine Laboratory, University of Hawaii.

Panchal, M. (2006). The automation of Nested Clade Phylogeographic. *Bioinformatics*, *23*(4), Aguirre, E., Godoy, J. a. (2009). Flujo génico: https://doi.org/10.1093/bioinformatics/btl614

Posada, D., Crandall, K. A., & Templeton, A. R. (2000). GeoDis: A program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology*, *9*(4), Aguirre, E., Godoy, J. a. (2009). Flujo génico: https://doi.org/10.1046/j.1365-294X.2000.00887.x

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., … Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*(3), 539–542.

Rozas, J. (2009). DNA Sequence Polymorphism Analysis Using DnaSP. In David Posada (Ed.), *Bioinformatics for DNA Sequence Analysis* (Vol. 537, p. Aguirre, E., Godoy, J. a. (2009). Flujo génico:). Humana Press. https://doi.org/https://doi.org/10.1007/978-1-59745-251-9_17

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S., & Sánchez-Gracia, A. (2019). DnaSP. Barcelona: Universitat de Barcelona.

Sambrook, J., Fritsch, E., & Maniatis, T. (1989). *Molecular cloning: a laboratory manual*. New York: Cold Spring Harbor Laboratory Press.

Tajima, F. (1984). Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics*, *123*.

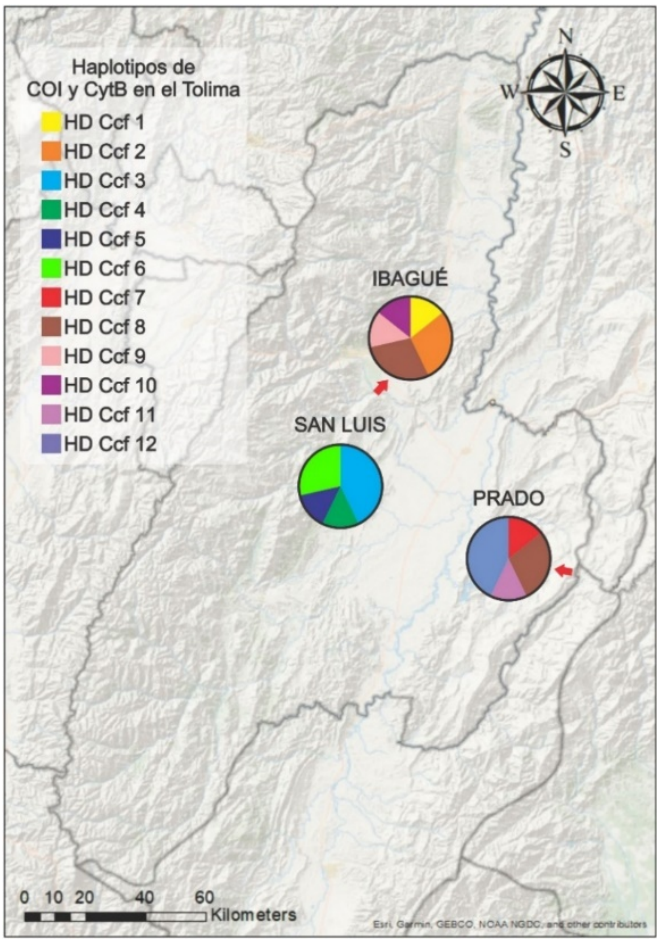
Templeton, A. R. (2004). Statistical phylogeography: Methods of evaluating and minimizing inference errors. *Molecular Ecology*, *13*(4), Aguirre, E., Godoy, J. a. (2009). Flujo génico: https://doi.org/10.1046/j.1365-294X.2003.02041.x

