SMM Conservation Fund Report:

Project Title:

How many Amazon river dolphin species are there? “Capturing” genomic and morphological evidence to clarify the *Inia*’s taxonomy to help their conservation

Abstract

For almost four decades the number of species described for the genus *Inia* has been subject to controversy. For the past 30 years, researchers have published morphological and genetic evidence supporting at least three species with different distribution: the Bolivian river dolphin (*Inia boliviensis*), the Araguaian river dolphin (*Inia araguaiaensis*) and the Amazon river dolphin, *Inia geoffrensis,* the latest with two subspecies, the nominal subspecies for the Amazon basin and *I. g. humboldtiana* for the Orinoco Basin. However, the Society of Marine Mammalogy currently only accepts one species (*Inia geoffrensis*) with two subspecies, which hinders the conservation of species that face different types of threats. Some of the arguments against accepting different species include the occurrence of hybrid zones and the lack of unified criteria to describe these species. Therefore, we have proposed test the monophyly of the four putative lineages within the genus using museomics, as well as 3D morphometrics. Unfortunately, the museomics portion of this study was not successful, further sequencing of mitochondrial genomes from these samples will be done. The 3D morphological analysis of the skulls showed distinctiveness of four groups based on the size of the rostrum, the number of teeth, and the width of the braincase. Although morphometric data supports four species, the morphology analysis showed some overlapping points, probably due to hybridization between lineages, based on the location where the specimen was collected. Although, our molecular studies did not produce a complete rangewide phylogenetic and population level study of *Inia*, we find evidence for at least three lineages that, along with morphological differentiation, could be elevated to species level, *I. humboldtiana, I. geoffrensis,* and *I. boliviensis*. We unfortunately did not generate enough information from our museum specimens to adequately evaluate *I. araguaiaensis* genetically, but morphological analysis does show significant differentiation from the other lineages.

Key-words: subspecies; South America; freshwater dolphin; evolution.

Introduction

River dolphins are some of the least known cetaceans, found only in large river systems in Asia and South America. River dolphin ancestors colonized epicontinental seas during the global high seas of the Middle Miocene and evolved into independent lineages, remaining in the river basins when waters receded (Hamilton et al., 2001). In South America, two river dolphin lineages are found, the franciscana and the Amazon River dolphin. The franciscana (*Pontoporia blainvillei*) evolved in the Paraná River Basin and is now found in the mouth of the La Plata River and along the coastal southeastern Brazil, northern Argentina, and Uruguay. The Amazon River dolphin (*Inia* sp.) is found in three distinct watersheds (Amazon, Tocantins-Araguaia, and Orinoco basins), and the number of distinct species contained within this lineage has been questioned for decades (Banguera-Hinestroza et al., 2002; Gravena et al., 2014; Hamilton et al., 2001; Hollatz et al., 2011; Hrbek et al., 2014; Pilleri & Gihr, 1977, 1981; Ruiz-García et al., 2008, 2018). River dolphins not only constitute some of the most intriguing groups in terms of their speciation within river systems, but they are also an interesting "natural experiment" to study the adaptation of cetaceans to freshwater environments (Caballero et al., 2015).

Unfortunately, river dolphins are among the most endangered cetaceans in the world (Avila et al., 2018). Due to habitat destruction and directed captures to use them as bait, *Inia* populations have suffered steep declines over the last few decades (da Silva, Freitas, et al., 2018), and the genus *Inia* was recently classified by the IUCN as "Endangered" (da Silva, Trujillo, et al., 2018). Governments in South America developed a Conservation Management Plan that was presented in 2020 at the Small Cetacean Subcommittee of the Scientific Committee of the International Whaling Commission (to which S. Caballero is a delegate) in order to try to develop effective measures to protect Amazon river dolphins against the many threats that are currently affecting their populations, which include not only direct takes, but also negative fisheries interactions, dam construction, noise pollution and habitat destruction in general. However, one of the questions that is hindering conservation plans for the Amazon River dolphin is the lack of taxonomic clarification within this genus. Clarification of the taxonomy within this genus will directly affect conservation initiatives that can be designed and applied depending on the extent of the distribution of each species/subspecies determined, guiding plans at the national or international level that can then be placed into action.

For almost four decades, the number of *Inia* species described has been subject to controversy. Pilleri and Gihr (1977, 1981) described two species within this genus: *Inia boliviensis* for the Beni river system in Bolivia and *Inia geoffrensis* for the Amazon and Orinoco river systems. In recent years, South American researchers have published evidence both morphological and genetic, supporting at least three species, the Bolivian river dolphin (*Inia boliviensis*), which is separated from other river dolphins by a waterfall barrier (Grabert, 1984;

Ruiz-García et al., 2008), the Araguaian river dolphin (*Inia araguaiaensis*) (Hrbek et al., 2014) in the Tocantins-Araguaia river basin, and the Amazon river dolphin, *Inia geoffrensis,* with two subspecies, the nominal subspecies for the Amazon basin and *I. g. humboldtiana* for the Orinoco Basin (Hamilton et al., 2001). However, the Society of Marine Mammalogy currently only accepts one species, *Inia geoffrensis,* with two subspecies. Some of the arguments against accepting different species include the occurrence of hybrid zones between *Inia boliviensis* and *Inia geoffrensis* in the Madeira River (Gravena et al., 2014, 2015) and the lack of unified criteria to describe these species.

Molecular studies have revealed differences among these lineages using a few individuals and short fragments of mitochondrial DNA data, microsatellite loci, and nuclear introns (Hamilton et al., 2001; Hrbek et al., 2014; Ruiz-García et al., 2008; Siciliano et al., 2016), but few analyses have included all lineages, and none have utilized recent next-generation sequencing methods to generate massive amounts of nuclear DNA data to address this question. Therefore, we would like to test the monophyly of the proposed lineages within the genus *Inia* to clarify the number of species within this genus and provide information that helps prioritize conservation measures for its declining populations. We do this by attempting to sequence *Inia* from across its range and pairing this with geomorphic morphometrics of skulls to compare morphological and genetic evidence for putative species.

Methodology

Molecular Analyses using Fresh Samples

Sampling, DNA extraction, DNA quality, and quantification evaluation

Thirty-six skin tissue or blood samples were obtained from dead stranded animals or from animals captured for satellite tagging, conducted by Fundación Omacha in Colombia, Peru and Bolivia. These samples were collected in 2020 and 2021. Table 1 and Figure 1 describe the current subspecies designation, the geographic location, and the number of samples used.

For DNA extraction, 100 µL of blood or less than 100 mg of skin tissue was used, using a commercial kit for DNA extraction from New England Scientific. DNA quality was evaluated by running them on a 0.8% agarose gel and by quantification on a spectrophotometer Nanodrop 2000 (Thermo Scientific). The DNA was diluted to a final concentration of 30 to 50 ng/µL. High DNA quality samples were vacuum dried and shipped to SNPSaurus at the University of Oregon (USA) for library preparation and sequencing.

DNA library construction and SNP (Single nucleotide polymorphism) loci generation

DNA libraries were constructed (SNPsaurus, LLC) following the protocols by Russello et al. (2015). DNA was fragmented with Nextera DNA Flexreagent (Illumina, Inc), and adaptors were ligated to the generated fragments according to the library kit protocol. These fragments were amplified by PCR for 27 cycles at 74°C with one of the primers complementing the adaptor and extending the amplification by 10 nucleotides with the selective sequence GTGTAGAGCC. In this way, only those libraries were selected for the following steps. Libraries were sequenced in Illumina HiSeq 4000 in one lane, limited to reads of 150 bps (University of Oregon).

Genotype analyses followed specific scripts (SNPsaurus, LLC) limiting the reads by the tool "cortando bbduk" (BBMap tools, http://sourceforge.net/projects/bbmap/): bbmap/bbduk.sh in=reads/run_2780/2780_CAAGTGTC-GT AAGGAG_S25_L003_R1_001_subset.fastq.gz out=reads/run_2780/2780_CAAGTGTC-GTAAGGAG_S25_L003_R1_001_t.fastq.gz ktrim=r k=17 hdist=1 mink=8 ref=bbmap/resources/nextera.fa.gz minlen=100 ow=t qtrim=r trimq=10. Afterwards, a *de novo* reference was created collecting 10 million reads in total, in an even way from all samples, excluding reads of less than 7 or higher than 700, following the standards established by SNPsaurus Company. SNP loci were identified and aligned among them to determine identical allelic haplotype and generate a unique representation. These were then mapped for all reads against the created reference with a threshold of 0.95 using bbmap (BBMaptools). Genotype identification was done using the script calvariants (BBMaptools). The variant call format (Variant Call Format - VCF) was filtered to remove alleles with a frequency of less than 3%. The VCF file was filtered again and converted to other formats using SNPfiltR (DeRaad, 2022) and the software PGDSpider (Lischer & Excoffier, 2012). Data on these other formats were used in phylogenetic and population structure analyses.

Phylogenetic Analysis

Phylogenetic analyses of the resulting SNP data generated above were run using Maximum Likelihood inference. For the ML tree reconstruction, we implemented RAxML v8.2.X (Stamatakis, 2014) using the GTR+G evolutionary model. For the full ML tree searches, we used 1,000 replicates of the rapid bootstrap algorithm of RAxML to account for uncertainty in the estimation of the tree topology (Stamatakis, 2014).

