Ancestral Hybridization Yields Evolutionary Distinct Hybrids Lineages and Species Boundaries in Crocodiles, Posing Unique Conservation Conundrums

Gualberto Pacheco-Sierra[1,2†], Ella Vázquez-Domínguez[1,3*†], Jessica Pérez-Alquicira[4†], Marco Suárez-Atilano[1†] and Jerónimo Domínguez-Laso[5]

[1] Departamento de Ecología de la Biodiversidad, Instituto de Ecología, Universidad Nacional Autónoma de Mexico, Ciudad de Mexico, Mexico, [2] Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de Mexico, Ciudad de Mexico, Mexico, [3] American Museum of Natural History, New York, NY, United States, [4] CONACYT - Laboratorio Nacional de Identificación y Caracterización Vegetal, Departamento de Botánica y Zoología, Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Zapopan, Mexico, [5] COMAFFAS AC, Programa Crocodylia, Chiapa de Corzo, Mexico

Interspecific hybridization can lead to adaptation and speciation, especially in the context of recent radiations. The emblematic Crocodylus (true crocodiles) is the most broadly distributed, ecologically diverse, and species-rich crocodylian genus. Nonetheless, their within-species evolutionary processes are poorly resolved mainly due to their potential for hybridization. Notably, the evolutionary outcomes when hybridization is ancient and involves long-lived species, like crocodiles, remain largely unexplored. Here, we evaluate the genomic admixture between the American (*Crocodylus acutus*) and the Morelet's (*Crocodylus moreletii*) species, and demonstrate that this hybridization system challenges the definition of species boundaries and poses a triple conservation conundrum: what has been recognized as *C. acutus* is actually two distinct species, therefore its taxonomic reassessment is needed; we identified two evolutionary distinct hybrids lineages, which are genetically discernible from the parental species; the remaining *C. moreletii* populations evidence its likely extinction as a species and/or evolution via hybridization. Hence, the crocodiles' distinct species and hybrids lineages warrant recognition and need urgent conservation efforts.

Keywords: crocodiles, *Crocodylus acutus*, *Crocodylus moreletii*, extinction, genomic admixture, hybridization, hybrids lineages, lineage divergence

INTRODUCTION

Since Darwin, hybridization has intrigued evolutionary biologists, recognized as a regular occurrence in nature and a significant source of variation (Roberts, 1919); a means for species to acquire beneficial alleles and the potential to adapt, contributing to the divergence and speciation of lineages (Mallet, 2007; Abbott et al., 2013; Sardell and Uy, 2016; Gompert et al., 2017). Extensive evidence indicates that recent, rapidly radiating, closely related species are most likely to hybridize, as well as those with short generation times (Brennan et al., 2012; Lamichhaney et al., 2017), and that interspecific hybrids are very rare at the population level (Mallet, 2005, 2007).

Contrastingly, the evolutionary outcomes when hybridization is ancient and involves long-lived species with deep divergences, like crocodiles, remain largely unexplored. In such systems, the likely prediction would be a hybrid zone that is shaped by multiple episodes of primary divergence and secondary contact, reflecting both recent and ancient hybridization, with different outcomes including wide admixture zones, genetic swamping of one species by another, or reproductive isolation and hybrid speciation. Hybrid zones that have persisted long enough can facilitate long-lived species to reach an equilibrium between migration and selection, enabling their progress toward speciation (Gompert et al., 2006, 2017; Mallet, 2007; Abbott et al., 2013; Sardell and Uy, 2016). Therefore, unraveling the dynamics and consequences of such introgression systems can enrich our understanding of the role of hybridization in shaping biodiversity, and help elucidate the evolutionary relationships within species.

Crocodiles (Crocodylidae) are a group of high conservation concern, in which all 23 species worldwide are in some protected category (IUCN red list). The species-rich genus *Crocodylus* comprises 12 of the so-called true crocodiles, an ecologically diverse group with a wide circumtropical distribution. The taxonomy of *Crocodylus* has been well resolved, however, knowledge of the phylogenetic relationships within these emblematic species is still lacking. Indeed, elucidating the crocodile's within-species evolutionary processes has been challenged by their potential to hybridize. As a result, single crocodile species are recognized that have extremely wide distributions, hindering adequate conservation strategies. Here we investigate the hybridization system between two non-sister, long-lived species, the American (*Crocodylus acutus*) and the Morelet's (*Crocodylus moreletii*) recognized crocodile species. The Morelet's crocodile has a narrow historical distribution (northeastern Mexico's Gulf through the Yucatan peninsula to northern Guatemala and central Belize), compared to the much broader range of *C. acutus* (Mexican Pacific, Yucatan peninsula, Central America to northern South America), whereas their historical region of sympatry encompasses the north of the Yucatan peninsula southwards along the Caribbean, where they are known to hybridize (Pacheco-Sierra et al., 2016 and references therein), but little is known about their evolutionary history or their diversification and hybridization patterns.