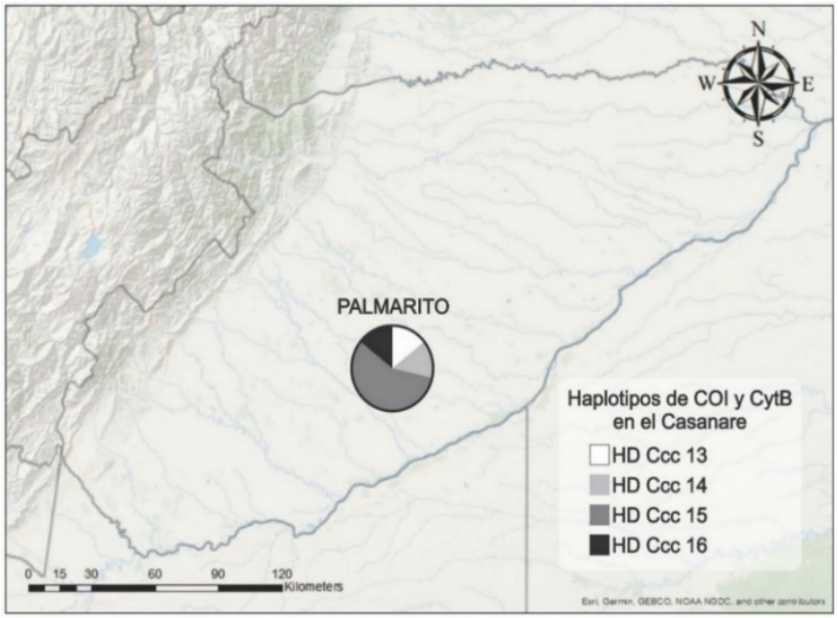
Venegas-Anaya, M., Crawford, A. J., Escobedo Galván, A. H., Sanjur, O. I., Densmore, L. D., & Bermingham, E. (2008). Mitochondrial DNA phylogeography of Caiman crocodilus in MesoAmerica and South America. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, *309*(10), 614–627. https://doi.org/10.1002/jez.502

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, *15*, Aguirre, E., Godoy, J. a. (2009). Flujo génico:

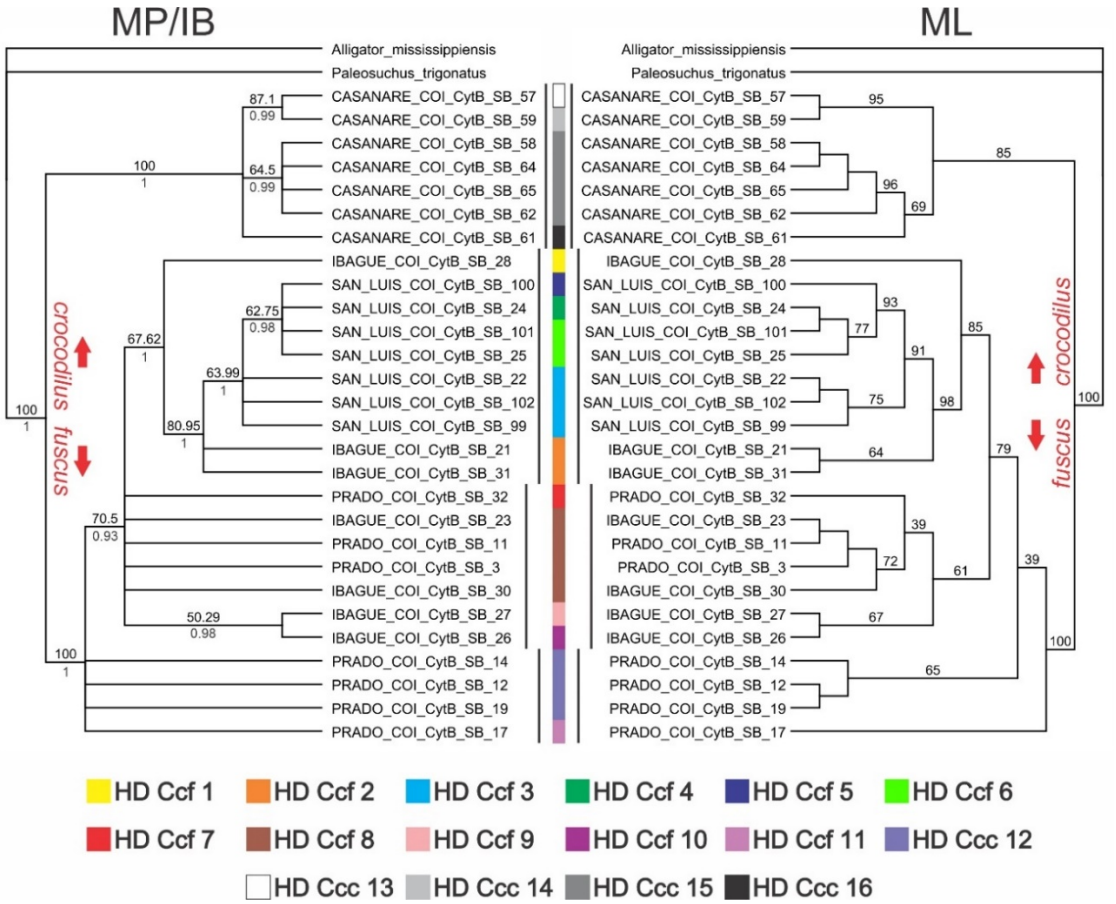
ANNEXES

Annex 1. Location of the 12 haplotypes of *Caiman crocodilus fuscus in Tolima* (above) and the 4 haplotypes of *C. crocodilus crocodilus* in Casanare (below).