Population Analysis

For the definition of population units using divergence of allelic frequencies (Dizon et al., 1992), identifying possible reproductive isolation (Waples, 1991), a variety of analyses were conducted using the SNPs generated above. As a first approximation, the software STRUCTURE v.2.3.4 (Pritchard et al., 2000) was used to infer the number of population units and the gene flow patterns of gene flow, examining the existence of different biological groups in the samples. Twenty independent runs were performed for K = 1–6 using 1,000,000 MCMC, with a burn-in of 100,000 iterations and considering a model of mixture with correlated frequencies. The posterior probability values were examined to evaluate convergence among independent runs. STRUCTURE Harvester (Earl & VonHoldt, 2011) was used to extract Q values from each run of K and to summarize results the software CLUMPP 1.1.2 (Jakobsson & Rosenberg, 2007) was used. Finally, the software DISTRUCT v.1.1 (Rosenberg, 2003) was used to plot and visualize the Q matrix containing ancestry probabilities for each sample in each defined population unit. The 'Evanno' method (Evanno et al., 2005) and the Puechmaille method (Puechmaille, 2016) were used to infer the most likely number of biological groups. using the software StructureSelector (Y. L. Li & Liu, 2018). The population structure patterns were also explored using a principal component analysis (PCA) and a discriminant analysis of principal components (DAPC) using the R package ADEGENET (Jombart, 2008).

Molecular Analyses using Museum Samples (Museomics):

Museum Sampling and Extraction

To extend sampling outside of the regions where we were able to obtain fresh tissue samples (see above), we sampled bone fragments, teeth or adherent osteological material (osteocrusts) from a total of 84 skulls across the range of *Inia* and including all four major lineages. These individual samples were collected from museums in the United States and Europe, with original localities and associated metadata shown in Table 2. DNA extraction was performed by phenol/chloroform isolation and subsequent cleaning with Qiagen MinElute spin columns as described by Hawkins et al. (Hawkins et al., 2022) for degraded historical DNA. The extractions were performed in the historical DNA clean room at the Smithsonian Museum Support Center, Laboratories of Analytical Biology (LAB).

Library preparation, hybridization, and sequencing

Between the 100 samples described on table 2, for seventeen samples (marked with a red X) not enough DNA concentration was obtained for the following protocols. Sixteen samples marked with a blue X, were not requested from the institutions of origin. After

extraction, for 67 museum-derived samples, we sheared the DNA using Covaris ME220 ultrasonicator (Covaris, Woburn, MA) to 200 bp fragments and constructed genomic libraries using New England Biolabs NEBNext Ultra II DNA Library Prep kits, according to the protocol provided by the manufacturer.

We constructed a panel of tiled 100 bp RNA-biotinylated baits (myBaits Custom Baits Kit, Daicel Arbor Biosciences, Ann Arbor, MI) derived from the SNPs identified above and a reference genome of *Inia geoffrensis* (mIniGeo1.hap1). Each locus included the SNP as well as 61 bp flanking each side of the SNP for a total of 122 bp. Due to repeated data surrounding some of the SNP loci, we were only able to generate complete RNA-biotinylated baits for 358 SNP loci, or about 10% of the loci from the initial SNPsaurus run. These baits were hybridized to our genomic libraries using the protocol outlined in myBaits manual version 5.03 with a hybridization temperature of 60ºC. Samples were grouped into pools and sequenced separately (one lane each, paired-end 150bp, with a quality score minimum 20, Illumina Nova Seq) by Breakthrough Genomics (Irvine, CA, USA).

Mapping and Variant Calling

After demultiplexing, Illumina paired-end reads were cleaned with BBDuk version 38.85 (Brian Bushnell, https://www.seqanswers.com/forum/bioinformatics/bioinformatics-aa/37399-introducing-bbduk-adapter-quality-trimming-and-filtering?t=42776) in Geneious Prime v.2025.1.2 (www.geneious.com) using default parameters. Reads were then aligned to the reference of 358 SNP loci/flanking regions of 122-bp each using default parameters in bwa-mem (H. Li, 2013; H. Li & Durbin, 2009). Going forward, only samples that mapped to at least 50% of the reference were kept for further processing. Variants were called using the "HaplotypeCaller" function in GATK4 (McKenna et al., 2010; Poplin et al., 2017; Van der Auwera & O'Connor, 2020) and the resulting VCF files were combined using "CombineGVCFs" and genotyped using "GenotypeGVCFs" in GATK4. Variants were then filtered using "VariantFiltration", keeping variants with a depth >10. SNPs were then pruned in bcftools (Danecek et al., 2021) to keep only one SNP per locus in a narrow region of 20 bp in the middle of each locus.

Phylogenetic Analysis

The resulting finalized VCF file was transformed into a phylip file using vcf2phylip (Ortiz, 2019) for the purposes of phylogenetic analysis. Maximum likelihood analysis was performed in IQTREE 2.1.2 (Minh et al., 2020) using default parameters as well as the GTR+ASC model to correct for ascertainment bias as a result of SNP selection. Ultrafast

bootstrapping (Hoang et al., 2018) was implemented with 1000 iterations to generate bootstrap scores for each node. The resulting tree was used to identify outliers, and an inspection of SNP data led to further exclusion of samples with >40% missing data. The analysis was then performed again as above.

Morphometric analysis

To avoid variability due to age, only adult skulls were examined. Specimens were considered adults when the basioccipital-basisphenoid suture was totally fused and closed (da Silva, 1994). Differences in tooth count was evaluated for both maxilla (TnS - number of upper teeth in the left and right sides of the maxilla) for the four putative species. The number of alveoli was counted on both the left and right sides of the skull. Because differences between sides are small (usually differing in only one tooth), we estimated an average of the teeth count for both sides, in order to produce a single value for each individual.

All the specimens used are listed in Table 2, with the respective geographical location plotted in Figure 2. The ones not marked with a red X are the ones used on the final analysis. All the skulls were scanned with Artec Space Spider Scanner (Artec 3D, Luxembourg). Scans were cleaned, prepared, and merged in the software Artec Studio, and exported in ply format before being further cleaned and decimated in Geomagic Wrap software (3D Systems). Models were decimated down to 1,500,000 triangles, reducing computational demands without compromising on detail for further morphometric analyses.

One hundred and twenty-one anatomical landmarks were manually inserted onto the surface of the scanned skulls using the "single point" option in the Stratovan Checkpoint software (Stratovan, Davis, CA). Of the 121 landmarks, 56 were inserted on both sides of the skull (left-hand side - LHS marked in purple, and right-hand side - RHS marked in green), and nine were placed on the medial line (marked in red), covering the dorsal and ventral regions of the skulls. The descriptions of the landmarks, as well as their position on the skull, are found in Table 3 and Figure 3. The semilandmarks are described in Table 4.

The selection of the landmarks was based on the descriptions of variation among taxa described by Pilleri and Gihr (1977) and da Silva (1994), as well as the variation observed by Coombs et al. (Coombs et al., 2020) among different cetacean species.

For the morphometric analyses, which were performed in the geomorph package v.3.1.0 (Adams et al., 2020), all landmarks must be present. To estimate the positions of missing landmarks, they were placed in the closest position to where they would originally be and marked as 'missing landmark' in Checkpoint software, which automatically generates a -9999 coordinate. Using the geomorph package, the estimate.missing function was used with the TPS (thin plate spline) method, where a subset of the most morphologically related and complete specimens is selected and aligned to the incomplete specimen using the common

landmarks. The software then applies a TPS spline transformation on the common landmarks, extending it to the unobserved one, thus estimating the missing data. Prior to statistical analyses, all specimens were superimposed to remove non-shape elements (i.e., size (scale), translation, and rotation (positioning)) from the data, using Generalized Procrustes Analysis (GPA) implemented by the 'gpagen' function of the geomorph package in R. The centroid size (the square-root of the sum of squared distances between the configuration center of gravity and each landmark) was extracted from the skull shape during the GPA and was used in subsequent analyses. A principal component analysis (PCA) was performed on the Procrustes-transformed morphometric data to evaluate the primary aspects of skull shape variation. Morphologies representing the extreme shapes along the principal component (PC) axes were generated to visualize shape variation. PCA analysis was performed using the 'gm.prcomp' function of the geomorph package.

The 'procD.lm' method, a permutation-based linear model, was used to test shape differences among putative taxa. We investigated the difference among species, adding the log-centroid size as a covariate, to control allometric effects. We also evaluated the interaction among species and sizes to account for differences in allometric relations. The permutation procedure was performed with 999 iterations to test hypotheses about shape variation while considering the multivariate nature of the data. The method uses residual randomization for permutations and calculates the effect of sizes (Z-scores) and significance levels based on F-distributions, providing a robust approach to investigating group differences.

To visualize differences among groups, we conducted a Canonical Variate Analysis (CVA) using the lineages as factors. Because CVAs are sensitive to high variable-to-specimen ratios due to the necessity to invert non-definite positive covariance matrices (Mitteroecker & Bookstein, 2011; Strauss, 2010), we used a pseudo-inversion method implemented in the package MORPHO (Schlager, 2017). The extremes of variation along both PCs and CVs were obtained to illustrate shape changes along those axes.

Results

Analysis of fresh samples

After filtering, a matrix of 3,255 SNPs was obtained for the 36 fresh samples analyzed, generating the Maximum Likelihood tree presented in Figure 4.

In this phylogenetic reconstruction, most of the individuals from the Orinoco region in Colombia group together in one clade (blue), including those from Arauca province, Guaviare Province, Meta River, and Serranía de la Macarena, as well as one sample from Puerto Carreño, Vichada province. In the other lineage, there is a branch separating the sample from Bolivia, and an additional clade showing reciprocal monophyly between three samples from

Puerto Carreño (Vichada province) in the Orinoco Basin and samples from the Colombian and Peruvian Amazon.

The analysis in STRUCTURE suggested the existence of at least four differentiated biological groups or population units (Figure 5).

The Puechmaille analysis (Figure 6) found support for 3 and 4 clusters when using the software StructureSelector. This method is more precise when using uneven numbers of samples and to deal with potential hierarchical variance due to possible population substructuring, as is likely the case for *Inia* sp.

The discriminant analysis of principal components (DAPC) showed similar results, with four biological groups or population units identified (Figure 7).