In the present study we analyze the entire hybrid zone between *Crocodylus acutus* and *C. moreletii* in Mexico, based on different types of genetic and genomic data, to explain the phylogenetic, phylogeographic, and genomic features that characterize this unique hybridization system. We explore the following key questions: (i) Can we elucidate the evolutionary relationships between and within these crocodile species? (ii) When did the hybridization process between *Crocodylus acutus* and *C. moreletii* start? (iii) Can different lineages be distinguished, both in terms of parental and hybrid individuals? (iv) Can species boundaries be defined in this hybridization system? Our ultimate goal was to discern if hybridization could be leading to the differentiation of evolutionary lineages and, if so, what would be the conservation consequences and protection policy challenges. We reveal, based on genome-wide DNA polymorphism data

and a most comprehensive sampling and genomic analyses, a complex evolutionary history that involves different lineages and evolutionary trajectories, including the diversification of new species and distinct hybrid individuals.
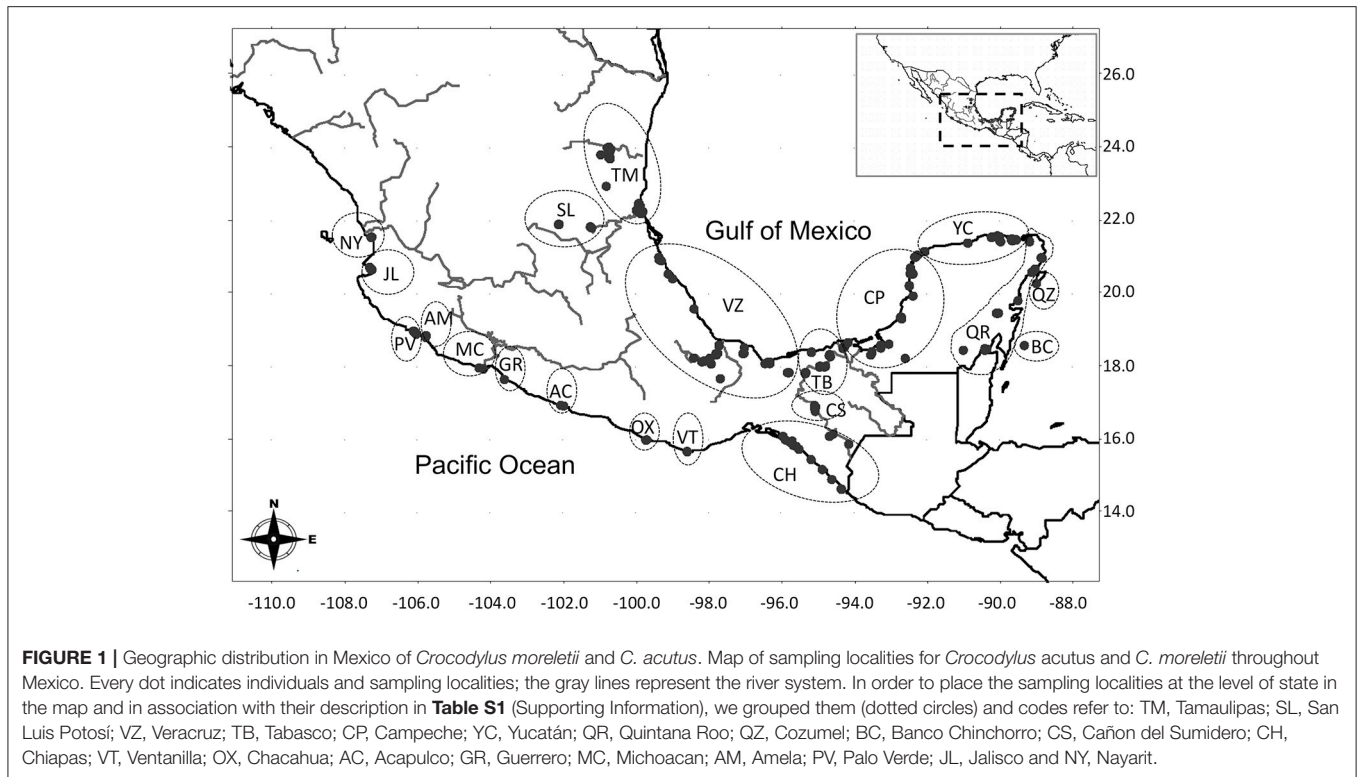
MATERIALS AND METHODS
Data Collection
We sampled 109 individuals for the present study which, together with 265 samples from Pacheco-Sierra et al. (2016), allowed us to analyze 374 crocodiles from 92 localities across the distribution of the recognized species *Crocodylus acutus* and *C. moreletii* in Mexico (**Figure 1**, **Table S1**). Thirty-five individuals from Pacheco-Sierra et al. (2016) were not included here because they were from localities that already had too many samples, and also because most analyses (see below) were done with subsets of the total individuals sampled. Crocodile individuals were captured and marked following a unique pattern by clipping tail scutes, following a numbered code, and liberated at the capture site (see Pacheco-Sierra et al., 2016 for details of the sampling scheme). Samples for genetic analyses included tail scutes and blood. Field work was performed in compliance with the Guidelines for use of live amphibians and reptiles in field and laboratory research (Beaupre et al., 2014). We extracted genomic DNA from each sample with the DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA, USA) following the manufacturer's protocol. We assessed the quantity and quality of the DNA in a microvolume spectrophotometer NanoDrop 2000 Thermo Scientific[TM].

Microsatellite Loci Amplification and mtDNA Sequencing
For the present study, we amplified 109 crocodile individuals from the Mexican Pacific coast and performed microsatellites genotyping using the same 12 polymorphic microsatellite loci (**Table S2**) and amplification protocols as in Pacheco-Sierra et al. (2016). We analyzed the microsatellite data (genotypes per individual) following the same sequence of analyses as in Pacheco-Sierra et al. (2016) for the new complete database (374 samples) encompassing the species' entire distribution.