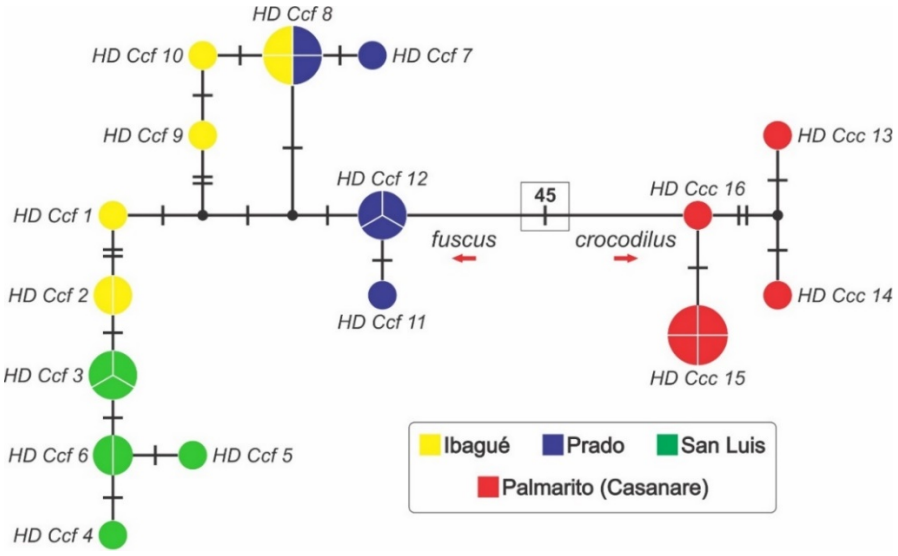




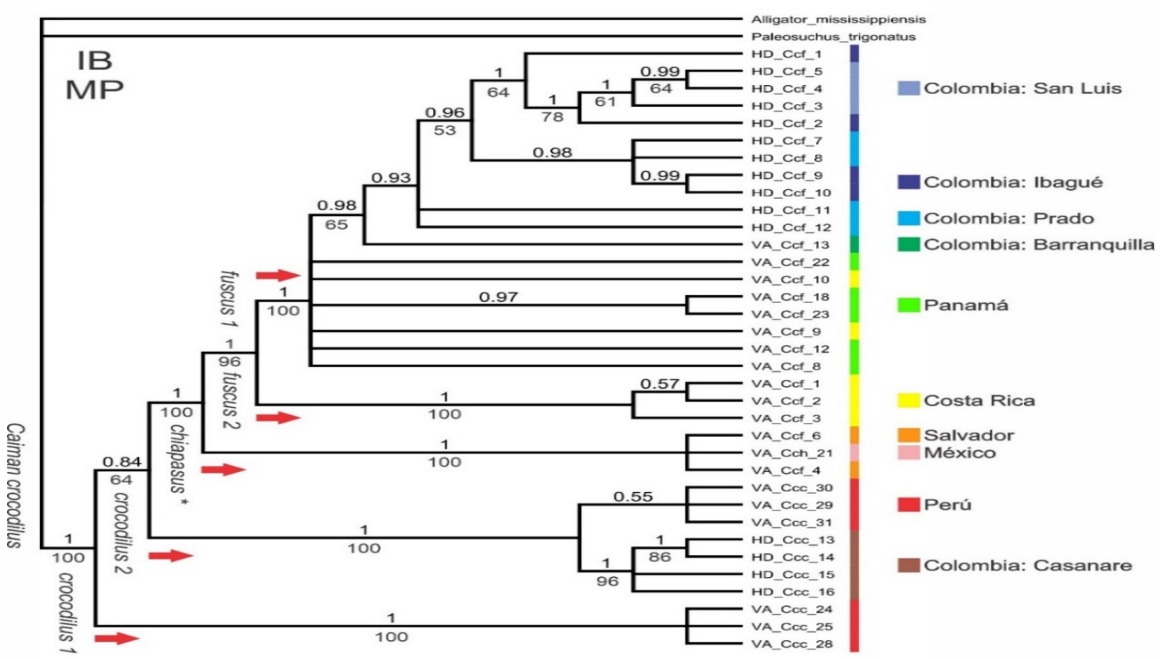
Annex 2. Phylogenetic relationships of populations sampled by Bayesian Inference, Maximum Parsimony and Maximum Likelihood for *COI* and *CytB*



Annex 3. Haplotype networks of Caiman crocodilus from 3 locations in Tolima and one in Casanare with the *COI* and *CytB* genes concatenated.