Museomics

We performed an initial run on one lane of Illumina NovaSeq with 44 pooled museum-derived samples using hybridization. Only 33 of these samples passed our cut-off of mapping to >50% of our reference. After conducting variant calling and filtering of SNPs, only 276 SNPs remained for further phylogenetic and population-level analysis. This was incredibly disappointing as the initial RAD Seq study above was able to generate 3255 SNPs. Initial phylogenetic analyses revealed numerous sample outliers which on inspection of the matrix still suffered from high levels of missing data. Ten samples were removed as they had >40% missing data and three others were removed due to potential contamination. The phylogenetic analysis was repeated. Figure 8 shows the unrooted maximum likelihood phylogenetic tree with only the remaining 20 samples included. The tree is supported by few characters, only 21 SNPs were deemed to be parsimony-informative. The Bolivian, Orinoco and Tocantins samples fall out separately from those of the Amazon basin, with only the Bolivian samples representing a well supported clade. The Bolivian samples support a distinct lineage similar to analyses of the fresh samples; both the one Orinoco sample and one Tocantins sample was positioned near the Bolivian lineage.

After the failure of this procedure to capture SNPs adequately, we decided to cease processing further samples. In order to rescue this part of the project, we will now concentrate on generating mitochondrial genomes using other funds and a leftover mito-baits kit that was purchased as part of another project. Mitochondrial DNA is often easier to sequence from historical samples, and this will add some context to our other molecular results from above.

3D Morphometrics

Emin-Lima et al. (2022) had already tested sexual dimorphism using t-tests performed on the cranial measurements of all *Inia* lineages, and none produced significant results. For this reason, we pooled males, females, and skulls with unknown sex assigned, to improve sample sizes.

For the shape variation analysis, using only 121 landmarks, the first two leading axes of the CVA showed more difference between the a priori defined lineages (CV1 = 40%; CV2 = 34%), excluding the probable hybrids from the Madeira and Tocantins rivers.

Using the landmarks and semilandmarks, CV1 explained 49% of the total variation, while CV2 explained 31% of the total variation (figure 9 and 10).

Using the landmarks and the semilandmarks (curves) described in table 1 and 3, respectively, CVA showed the two first axes as the most variable between lineages, CV1 explained 41% of the total variation, while CV2 explained 32% of the total variation (figure 9). The Analysis of Variance (MANCOVA) using residual randomization, for the lineages excluding the size of each specimen, showed that differences between lineages are explained by 15% of variance. If we include the size variation in the MANCOVA analysis, 14% of the variance is explained by the different lineages, and 13% are explained by the size, which means that allometry of the specimens are the same.

Discussion

As it has already been described by different authors, *Inia* dolphins from different river basins have different skull morphological characteristics as well as genetic differences at the genomic level.

Using fresh samples in genomic analysis it was possible to detect biological groups differentiating samples collected in the Colombian and Peruvian Amazon (dark purple color) from all samples from the Orinoco Region (blue and orange). Among the samples from the Orinoco Region, there is one biological unit which groups samples from Arauca River, Serranía de la Macarena, and Guaviare River, as well as some samples collected in the Meta River (Vichada Province) (blue color). However, it is also interesting to observe an additional biological group with samples from the Orinoco River, collected close to Puerto Carreño, Vichada Province (orange color). These results support the existence of one clear biological group or unit in the Colombian and Peruvian Amazon. In the Orinoco River, there were two lineages identified. One grouped samples from various rivers in the Orinoco Basin, including samples from Arauca and Guaviare River, as well as samples from Meta River and from the Guayabero River in Serranía de la Macarena. The other cluster was formed by samples from

the Orinoco River, collected close to Puerto Carreño, Vichada province. This second group in the Orinoco Basin, which appear differentiated from the Amazon River samples, appears to occupy the upper Meta River, close to the Cassiquiare Canal, the only water passage connecting the Orinoco and the Negro River, which is part of the Amazonas River basin. This result suggests two possible hypotheses. The first one, is that in this area there is probably recent or current connectivity between the populations of the upper Negro River and the Orinoco River, with some level of migration that should be tested in the future. Alternatively, the second hypothesis could support population differentiation between the Orinoco and Negro River populations with possible secondary contact between populations in this region.

In addition, the Bolivian sample (green bar on figure 5, and number 4 in figure 7) is also distinct from all the others. The small number of SNPs generated from the museum specimens also shows a distinct Bolivian clade (green dots in figure 8) but does not show the Orinoco or Tocantins samples as separate from those of Amazon basin, although this data is flawed as discussed above.

The 3D morphological analysis of the skulls showed distinctiveness of four groups based on the size of the rostrum, the number of teeth, and the width of the braincase (see figure 11). These features have been indicated by other authors (da Silva, 1994; Pilleri & Gihr, 1977) and we will still have to test these results statistically before publication. Besides that, using only the CVA analysis it is already possible to define four morphologically distinct groups, with some overlapping areas in morphospace, probably due to hybridization between lineages, based on the location where the specimen was collected. Hybridization between *I. geoffrensis* and *I. boliviensis* in the Madeira River have already been described (Gravena et al., 2015), and the divergence time between these species have been estimated at approximately 2.8 million years (Hrbek et al., 2014). Both species probably remained separated due to a barrier formed before the Madeira River drainage was running, which allowed morphological and genetic differentiation. Probably the increase in the water level, during annual events of extreme flood, over the period since the formation of the Madeira River, allowed botos from the Bolivian rivers to overcome the rapids, having access to populations already differentiated below them (Gravena et al., 2014), generating a secondary contact between putative species.

Although, our molecular studies did not produce a complete rangewide phylogenetic and population level study of *Inia*, we find evidence for at least three lineages that, along with morphological differentiation, could be elevated to species level, *I. humboldtiana, I. geoffrensis,* and *I. boliviensis*. However, our molecular analysis, did indicate some degree of mixing in the Puerto Carreño area, requiring further study. We unfortunately did not generate enough information from our museum specimens to adequately evaluate *I. araguaiaensis*

genetically, but morphological analysis does show significant differentiation from the other lineages.

Conclusions and Next Steps

As the museomics portion of this study was not successful, further sequencing of mitochondrial genomes from these samples will be done with other existing funds in attempt to rescue the historical samples part of the project before publication. This will create a mitochondrial tree that will encompass the entire range of *Inia*, even though this will only encapsulate mitochondrial data. Although morphometric data supports four species, with some degree of potential hybridization, more genomic data from other portions of the range, especially the Brazilian Amazon and Tocantins-Araguaia Basin will be needed to place the other putative lineages in context and fully resolve this species complex. We intend to publish these results in two papers, one highlighting the morphometric data, and another examining the RAD-Seq data from the fresh samples along with the future mitochondrial data from the museum specimens.

References

Adams, D. C., Collyer, M. L., & Kaliontzopoulou, A. (2020). *Geomorph: Software for geometric morphometric analyses* (R package version 3.3.1). https://cran.r-project.org/package=geomorph

Avila, I. C., Kaschner, K., & Dormann, C. F. (2018). Current global risks to marine mammals: taking stock of the threats. *Biological Conservation*, *221*, 44–58.

Banguera-Hinestroza, E., Cárdenas, H., Ruiz-García, M., Marmontel, M., Gaitán, E., Vázquez, R., & García-Vallejo, F. (2002). Molecular identification of evolutionarily significant units in the Amazon river dolphin *Inia* sp. (Cetacea: Iniidae). *Journal of Heredity*, *93*(5), 312–322. https://doi.org/10.1093/jhered/93.5.312

Caballero, S., Duchene, S., Garavito, M. F., Slikas, B., & Baker, C. S. (2015). Initial evidence for adaptive selection on the NADH subunit two of freshwater dolphins by analyses of mitochondrial genomes. *PLoS ONE*, *10*.

Coombs, E. J., Clavel, J., Park, T., Churchill, M., & Goswami, A. (2020). Wonky whales: the evolution of cranial asymmetry in cetaceans. *BMC Biology*, *18*(1), 86. https://doi.org/10.1186/s12915-020-00805-4

da Silva, V. M. F. (1994). Aspects of the biology of the Amazonian dolphins genus *Inia* and *Sotalia fluviatilis*. In *Biology: Vol. Ph.D.*

da Silva, V. M. F., Freitas, C. E. C., Dias, R. L., & Martin, A. R. (2018). Both cetaceans in the Brazilian Amazon show sustained, profound population declines over two decades. *PLoS ONE*, *02*, 1–12. https://doi.org/10.1371/journal.pone.0191304

da Silva, V. M. F., Trujillo, F., Martin, A. R., Zerbini, A. N., Crespo, E., Aliaga-Rossel, E., & Reeves, R. R. (2018). *Inia geoffrensis*. *The IUCN Red List of Threatened Species 2018*.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2). https://doi.org/10.1093/gigascience/giab008

DeRaad, D. A. (2022). snpfiltr: An R package for interactive and reproducible SNP filtering. *Molecular Ecology Resources*, *22*(6), 2443–2453. https://doi.org/10.1111/1755-0998.13618

Dizon, A. E., Lockyer, C., Perrin, W. F., Demaster, D. P., & Sisson, J. (1992). Rethinking the stock concept: A phylogeographic approach. *Conservation Biology*, *6*(1), 24–36. https://doi.org/10.1046/j.1523-1739.1992.610024.x

Earl, D. A., & VonHoldt, B. M. (2011). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Emin-Lima, R., Machado, F. A., Siciliano, S., Gravena, W., Aliaga-Rossel, E., de Sousa e Silva, J., Hingst-Zaher, E., & de Oliveira, L. R. (2022). Morphological disparity in the skull of Amazon River dolphins of the genus *Inia* (Cetacea, Iniidae) is inconsistent with a single taxon. *Journal of Mammalogy*, *5*(X), 1–12. https://doi.org/10.1093/jmammal/gyac039

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*(8), 2611–2620.

Grabert, H. (1984). Migration and speciation of the South American Iniidae (Cetacea, Mammalia). *Z Säugetierkunde*, *49*, 334–341.