We amplified the mitochondrial cytochrome *b* gene (cyt *b*) of a subset of 271 individuals, with the primers crCYTBF and crCYTBR (Weaver et al., 2008), and the gene region encompassing tRNA[Pro]-tRNA[Phe]-d-loop with primers L15459 and CR2HA (Glenn et al., 2002; Ray and Densmore, 2002). Amplification of both fragments was carried out in 25 μL reaction volumes containing: 25-50 ng template DNA, 1 unit of Taq DNA polymerase (Vivantis, Selangor, Malaysia), 2 mM MgCl$_2$, 0.2 mM dNTPs, 2.5 μL PCR Buffer, 1 μL Bovine serum albumin (BioLabs, Ispwich, MA, USA) and 0.5–1.5 μM of each primer. Polymerase chain reaction conditions were performed in a PTC-100 thermocycler (M.J. Research) and consisted of an initial denaturation step at 94°C for 2 min, followed by 35 cycles of 94°C for 30 sec, 58°C for 1 min, 72°C for 45 sec and a final extension at 72°C for 5 min. Sample purification and sequencing was performed by the UW High Throughout Genomics Center, Seattle, WA. Forward and reverse sequences for each individual were aligned and edited manually using Geneious 9.0.5 (http://

FIGURE 1 | Geographic distribution in Mexico of *Crocodylus moreletii* and *C. acutus*. Map of sampling localities for *Crocodylus* acutus and *C. moreletii* throughout Mexico. Every dot indicates individuals and sampling localities; the gray lines represent the river system. In order to place the sampling localities at the level of state in the map and in association with their description with **Table S1** (Supporting Information), we grouped them (dotted circles) and codes refer to: TM, Tamaulipas; SL, San Luis Potosí; VZ, Veracruz; TB, Tabasco; CP, Campeche; YC, Yucatán; QR, Quintana Roo; QZ, Cozumel; BC, Banco Chinchorro; CS, Cañon del Sumidero; CH, Chiapas; VT, Ventanilla; OX, Chacahua; AC, Acapulco; GR, Guerrero; MC, Michoacan; AM, Amela; PV, Palo Verde; JL, Jalisco and NY, Nayarit.

www.geneious.com; Kearse et al., 2012). All chromatograms were reviewed to visually identify heterozygous sites and errors in base calling.

We successfully amplified 1,009 bp and 721 bp of cyt *b* and of tRNA^{Pro}-tRNA^{Phe}-d-loop regions, respectively, across 271 individuals. The concatenated dataset included 123 polymorphic sites, 112 parsimony informative sites and 86 unique haplotypes (GenBank Accession numbers Cytb: MH749473-MH749743; dloop: MH749744-MH750014). The estimated Tr/Tv ratio was 3.74, with mean nucleotide frequencies as 29.13% A, 30.08% C, 13.95% G and 26.84% T. The best-fit model selected for our data was TIM3+I+G (pinv=0.66; α =0.67), and the closest equivalent, GTR+I+G, was used in subsequent analyses.

Genotyping by Sequencing (GBS) and SNP Calling

High quality DNA (0.5 μg/ul) of a subset of 190 individuals was sent to the Genomic Diversity Facility (GDF) at Cornell University for Genotyping by Sequencing library construction (GBS, Elshire et al., 2011) and sequencing services. Libraries were prepared for two plates (190 samples), which included two blanks as a negative control, using the EcoT221 enzyme for genome complexity reduction. A sample-specific barcoded adapter and a common adapter were ligated to the sticky ends of fragments, and samples were pooled and purified. Pooled samples were amplified through PCR reaction, purified and quantified for sequencing on the Illumina Hi-seq2500 (Illumina Inc, USA). The whole genome sequence (WGS) of *Crocodylous porosus* (Green et al., 2014) was used as a reference genome. The software TASSEL v.3.0 (Glaubitz

et al., 2014) was used by GDF-Cornell University for calling Single nucleotide polymorphisms (SNPs). Next, we performed the following filtering processes with VCFTOOLS (Danecek et al., 2011): minor allele frequency = 0.05, depth = 8, maximum missing data was allowed for 18 samples, and only biallelic SNPs were included. We further eliminated all the missing data for the reference and alternate alleles. We obtained high quality data for 172 samples and 264,987 raw SNPs, for a final data set of 12,879 SNPs after filtering. For the different genomic analyses performed, we used PGDSPIDER (Lischer and Excoffier, 2012) to convert vcf files to other input file formats; GENEPOPEDIT (Stanley et al., 2017) was used to generate the input files for the BGC program (Gompert and Buerkle, 2012).