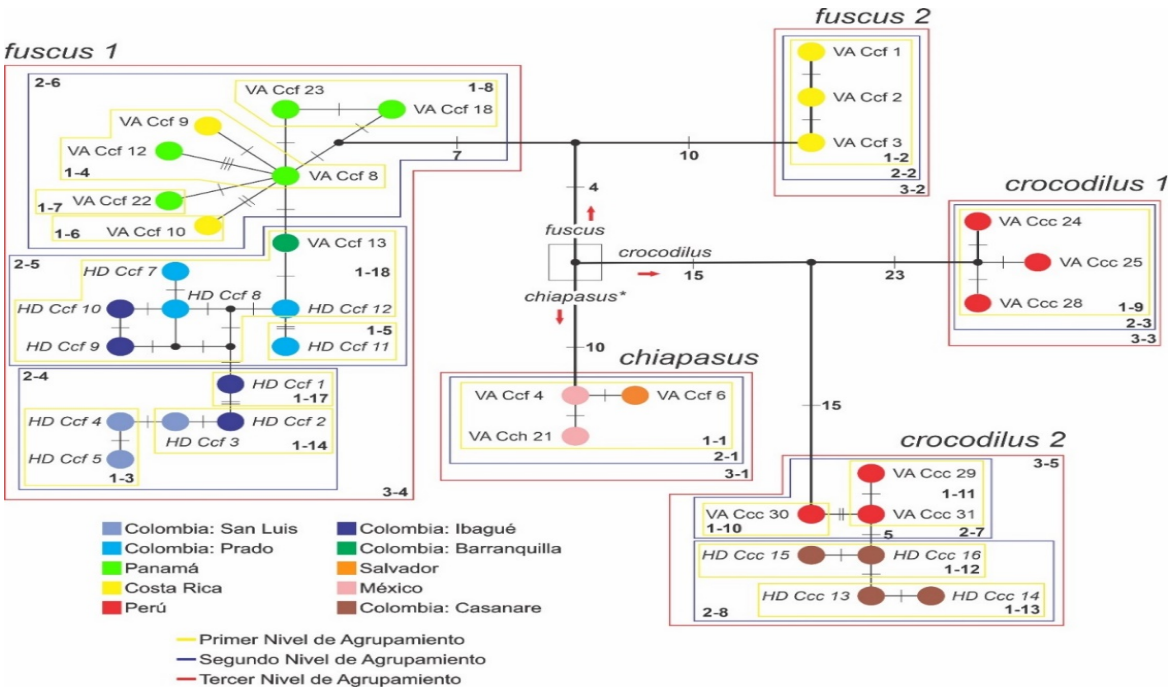


Annex 4. Phylogeny reconstructed by IB and MP from the concatenated sequences of the *COI* and *CytB* genes available for *Caiman crocodilus*



VA: Venegas-Anaya et al., (2008), HD: Authors

Annex 5. Haplotype networks and haplotype grouping of the *COI* and *CytB* genes available for *Caiman crocodilus* by TCS



Acknowledgements

I would like to thank my thesis advisors Carlos Prada, associate professor at the University of Tolima, and Sergio Balaguera, associate professor at the University of Ibague, who supported and guided me during the completion of this project. I would like to thank my thesis partner Carlos Felipe Hernandez with whom I developed this project from beginning to end, master's student Juan Felipe Moncada for his collaboration, Sigifredo Clavijo for his support and collaboration during the field work, Cristina Mora with whom I began this project, and Dr Carlos Piña.

I would like to thank Rafael Antelo from the Palmarito Foundation who donated tissue samples. To the laboratory of tropical parasitology of the University of Tolima and the laboratory of molecular biology of the University of Ibague, where the laboratory work was done. The IUCN-SSC Crocodile Specialist Group provided a SRAS grant to assist the project.