Gravena, W., da Silva, V. M. F., da Silva, M. N. F., Farias, I. P., & Hrbek, T. (2015). Living between rapids : genetic structure and hybridization in botos (Cetacea : Iniidae : *Inia* spp .) of the Madeira River , Brazil. *Biological Journal of the Linnean Society*, *114*, 764–777. https://doi.org/https://doi.org/10.1111/bij.12463

Gravena, W., Farias, I. P., da Silva, M. N. F., da Silva, V. M. F., & Hrbek, T. (2014). Looking to the past and the future: were the Madeira River rapids a geographical barrier to the boto (Cetacea: Iniidae)? *Conservation Genetics*, *15*(3), 619–629. https://doi.org/10.1007/s10592-014-0565-4

Hamilton, H., Caballero, S., Collins, A. G., & Brownell Jr, R. L. (2001). Evolution of river dolphins. *Proceedings of the Royal Society of London Series B: Biological Sciences*, *268*(1466), 549–556. https://doi.org/https://doi.org/10.1098/rspb.2000.1385

Hawkins, M. T. R., Flores, M. F. C., McGowen, M., & Hinckley, A. (2022). A comparative analysis of extraction protocol performance on degraded mammalian museum specimens. *Frontiers in Ecology and Evolution*, *10*. https://doi.org/10.3389/fevo.2022.984056

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Hollatz, C., Vilaça, S. T., Redondo, R. A. F., Marmontel, M., Baker, C. S., & Santos, F. R. (2011). The Amazon River system as an ecological barrier driving genetic differentiation of the pink dolphin (*Inia geoffrensis*). *Biological Journal of the Linnean Society*, *102*, 812–827.

Hrbek, T., da Silva, V. M. F., Dutra, N., Gravena, W., Martin, A. R., & Farias, I. P. (2014). A new species of river dolphin from Brazil or: how little do we know our biodiversity. *PLoS One*, *9*(1), 1–12. https://doi.org/10.1371/journal.pone.0083623

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics (Oxford, England)*, *23*(14), 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. http://arxiv.org/abs/1303.3997

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, Y. L., & Liu, J. X. (2018). StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology Resources*, *18*(1), 176–177. https://doi.org/10.1111/1755-0998.12719

Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*(2), 298–299. https://doi.org/10.1093/bioinformatics/btr642

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Mitteroecker, P., & Bookstein, F. L. (2011). Linear Discrimination, Ordination, and the Visualization of Selection Gradients in Modern Morphometrics. *Evolutionary Biology*, *38*(1), 100–114.

Ortiz, E. M. (2019). *vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis.* . 10.5281/zenodo.2540861.

Pilleri, G., & Gihr, M. (1977). Observations on the Bolivian, *Inia boliviensis*, (D'Orbigny, 1834) and the Amazonian bufeo, *Inia geoffrensis* (Blainville, 1817), with a description of a new subspecies (*Inia geoffrensis humboldtiana*). In G. Pilleri (Ed.), *Investigations on Cetacea* (Vol. 8, Issue 1, pp. 11–76).

Pilleri, G., & Gihr, M. (1981). Additional considerations on the taxonomy of the genus Inia. *Investigations on Cetacea*, *12*, 15–27.

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M. J., Neale, B., MacArthur, D. G., & Banks, E. (2017). *Scaling accurate genetic variant discovery to tens of thousands of samples*. bioRxiv. https://doi.org/10.1101/201178

Pritchard, J. K., Stephens, M. J., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1461096&tool=pmcentrez&rendertype=abstract

Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators

alleviate the problem. *Molecular Ecology Resources*, *16*(3), 608–627. https://doi.org/10.1111/1755-0998.12512

Rosenberg, N. A. (2003). DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, *4*(1), 137–138. https://doi.org/10.1046/j.1471-8286.2003.00566.x

Ruiz-García, M., Caballero, S., Martinez-Agüero, M., & Shostell, J. M. (2008). Molecular differentiation among *Inia geoffrensis* and *Inia boliviensis* (Iniidae, Cetacea) by means of nuclear intron sequences. In V. T. Koven (Ed.), *Population Genetics Research Progress* (pp. 177–203). Nova Publishers Inc.

Ruiz-García, M., Escobar-Armel, P., de Thoisy, B., Martínez-Agüero, M., Pinedo-Castro, M., & Shostell, J. M. (2018). Biodiversity in the Amazon: Origin Hypotheses, Intrinsic Capacity of Species Colonization, and Comparative Phylogeography of River Otters (*Lontra longicaudis* and *Pteronura brasiliensis*, Mustelidae, Carnivora) and Pink River Dolphin (*Inia* sp., Iniidae, Cetacea). *Journal of Mammalian Evolution*, *25*(2), 213–240. https://doi.org/10.1007/s10914-016-9375-4

Russello, M. A., Waterhouse, M. D., Etter, P. D., & Johnson, E. A. (2015). From promise to practice: Pairing non-invasive sampling with genomics in conservation. *PeerJ*, *2015*(7). https://doi.org/10.7717/peerj.1106

Schlager, S. (2017). Morpho and Rvcg – Shape Analysis in R: R-Packages for Geometric Morphometrics, Shape Analysis and Surface Manipulations. In G. Zheng, S. Li, & G. Székely (Eds.), *Statistical Shape and Deformation Analysis* (pp. 217–256). Academic Press. https://doi.org/10.1016/B978-0-12-810493-4.00011-0

Siciliano, S., Hugo, V., Renata, V., Costa, A. F., Sartor, J., Dorneles, T., & Oliveira, R. De. (2016). New genetic data extend the range of river dolphins *Inia* in the Amazon Delta. *Hydrobiologia*. https://doi.org/10.1007/s10750-016-2794-7

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Strauss, R. (2010). Discriminating groups of organisms. In A. M. T. Elewa (Ed.), *Morphometrics for nonmorphometricians* (pp. 73–91). Springer.

Van der Auwera, G. A., & O'Connor, B. D. (2020). *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* (1st ed.). O'Reilly Media, Inc.

Waples, R. S. (1991). Genetic methods for estimating the effective size of cetacean populations. *Report of the International Whaling Commission*, *13*, 279–300.

Table 1 - Fresh skin and blood samples collected in the Amazon, the geographic location (numbers between parentheses corresponding to figure 1), and the number of samples (N) used in this study.

Taxa	**Geographic Location**	**N**
Inia geoffrensis geoffrensis	(1) Ucayali River, Peruvian Amazon	1
Inia geoffrensis geoffrensis	(2) Puerto Nariño, Colombian Amazon	4
Inia geoffrensis humboldtiana	(3) Puerto Carreño, Vichada Province, Colombia	4
Inia geoffrensis humboldtiana	(4) Meta River, Vichada, Province, Colombia	5
Inia geoffrensis humboldtiana	(5) Arauca River, Arauca province, Colombia	15
Inia geoffrensis humboldtiana	(6) Serranía de la Macarena, Meta Province, Colombia	4
Inia geoffrensis humboldtiana	(7) Guaviare River, Guaviare Province, Colombia	2
Inia geoffrensis boliviensis	(8) Bolivian Amazon	1
	TOTAL	**36**

Table 2 – DNA (🧬) and morphology (🦷) samples obtained from various museums and collections. Each sample's identification (ID) corresponds to the institution's acronym and deposit number. Information regarding location, country, and sex was provided by the institutions. Samples marked with a red X were not used in one of the methodologies, the concentration of DNA was not enough, or represented juvenile or broken skulls. Samples with a blue X were not requested from the institutions of origin. Museum IDs are at the end of the table.

N°	ID	Location	Country	Sex	🧬	🦷
1	LACM19588	Ucayali River	Peru	M		X
2	LACM19589	Amazon River	Peru	F		
3	LACM19590	Amazon River	Peru	M		
4	LACM19591	Amazon River	Colômbia	F		
5	LACM19593	Amazon River	Colômbia	M		
6	LACM19595	Amazon River	Colômbia	U		
7	LACM19596	Amazon River	Colômbia	U		
8	LACM27060	Amazon River	Colômbia	M		
9	LACM27063	Amazon River	Colômbia	F		X
10	LACM27064	Amazon River	Brazil	U		X
11	LACM27067	Ucayali River	Peru	M		
12	LACM27073	Amazon River	Peru	M		X
13	LACM27074	Amazon River	Peru	M		
14	LACM27075	Amazon River	Peru	F		
15	LACM28257	Guanare Viejo River	Venezuela	U	X	
16	LACM52457	-	Peru	M		
17	LACM54440	Amazon River	Peru	M		
18	LACM54441	-	Peru	F		
19	LACM54453	Orinoco River	Venezuela	F		
20	LACM72146	Amazon River	Peru	M		
21	MCZ31708	Tapajós River	Brazil	F		X
22	MCZ31709	Tocantins River	Brazil	M		
23	MCZ31710	Tocantins River	Brazil	F		