Genetic Structure and Hybridization Patterns

We examined structure and differentiation of hybrids at different genetic (microsatellite and SNPs loci and mitochondrial haplotypes) levels, with an array of analytical procedures. First, a spatially independent analysis based on a Bayesian inference of admixture proportions, that is the proportion of each individual's genome derived from each source population i (q_i), was performed using STRUCTURE 2.3.4 (Pritchard et al., 2000) for the entire dataset (374 individuals) and their microsatellites genotypes. The probability of individual assignment into population clusters (K) was estimated without prior information on the origin of individuals, conducting several tests with a maximum number of populations from $K = 2$ to $K = 10$, with the admixture model and correlated allelic frequencies,

a length value set at 1,000,000 iterations, a burn-in period of 100,000 MCMC repetitions. For the mtDNA sequences, we set a maximum number of populations from $K = 2$ to $K = 6$, with the non-admixture model and correlated allelic frequencies, a length value set at 100,000 iterations, a burn-in period of 1,000 MCMC repetitions, and 10 simulations per K. Each test yielded a log likelihood value of the data (Ln probability for K) and we used the Evanno's ΔK test to estimate the maximum number of clusters (Evanno et al., 2005); data were processed with STRUCTURE HARVESTER (Earl and vonHoldt, 2012) and results were summarized and averaged using CLUMPP 1.1.2 (Jakobsson and Rosenberg, 2007). For the SNPs data, we used FASTSTRUCTURE (Raj et al., 2014), with default convergence criterion and priors, and five replicates for each K-value. The optimal K-value was estimated with the Python-based tool CHOOSEK (included in the FASTSTRUCTURE package) and results of all replicates were summarized using CLUMPAK (Kopelman et al., 2015).

Next, we performed a Maximum-likelihood estimation of hybrid indexes (h) with the R package INTROGRESS (Gompert and Buerkle, 2010) to explore hybridization based on microsatellites loci. Non-admixed individuals were identified from admixture proportions based on the STRUCTURE results, which were then used to define the source populations to polarize the hybrid index, where a value of $h = 0$ corresponds to non-admixed *C. moreletii* and $h = 1$ corresponds to non-admixed *C. acutus* individuals (see Pacheco-Sierra et al., 2016 for details). In addition, we used the Bayesian genomic cline model implemented in the BGC software (Gompert and Buerkle, 2012) to estimate a hybrid index with SNPs. Non-admixed and hybrid individuals were also defined based on the microsatellite data. In total 125 samples were included as hybrids, while 18 and 29 as non-admixed (parental), for *C. moreletii* and *C. acutus* respectively. We used the model of genotype uncertainty, which is recommended for next generation sequencing data, based on 100,000 MCMC repetitions, recording samples from the posterior distribution every 5th step, and with a burn-in period of 30,000. The hybrid index included values from 0.80 to 0.38, where higher and lower values indicated genetic similarities to non-admixed (parental) *C. acutus* and *C. moreletii*, respectively.

Genetic Diversity

We estimated genetic diversity as number of alleles, effective and private alleles, observed and expected heterozygosity values, and inbreeding (F_{IS}) coefficients using GENALEX 6.5 (Peakall and Smouse, 2012) and GENEPOP 4.6.9 (Rousset, 2017), per genetic group identified (see Results) with STRUCTURE (microsatellites) and FASTSTRUCTURE (SNPs), respectively. To explore the pattern of genetic structure between genetic groups, we also estimated Nei's genetic distances (Nei et al., 1983), migration (Nm), and population subdivision (F_{ST}, Wright, 1951; R_{ST}, Slatkin, 1991).

The results of hybrid index and genetic structure from microsatellites data also served as a basis to establish genetic clusters across all the individuals of *Crocodylus* and further analyze hybridization patterns with the mitochondrial data (mtDNA) (mtDNA analyses are explained below). For that, we

performed a Principal Component Analysis (PCA) to examine patterns of variation and covariation between the observed polymorphic sites for the unique haplotypes. We constructed plots based on the first two-principal components, incorporating the hybrid differentiation information: the hybrid index (h) and the assignment proportion (qi) per individual. Next, we performed a Canonical Discriminant Analysis (CDA), based on Fisher linear discriminant analysis (Fraley and Raftery, 2002), following the same classification criteria (h and qi as in the PCA), in order to explore if the CDA could distinguish the main groups we had hypothesized (non-admixed and hybrids clusters); all analyses were performed in R.

Phylogenetic (mtDNA) and Diversity Analyses

We further explored the hybridization process between *C. acutus* and *C. moreletii* and how they are genetically structured, spatially and temporally, with the aim of evaluating the historical onset of hybridization and diversification and the hybrid zone processes. First, to explore the contribution of each species into the hybridization process, we evaluated the genealogical relationships between individuals via phylogenetic analyses both for each mitochondrial gene separately and for the concatenated set, considering individuals as operational units (OTU's). We used the Akaike information criterion scores as implemented in the program jModelTest 0.1.1 (Posada, 2008) to select the best fitting model of sequence evolution for our datasets, which were subsequently used for maximum-likelihood (ML) and Bayesian inference (BI) methods.

ML was conducted with PhyML 3.0 (Guindon et al., 2010), using a Nearest-Neighbor Interchange and Subtree-Prune and Regraft moves (NNI+SPR) for branch length and topology optimization. Clade support was assessed with 1,000 non-parametric bootstrap replicates. For the Bayesian inference, implemented in MrBayes 3.1 (Huelsenbeck and Ronquist, 2001), we used default settings and four chains sampled every 1,000 generations for 50 million generations. Convergence and stationarity within chains were visualized with TRACER 1.4 (Drummond and Rambaut, 2007); 25% of generations were discarded as burn-in. We estimated the 50% majority-rule consensus topology and posterior probabilities for each node with the remaining trees. We included sequences from *C. intermedius* (NC015648) and *C. rhombifer* (NC024513), the sister species of *C. acutus* and *C. moreletii*, respectively, and chose the African *C. niloticus* (AJ810452) as outgroup of the American crocodilians (Oaks, 2011). Second, to investigate the relationship at the level of mitochondrial haplotypes, we constructed an unrooted network among unique haplotypes, using POPART (available at: http://popart.otago.ac.nz), based on the median-joining method (Bandelt et al., 1999). We integrated the information about the admixture proportion (qi) from STRUCTURE ($K = 5$) with the haplotype network.