24	MCZ60060	-		U	X	X
25	MCZ142	-		U		
26	AMNH147502	Ucayali River	Peru	U		
27	AMNH147503	Ucayali River	Peru	U		
28	AMNH209101	Guaporé-Iténez River	Bolívia/Brazil	M		
29	AMNH209102	Guaporé-Iténez River	Bolívia/Brazil	F		X
30	AMNH209103	Guaporé-Iténez River	Bolívia/Brazil	F		
31	AMNH209104	Guaporé-Iténez River	Bolívia/Brazil	U		X
32	AMNH209105	Baurés River	Bolívia	M		
33	AMNH209106	Baurés River	Bolívia	F		
34	AMNH93412	Andirá/Amazonas River	Brazil	M	X	X
35	AMNH93413	Andirá/Amazonas River	Brazil	M		X
36	AMNH93414	Andirá/Amazonas River	Brazil	F		
37	AMNH93415	Andirá/Amazonas River	Brazil	F		
38	AMNH93416	Amazon River	Brazil	F		X
39	AMNH95753	Tapajós River	Brazil	F	X	
40	AMNH98695	-	Peru	U		X
41	USNM571366	Branco River, Roraima	Brazil	M		X
42	USNM239663	Tapajós River	Brazil	U		X
43	USNM239667	Tapajós River	Brazil	M		
44	USNM395602	San Fernando de Apuré River	Venezuela	F		
45	USNM395614	-		F		
46	USNM395415	San Fernando de Apuré River	Venezuela	F		
47	USNM395416	San Fernando de Apuré River	Venezuela	U		
48	USNM396166	San Fernando de Apuré River	Venezuela	F		
49	USNM406801	Manapiare River	Venezuela	U	X	
50	USNM49582	Purus River	Brazil	M	X	
51	USNM594663	Nanay River	Peru	U		
52	CAS13135	Amazon River	Peru	M		
53	CAS15977	San Fernando de Apuré River	Venezuela	F	X	
54	CAS16007	Cuieiras, Negro River	Brazil	M		
55	CAS16450	Solimões River	Colômbia	M		X
56	CAS16631	Amazon River	Colômbia	M	X	
57	CAS16637	Amazon River	Peru	M		
58	RMNH12400	Tocantins River	Brazil	M		
59	RMNH16109	San Fernando de Apuré River	Venezuela	F		
60	RMNH16936	Apuré River	Venezuela	M	X	
61	RMNH17613	Negro River	Venezuela	U	X	
62	RMNH17614	Manapiare River	Venezuela	U	X	
63	RMNH17615	Manapiare River	Venezuela	U		
64	RMNH17771	Apuré River	Venezuela	U		
65	SMNS45655	Ibaré River	Bolívia	M		X
66	SMNS45656	Ibaré River	Bolívia	F		
67	SMNS45657	Ibaré River	Bolívia	F		
68	SMNS45658	Mamoré River	Bolívia	F		
69	SMNS45659	Mamoré River	Bolívia	U		X
70	SMNS45660	Mamoré River	Bolívia	U		

71	SMNS45661	Mamoré River	Bolívia	F	X	
72	SMNS45662	Mamoré River	Bolívia	M	X	
73	SMNS45664	Apuré River	Venezuela	U	X	
74	SMNS45665	Orinoco River	Venezuela	U	X	
75	SMNS45666	Negro River	Venezuela	U	X	
76	NRM558294	Solimões River	Brazil	F	X	
77	NRM558295	Tocantins River	Brazil	M		
78	NRM558296	Tocantins River	Brazil	U		
79	NRM558297	Amazon River	Brazil	F		
80	NRM558298	Rio Madeira	Brazil	M		
81	NRM558299	Rio Madeira	Brazil	F		
82	NRM558300	Amazon River	Brazil	U		
83	NRM558301	Amazon River	Brazil	U		
84	NRM558415	Solimões River	Brazil	M		X
85	IDSM_IA01	Araguaia-Tocantins Basin	Brazil	F	X	
86	IDSM_IA03	Araguaia River	Brazil	M	X	
87	IDSM_IA04	Araguaia River	Brazil	F	X	
88	IDSM_IA05	Araguaia River	Brazil	U	X	
89	IDSM_IA06	Araguaia River	Brazil	M	X	
90	IDSM_IA07	Araguaia River	Brazil	F	X	
91	IDSM_Ig02	Araguari River	Brazil	U	X	
92	IDSM_Ig_Cass	Cassiporé River	Brazil	F	X	
93	Ig01_Coari	Solimões River	Brazil	U	X	
94	Ig02_Coari	Solimões River	Brazil	U	X	
95	NHMUK 1937.5.26.1	Amazon River	Brazil	F	X	
96	NHMUK 1856.8.2.1	Solimões River	Brazil	U	X	
97	NHMUK C.1939.5.13.1	Negro River	Brazil	U	X	
98	NHMNUK GERM1169b	Solimões River	Brazil	U	X	
99	NHMD CN17x	Apuré River	Venezuela	U	X	
100	NHMD CN18x	Apuré River	Venezuela	U	X	

Sex F-female, M-male, U-unknown; Museums: USNM (*National Museum of Natural History*, Smithsonian); LAC (*Los Angeles County Museum of Natural History*); MCZ (*Museum of Comparative Zoology*, Harvard); CAS (*California Academy of Science*, San Francisco); AMNH (*American Museum of Natural History*, New York); RMNH (*Naturalis Biodiversity Center*, Netherlands); SMNS (*Stuttgart State Museum of Natural History*, Germany); NRM (*Swedish Museum of Natural History*); IDSM (Instituto de Desenvolvimento Sustentável Mamirauá); NHMUK (National History Museum of London).

Table 3 – Description of landmarks placed on each specimen. Includes a description and the number of the corresponding left-hand side (LHS - purple) and right-hand side (RHS - green) landmarks. Midline landmarks are shown in red.

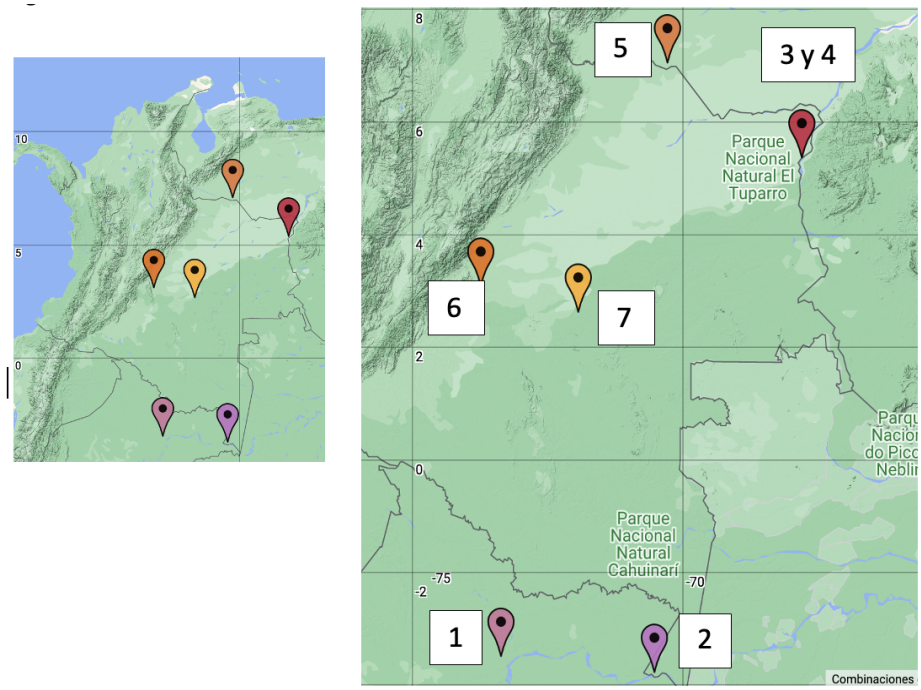
Landmark description	**Number on LHS**	**Number on RHS**
Nasal anterior	1	66
Left anterior lateral nasal	2	67
Posterior lateral corner of the nasal	3	68
Posterior point of the nasal	4	69
Posterior dorsal premaxilla	5	70

Dorsal medial maxilla (suture with nasal and premaxilla)	6	71
Nasal-frontal-maxilla suture (posterior medial maxilla)	7	72
Dorsal lateral posterior orbit maxilla suture	8	73
Front lacrimal medial	9	74
Frontal-jugal-maxilla ventral suture	10	75
Posterior tooth row lateral maxilla	11	76
Jugal anterior dorsal	12	77
Jugal anterior ventral	13	78
Jugal posterior ventral	14	79
Anterior medial frontal	15	80
Posterior medial frontal	16	81
Lateral posterior frontal (posterior lateral parietal suture)	17	82
Postorbital process/bar tip (anterior on crest)	18	83
Lateral frontal (on orbit)	19	84
Anterior dorsal corner of frontal (on orbit) – lacrimal suture	20	85
Anterior medial parietal – frontal suture	21	86
Posterior medial parietal	22	87
Posterior lateral parietal (squamosal/occipital suture)	23	88
Anterior lateral parietal (on vault)	24	89
Dorsal anterior lateral parietal (suture with frontal)	25	90
Dorsal anterior squamosal suture (with parietal, maybe alisphenoid/frontal)	26	91
Medial anterior zygomatic vault junction (squamosal)	27	92
Anterior dorsal jugal-squamosal suture	28	93
Ventral anterior lateral squamosal suture	29	94
Anterior medial most point of the mandibular articular process.	30	95
Posterior lateral-most point of the mandibular articular process	31	96
Lateral posterior squamosal (occipital suture)	32	97
Posterior medial dorsal squamosal (parietal/occipital suture)	33	98
MIDLINE: posterior margin of skull roof	*34*	*34*
Medial anterior supraoccipital (frontal-occipital suture, usually)	35	99
MIDLINE: dorsal/superior margin of foramen magnum	*36*	*36*
Lateral dorsal occipital	37	100
Dorsal medial occipital condyle	38	101
Dorsal lateral occipital condyle	39	102
Tip of paraoccipital process - lateral tip	40	103
Lateral ventral occipital squamosal suture	41	104
Ventral medial occipital condyle	42	105
Ventral lateral occipital condyle	43	106
MIDLINE: ventral margin of foramen magnum	*44*	*44*
MIDLINE: anterior basioccipital	*45*	*45*
Lateral anterior basioccipital	46	107
MIDLINE: anterior most point of basisphenoid, just posterior to the pterygoids and palate	*47*	*47*
Lateral anterior basisphenoid	48	108
Lateral posterior basisphenoid	49	109
MIDLINE: Medial posterior basisphenoid	*50*	*50*
MIDLINE: Posterior ventral medial point of palate	*51*	*51*
MIDLINE: Palatine anterior midline ventral suture	*52*	*52*

Pal-pterygoid suture	53	110
Pal-max lateral posterior suture	54	111
Pterygoid posterior	55	112
Ventral posterior pterygoid	56	113
MIDLINE: Maxilla ventral midline posterior suture	*57*	*57*
Premaxilla ventral midline posterior suture	58	114
Palato anterior lateral maxilla suture	59	115
Premaxilla ventral anterior	60	116
Premaxilla anterior lateral ventral suture	61	117
Anterior-most point of palatal surface immediately posterior to tooth row	62	118
Tip of rostrum, anterior dorsal side, anterior midline of tooth row (usually premaxilla)	63	119
Anterior dorsal premaxilla	64	120
Anterior lateral maxilla	65	121

Table 4 – Description of the semilandmarks placed on each specimen. The number of the corresponding left-hand side (LHS) and right-hand side (RHS) curves are specified in the first column. The first and last landmarks used to construct the curve are specified as lm1 and lm2 respectively, followed by the number of semilandmarks used on each curve, and the name of the bone marked.