Finally, we performed a ML analysis with the SNPs data using FASTTREE 2.1 (Price et al., 2009), with the NNI+SPR search for topology and branch-length optimization and a General-Time Reversible with single rate per-site (GTR+CAT) nucleotide

substitution model. Because FASTTREE considers those SNP's identified as fixed within individuals (i.e., homozygous), but polymorphic among individuals, only 78% of the total alignment (10,045) was included in this analysis. We constructed a SNP-based outgroup from the Australian saltwater crocodile *Crocodylus porosus* whole genome (ftp://ftp.crocgenomes.org/pub/ICGWG/Genome_drafts/crocodile.current/croc_sub2.assembly.fasta.gz; Green et al., 2014) with SAMTOOLS 1.X (Li, 2011). With the aim to compare the phylogenetic information provided by nuclear and mtDNA, we performed a second ML analysis with the mitochondrial sequences but including only the 172 individuals for which we had SNPs genotypes; *C. porosus* (DQ273698) was used as outgroup.

Divergence Time Estimation

In order to establish the time frame of the hybridization process between *C. acutus* and *C. moreletii* along their distribution area, we estimated divergence times (time to the most recent common ancestor, TMRCA). The analysis was done using a coalescent constant size tree prior and a relaxed-clock dating implemented in BEAST 1.8.2 (Drummond et al., 2012), which allows estimation of divergence times incorporating simultaneously the rate heterogeneity and the uncertainty of the substitution parameters, tree topology, and calibration ages considering a Bayesian framework (Drummond et al., 2006; Brandley et al., 2010). We used unique haplotypes from a modified concatenated dataset discarding tRNA's and including only 1,576 bp of cyt *b*+d-loop mtDNA; the latter with the purpose of incorporating sequences from Oaks (2011) and the following calibration points: the *Mecistops-Crocodylus* split, the split between *C. niloticus* and American crocodiles, the divergence within the *C. intemedius-C. acutus*, and within *C. rhombifer-C. moreletii* clades. Final estimation included the GTR+I+G model of evolution across all gene and codon positions, 100,000,000 generations sampled every 5,000th and 25% of initial generations discarded as burnin. Convergence and stationarity were visualized with Tracer v.1.6 (http://tree.bio.ed.ac.uk/software/tracer/).

Bayesian Species Delimitation With Genomic Data

Taking into account the phylogenetic position of individuals from the Caribbean islands and the Pacific, which group into different clades and could potentially be distinct lineages within *C. acutus*, as well as a separated set of hybrids (see Results), we performed a coalescent Bayesian species delimitation analysis. We constructed a genome subset that retained a total of 500 SNPs under linkage equilibrium, avoiding missing data ("N" bases) and including only individuals with high values of assignment probability (>0.85, $K = 5$) to each cluster inferred with FASTSTRUCTURE ($n = 89$).

We performed the Bayes Factor Species Delimitation method (Leaché et al., 2014) with the SNAPP 1.1.1 and MODEL-SELECTION 1.3.4 modules (Bryant et al., 2012) included in BEAST 2.4.5 (Bouckaert et al., 2014). We tested two species (evolutionary units) models, based on our clustering and phylogenetic results: a) four lineages hypothesis: non-admixed (parental) *C. moreletii*, Yucatan peninsula hybrids, hybrids, and

C. acutus (Caribbean+Pacific); and b) five lineages hypothesis: non-admixed (parental) *C. moreletii*, Yucatan peninsula hybrids, hybrids, Caribbean islands *C. acutus*, and non-admixed Pacific *C. acutus*. We conducted a path sampling with 14 steps (100,000 MCMC runs and 10,000 pre burn-in generations each) to estimate marginal likelihoods for each hypothesis. The best-supported species model was based on the comparison of the Bayes factor support between models. The best-supported species tree from the final path-sampling step (10% burn-in) was visualized with DENSITREE 2.2.1 (Bouckaert, 2010).
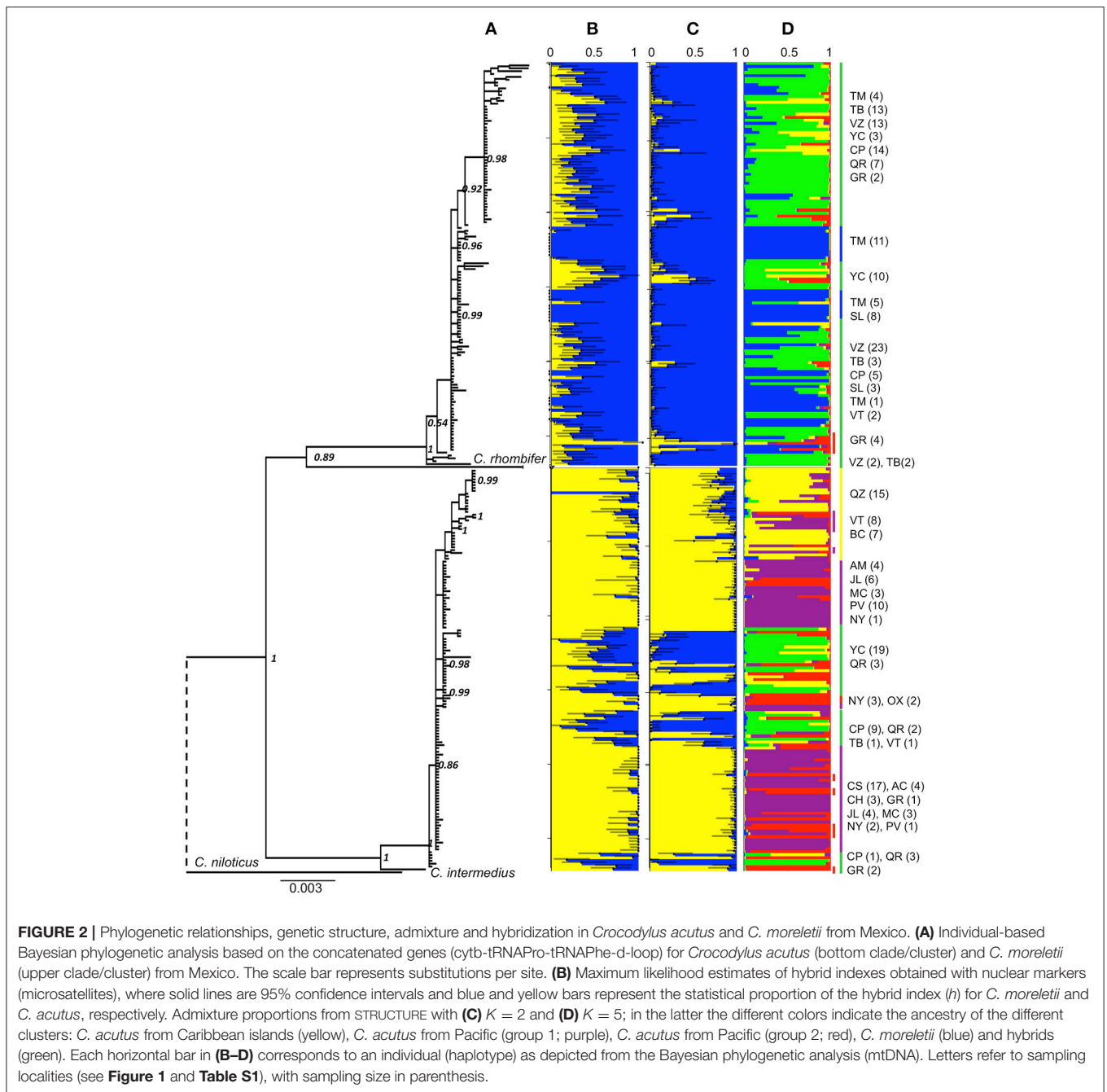
RESULTS

We amplified 12 microsatellites loci for the entire 374 crocodile individuals, amplified 1,730 bp of mtDNA across 271 individuals, and obtained high GBS quality data for 172 samples (12,879 SNPs).

Genetic Structure and Hybrid Indices

The results of the Bayesian inference of admixture proportions (STRUCTURE; using microsatellites) for the entire data set (374 individuals) showed a sequential admixture proportion, where $K = 2$ to $K = 5$ (qi) gradually separate the non-admixed from the hybrid individuals (**Figure S1**). The number of clusters detected by the Evanno's test for the mtDNA dataset (271 individuals) that best explain the results indicated a $K = 3$ [Ln Pr($K = 3$) $= -5567.14$], and for the microsatellites it was 5 [Ln Pr(K=5) $= -11929.00$]. In the latter, the Gulf of Mexico and the Pacific populations form five distinguishable separate clusters (**Figures 2C,D**, and **Figures S2B,C**): two different groups of non-admixed *C. acutus* from the Pacific (group 1, purple; group 2, red), non-admixed *C. acutus* from the Caribbean islands (yellow), non-admixed *C. moreletii* from the Gulf of Mexico (blue), and hybrids predominantly from the Gulf of Mexico but also within some Pacific localities (green). Indeed, this pattern is highly concordant with the hybrid index results (h): a few non-admixed and isolated individuals for both *C. moreletii* and *C. acutus* and a range of hybrids throughout the entire distribution (**Figure 2B**, and **Figure S2A**). The Bayesian assignment results based on SNPs for the subset of 192 individuals, the ChooseK method showed that the number of clusters that best explain the structure also was $K = 5$ (Ln $= -0.61091$; **Figure 3**). Interestingly, three groups are concordant with the microsatellite structure patterns (**Figure 3B,C**): non-admixed *C. moreletii* (nACm, blue), hybrids (Hyb, green), non-admixed *C. acutus* from the Caribbean islands (nACiCa, yellow), whereas only one group for the Pacific individuals (non-admixed Pacific *C. acutus*; nAPCa, red) and a second differentiated hybrids group (Yucatan peninsula hybrids; YpHyb, purple) were depicted. The hybrid index results based on microsatellites and SNPs were compared with a linear regression using the R software (R Development Core Team, 2016), which was highly significant [$R^2 = 0.79$, $F_{(1,170)} = 659$, $P < 0.05$] (**Figure S3**).