Curves	**lm1**	**lm2**	**Number of semilandmarks**	**Bone**
1	20	17	9	left frontal-parietal
2	17	16	5	left frontal posterior
3	17	41	14	left parietal-zygomatic arch
4	33	24	13	left squamosal suture
5	17	15	9	left frontal anterior
6	11	64	28	left maxila last tooth to point
7	59	60	29	left ventral maxila
8	5	63	43	left pre-maxila
9	85	82	9	right frontal-parietal
10	82	81	5	right frontal posterior
11	82	104	14	right parietal-zygomatic arch
12	98	89	13	right squamosal suture
13	82	80	9	right frontal anterior
14	76	120	28	right maxila last tooth to point
15	115	116	29	right ventral maxila
16	70	119	43	right pre-maxila



*(8) *Inia geoffrensis boliviensis* (BOLIVIA)

Figure 1 - Sampling locations of fresh tissue used in this work. The number of the geographic location is the same as that used in Table 1.

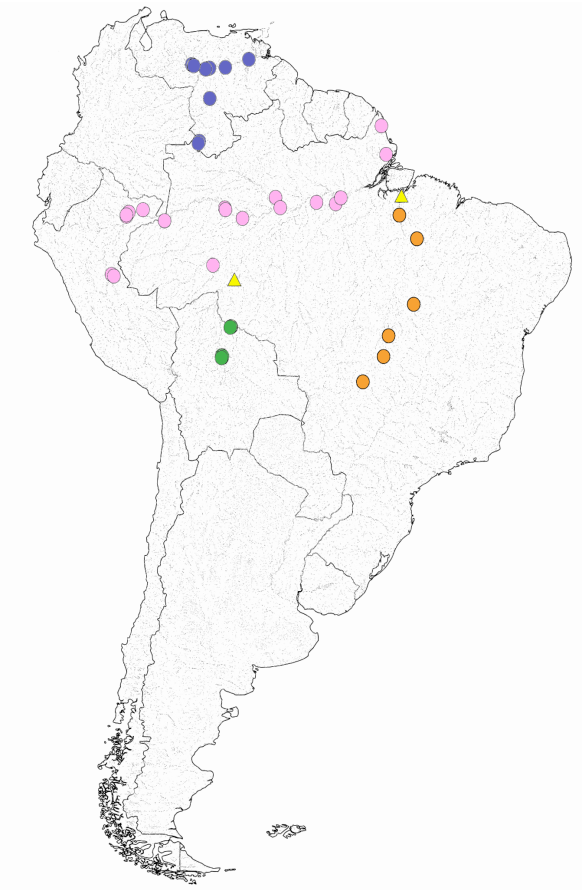


Figure 2 - Map showing the geographic distribution of samples obtained from different museums in the United States, Europe, and Brazil, according to Table 1. The geographic points were estimated based on descriptions and information provided by collectors at the time the specimens were deposited in the aforementioned collections. Different colours identify the four probable lineages according to their distribution: *Inia geoffrensis* in pink, *Inia boliviensis* in green, *Inia araguaiaensis* in orange, and *Inia geoffrensis humboldtiana* in blue. The yellow triangle indicates the area with possible hybridization between lineages.

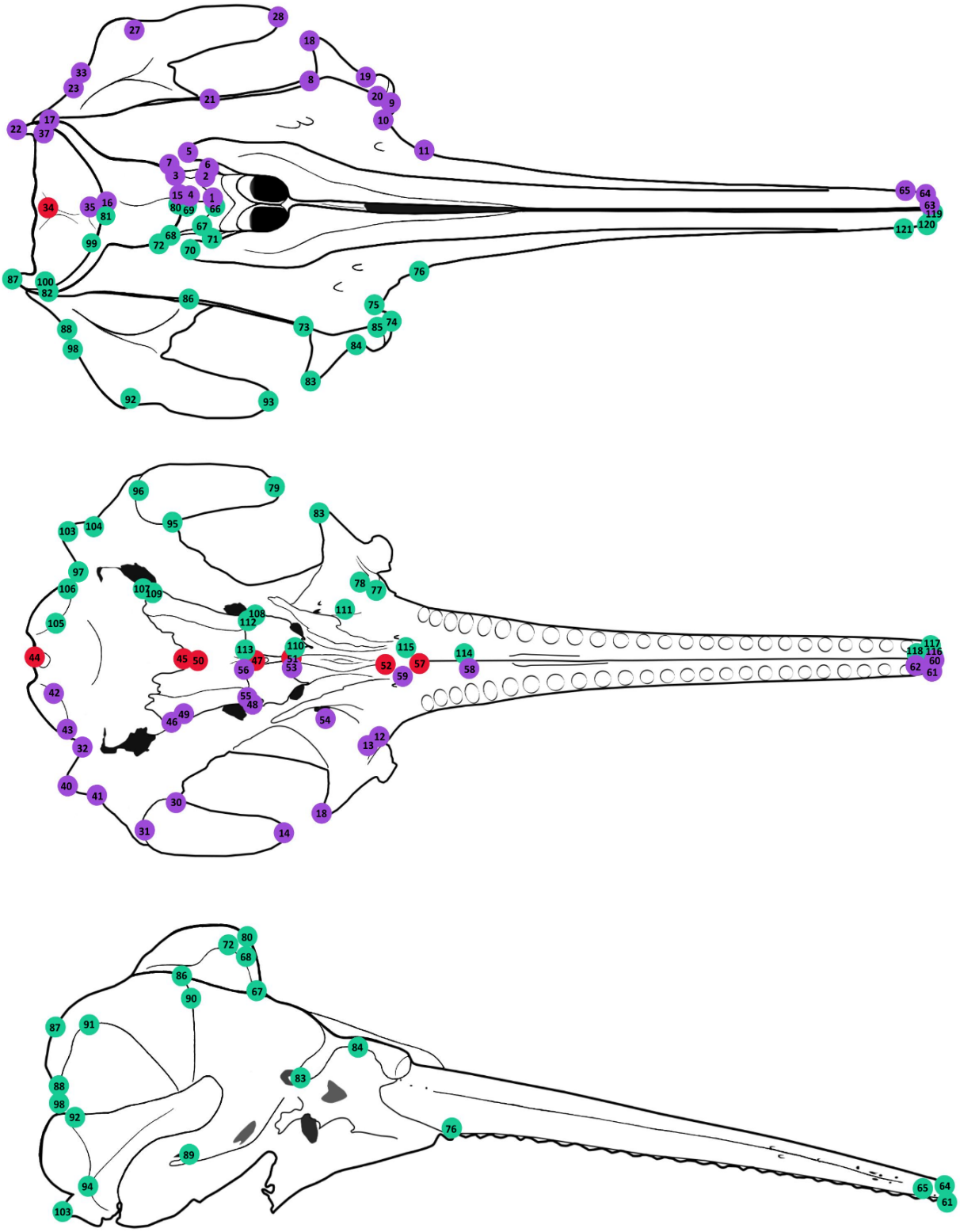


Figure 3 - Position of the 121 landmarks used on 3D morphological analysis on dorsal, ventral and lateral views of the skull of *Inia*. Landmarks on the left-hand side - LHS are marked in purple, and on the right-hand side - RHS are marked in green. The red ones are the medial line landmarks. The numbers on each side of the skull are described in Table 3. (Figure from Maira Laeta, modified by Waleska Gravena)

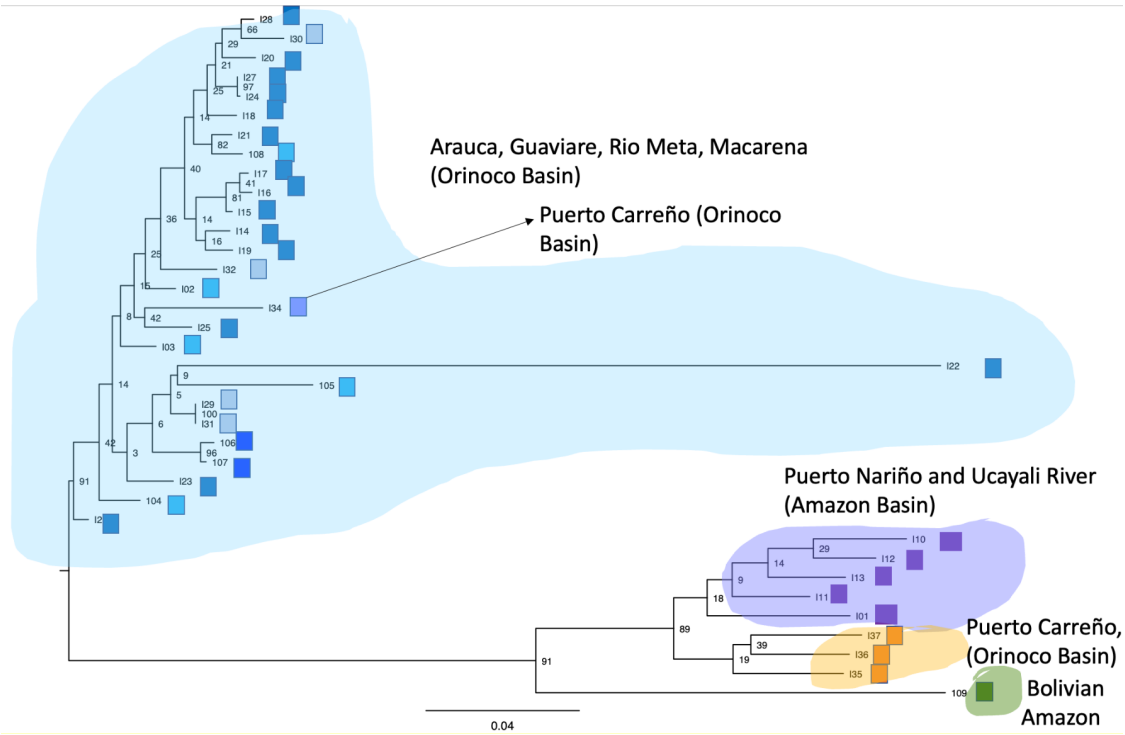


Figure 4 – Mid-point rooted, maximum likelihood tree obtained from analysis of 3255 SNPs from 36 fresh samples. The blue squares represent samples from Arauca, Guaviare, Rio Meta, Macarena, and only one sample from Puerto Carreño (with the arrow), in the Orinoco Basin. The purple squares represent samples from Puerto Nariño, and Ucayali River, in the Amazon Basin. Orange squares represent three samples from Puerto Carreño, Orinoco Basin. The green square represents the sample from the Bolivian Amazon.