Genetic Diversity and mtDNA Hybridization

Genetic diversity based on microsatellite loci was lowest in the non-admixed *C. moreletii* population ($H_O = 0.227$; $H_E = 0.380$),

FIGURE 2 | Phylogenetic relationships, genetic structure, admixture and hybridization in *Crocodylus acutus* and *C. moreletii* from Mexico. **(A)** Individual-based Bayesian phylogenetic analysis based on the concatenated genes (cytb-tRNAPro-tRNAPhe-d-loop) for *Crocodylus acutus* (bottom clade/cluster) and *C. moreletii* (upper clade/cluster) from Mexico. The scale bar represents substitutions per site. **(B)** Maximum likelihood estimates of hybrid indexes obtained with nuclear markers (microsatellites), where solid lines are 95% confidence intervals and blue and yellow bars represent the statistical proportion of the hybrid index (h) for *C. moreletii* and *C. acutus*, respectively. Admixture proportions from STRUCTURE with **(C)** $K = 2$ and **(D)** $K = 5$; in the latter the different colors indicate the ancestry of the different clusters: *C. acutus* from Caribbean islands (yellow), *C. acutus* from Pacific (group 1; purple), *C. acutus* from Pacific (group 2; red), *C. moreletii* (blue) and hybrids (green). Each horizontal bar in **(B–D)** corresponds to an individual (haplotype) as depicted from the Bayesian phylogenetic analysis (mtDNA). Letters refer to sampling localities (see **Figure 1** and **Table S1**), with sampling size in parenthesis.

which showed no private alleles (**Table S3**); hybrids showed intermediate values. All genetic groups showed high inbreeding ($F_{IS} = 0.377$–0.448). The genetic differentiation results (F_{ST} and R_{ST}) were highly significant between genetic groups ($P < 0.001$; **Table S4**). Accordingly, the highest number of migrants and lower genetic distance were between adjacent populations along the geographic gradient (**Table S5**). Genetic diversity and group genetic differentiation based on SNPs were concordant, where non-admixed *C. moreletii* (nACm) showed the lowest values ($H_O = 0.130$; $H_E = 0.135$), seen also for Pacific *C. acutus* (nAPCa; $H_E = 0.109$), while hybrids showed the highest values ($H_O =$

0.257; $H_E = 0.275$). Also, nACm and both groups of hybrids (Hyb, YpHyb) exhibited heterozygosity deficiency (**Table S6**). Differentiation also followed an isolation by distance pattern, which was highest between nACm and nAPCa groups ($F_{ST} = 0.806$), while the lowest was between both hybrids YpHyb and Hyb ($F_{ST} = 0.127$) (**Table S7**).

The PCA (polymorphic sites from the mtDNA haplotypes; see results below) showed that the two first components explained 63% of the variation. When organized based on the classification of the hybrid index and the admixture proportions ($K = 5$), the PCA did not show a clear differentiation between clusters
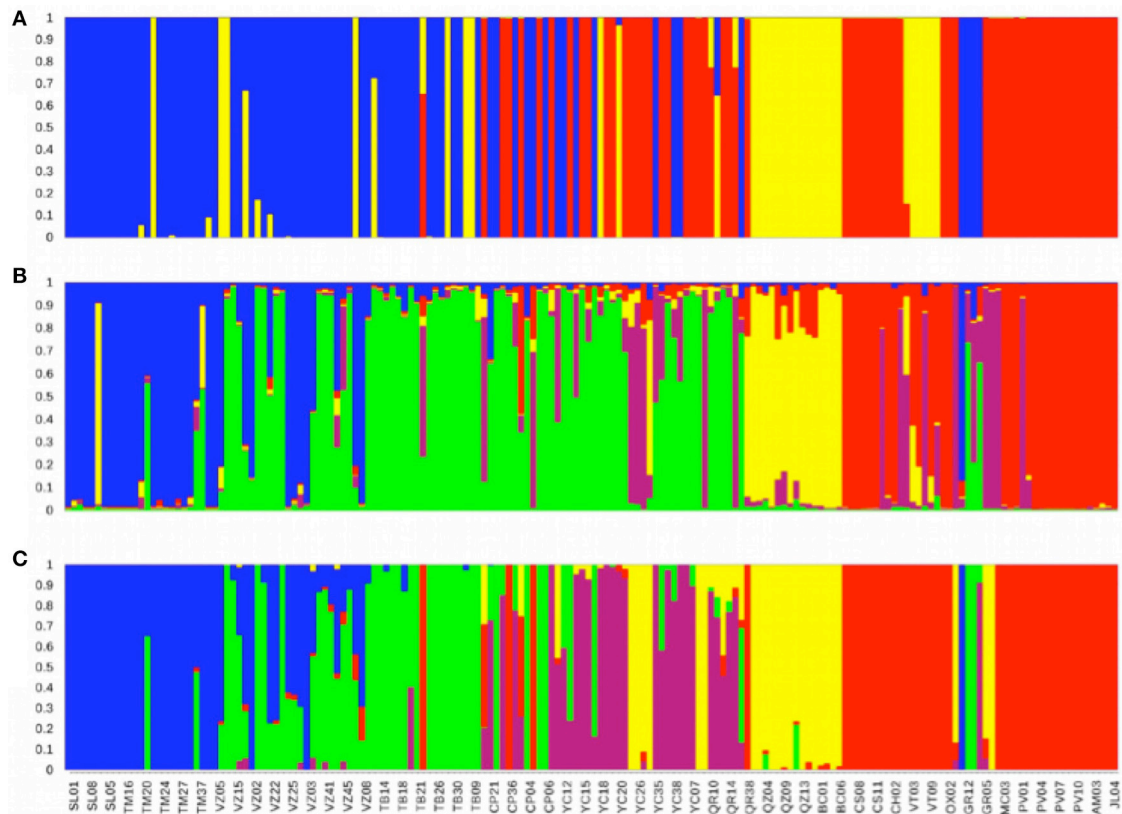
FIGURE 3 | Genetic structure, admixture, and hybridization in *Crocodylus acutus* and *C. moreletii*. Admixture proportions results for the 190 individuals subset, representing the maximum number of clusters for each genetic marker, based on **(A)** mtDNA ($K = 3$; STRUCTURE), **(B)** microsatellites ($K = 5$; STRUCTURE), and **(C)** SNPs ($K = 5$; FASTSTRUCTURE). Samples are ordered following the geographic cline from northeast to southeast along the Gulf of Mexico and Caribbean, following then along the Pacific from southwest to northwest (see **Figure 1**).