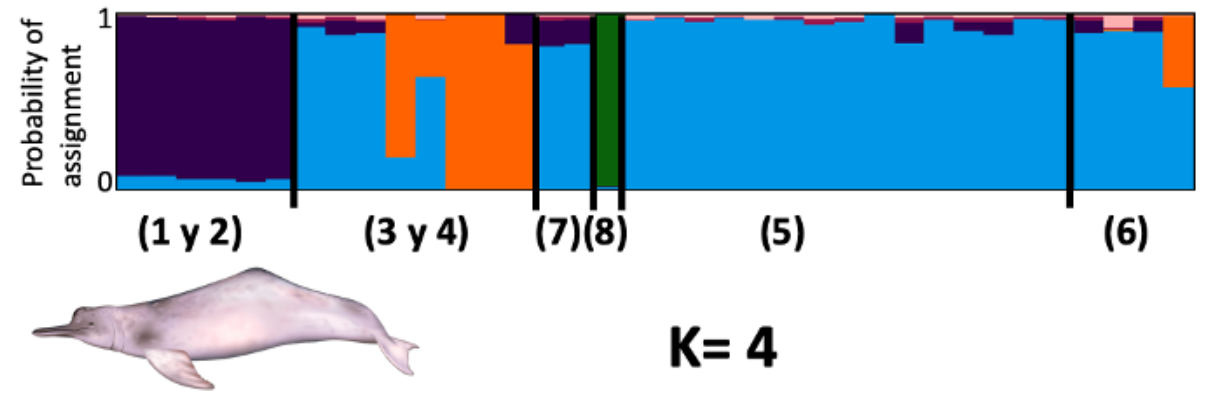


Figure 5 - STRUCTURE plot, showing four distinct biological groups or population units. Numbers under the plot, are the same listed in table 1 and represent: (1) Ucayali River, Peruvian Amazon; (2) Puerto Nariño, Colombian Amazon; (3) Puerto Carreño, Colombia, Orinoco Basin; (4) Meta River, Colombia, Orinoco Basin; (5) Arauca River, Colombia, Orinoco Basin; (6) Serranía de la Macarena, Colombia, Orinoco Basin; (7) Guaviare River, Colombia, Orinoco Basin; and (8) Bolivian Amazon.

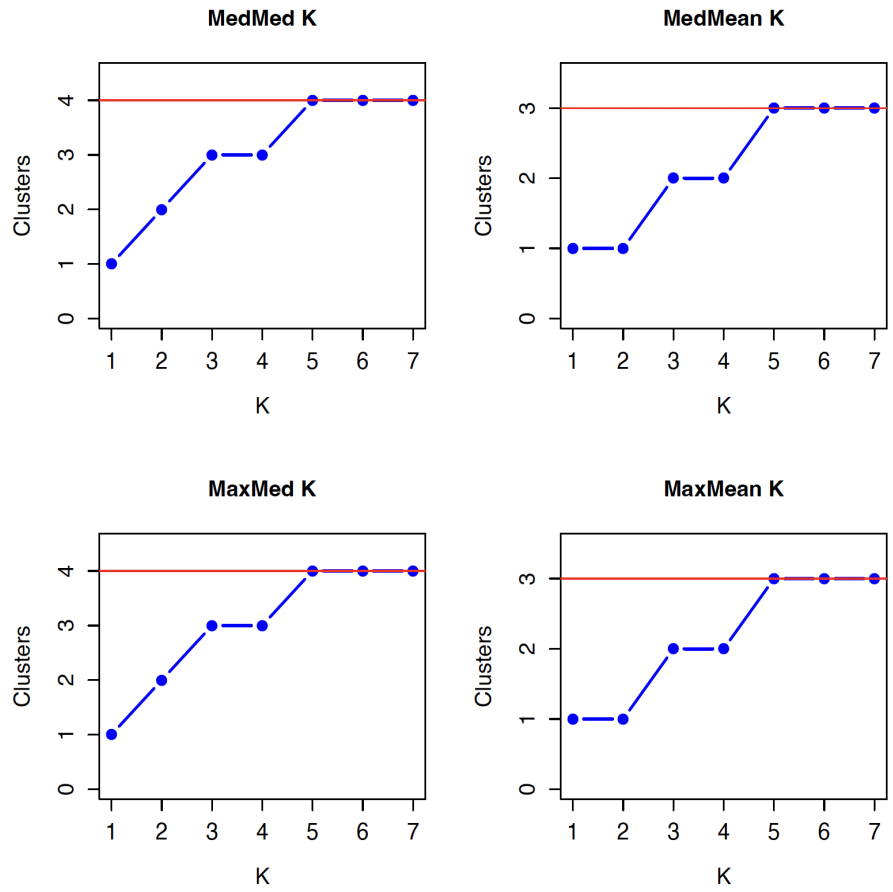


Figure 6 - Puechmaille analysis supporting K= 3 and K= 4 clusters when using the software StructureSelector.

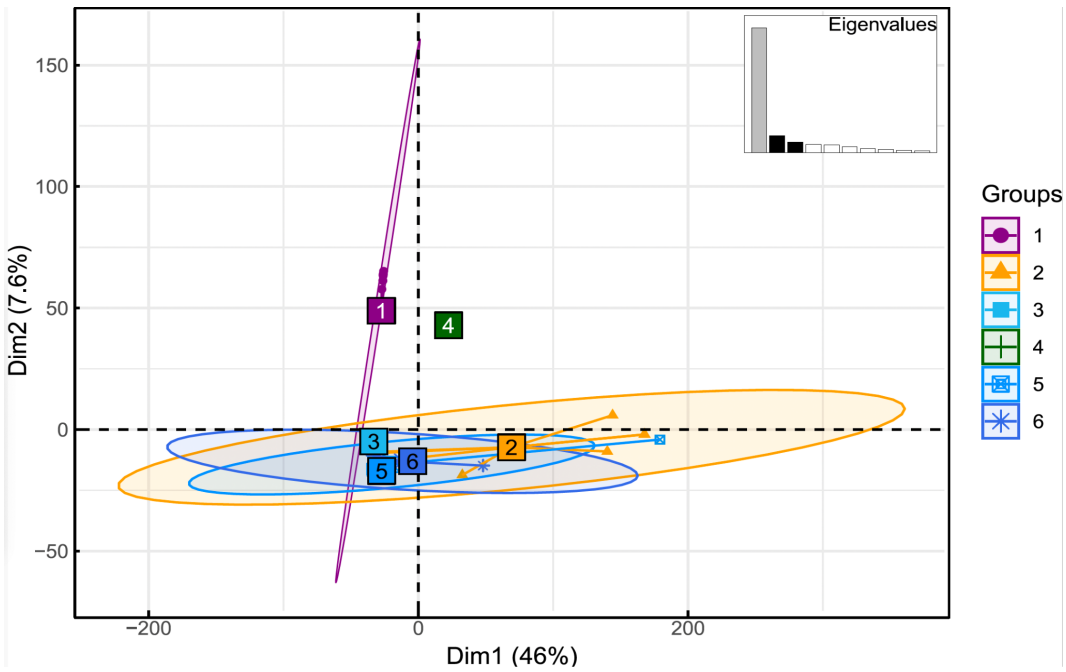


Figure 7 - Discriminant analysis of principal components (DAPC) with fresh samples described in Table 1. The numbers in the graph represent: (1) Ucayali River, Peruvian Amazon and one sample of Puerto Nariño, Colombian Amazon (Amazon Basin); (2) Puerto Carreño, Colombia Orinoco Basin;

(3) Meta River, Colombia, Orinoco Basin; (4) Bolivian Amazon; (5) Arauca River, Colombia, Orinoco Basin; (5) Serranía de la Macarena, Colombia, Orinoco Basin; and (6) Guaviare River, Colombia, Orinoco Basin.

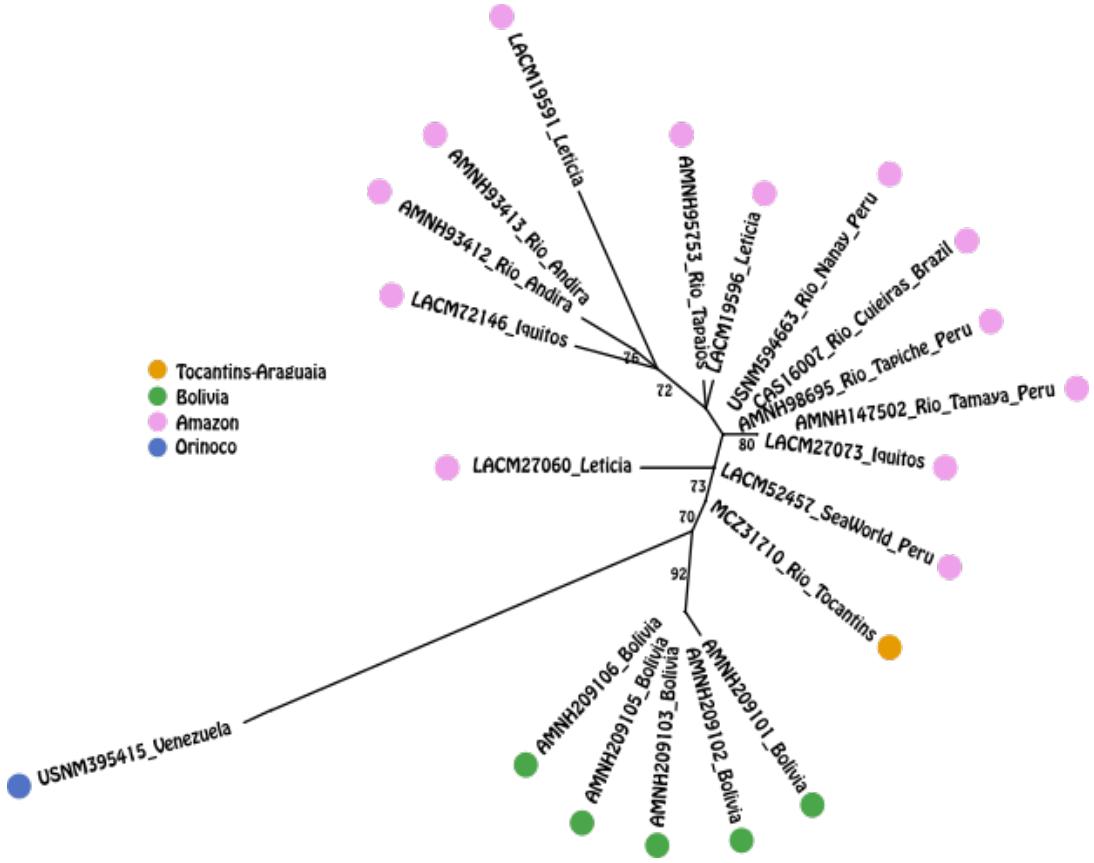


Figure 8 - Unrooted maximum-likelihood tree of 276 SNPs generated as part of the museomics study. River basins are shown in color to the right of each individual specimen, according to the caption on the upper left and the map (figure 2). All bootstrap scores >70 are shown.

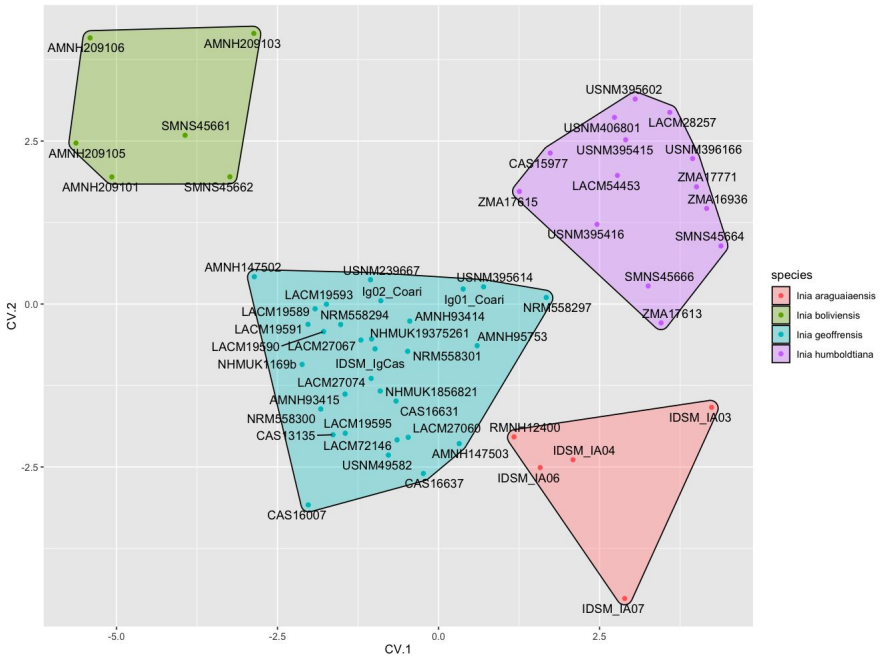


Figure 9 - CVA using only the landmarks of specimens of the four proposed lineages within the genus *Inia*. The lineages are represented by colors according to the caption on the right side.

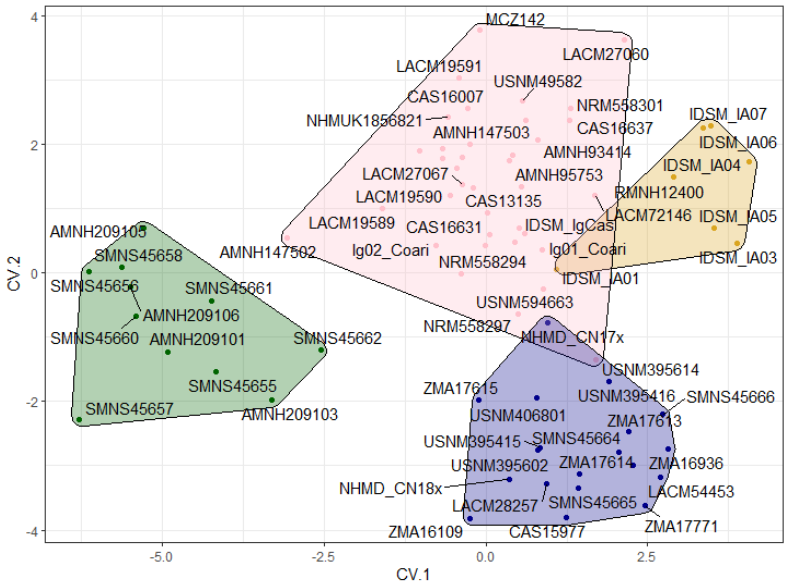


Figure 10 - CVA using only landmarks of all specimens of the four proposed lineages and the probable hybrids within the genus *Inia*. For this analysis the probable hybrids from the Madeira River and Tocantins River were included. The lineages are *Inia geoffrensis* (pink), *Inia boliviensis* (green), *Inia araguaiaensis* (orange) and *Inia geoffrensis humboldtiana* (blue).

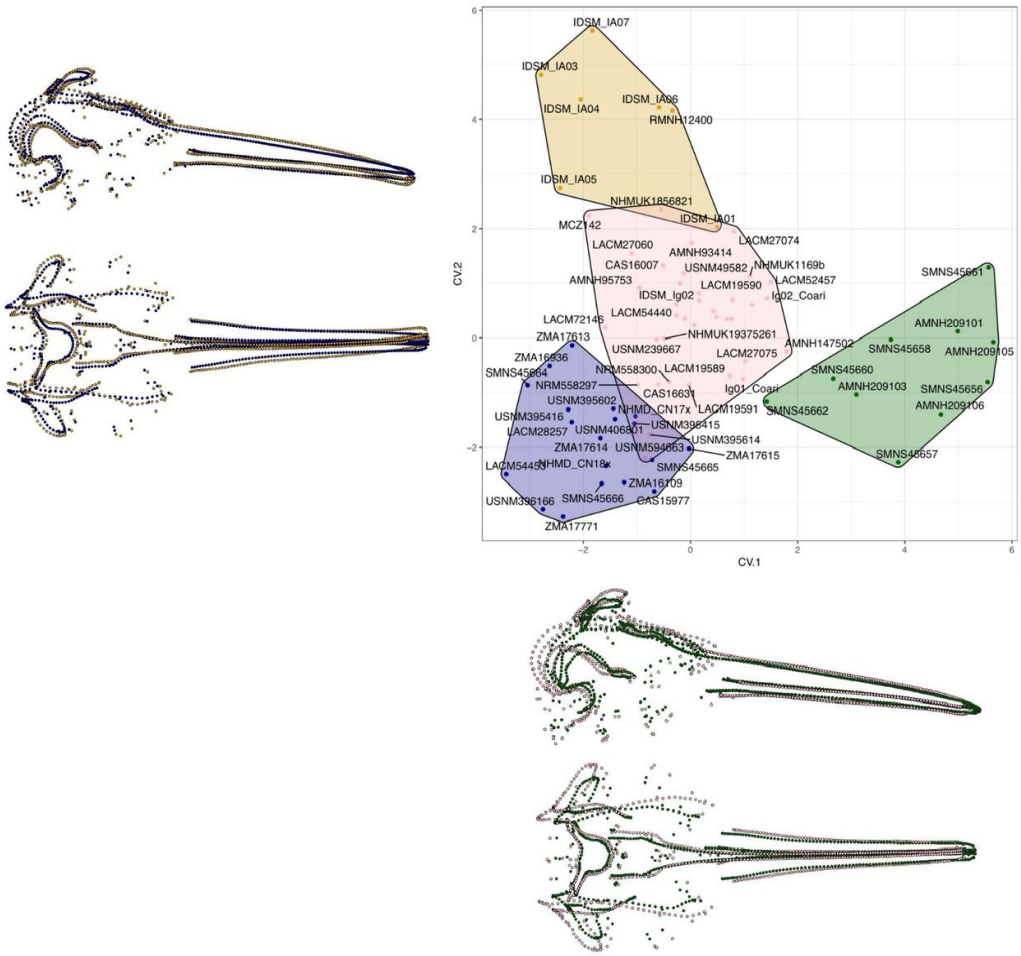


Figure 11 - CVA analysis using landmarks and semilandmarks of the four proposed lineages within the genus *Inia*. The lineages are *Inia geoffrensis* (pink), *Inia boliviensis* (green), *Inia araguaiaensis* (orange) and *Inia geoffrensis humboldtiana* (blue). Dotframes depict morphological differences between extreme groups according to lineage colors.