(**Figure S4A**). Contrastingly, the different groups can be readily observed with the CDA for $K = 5$ (Wilk's $\lambda = 0.025$, $P < 0.001$), where 80.1% of individuals are classified correctly and the hybrids group is significantly separated (**Figure S4B**, **Table S8**).
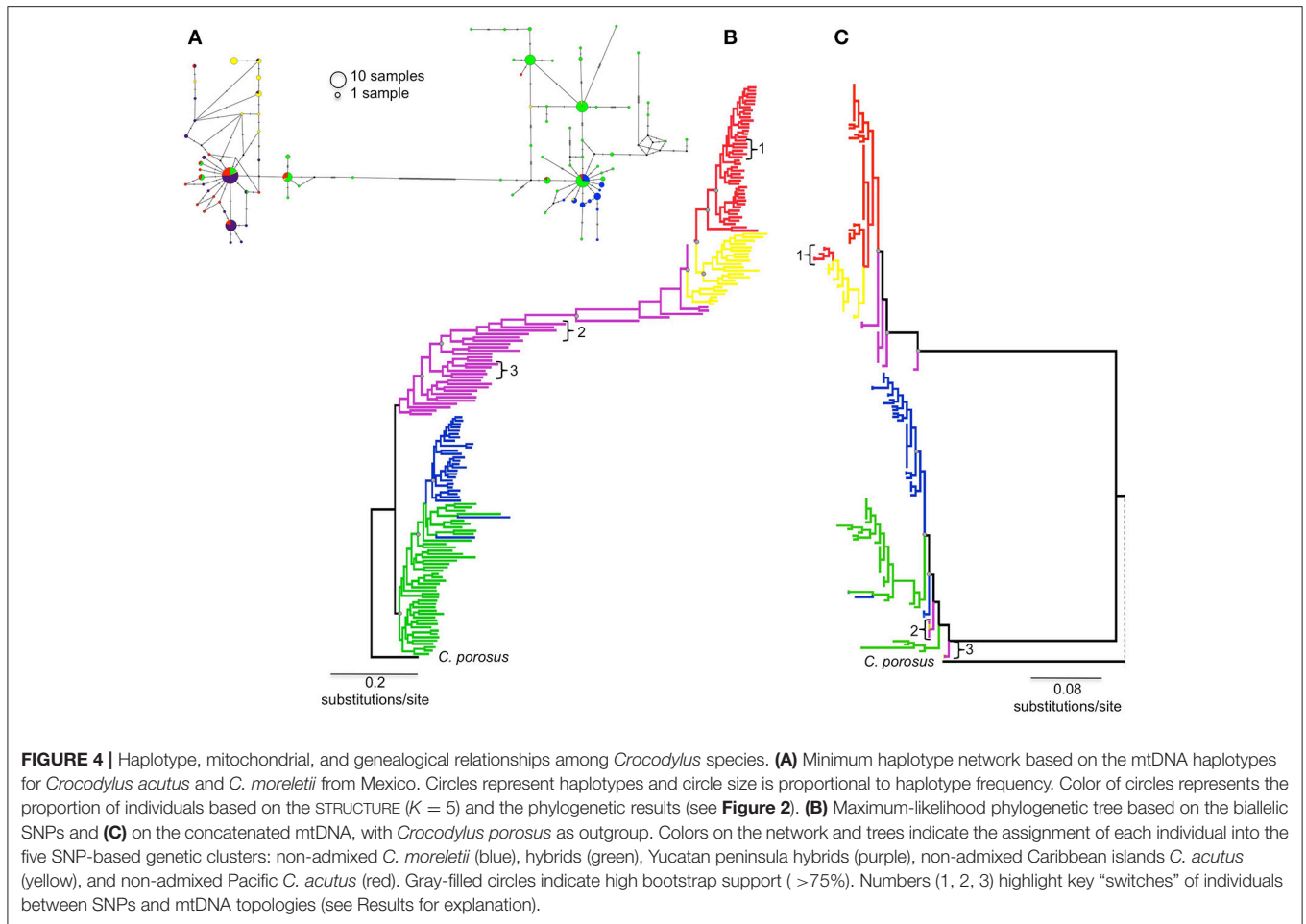
Phylogenetic (mtDNA) and Diversity Analyses

Phylogenetic inference methods (BI and ML) based on mtDNA showed the same topology, both with each gene separately (not shown) and with the concatenated dataset (**Figure S5**). Our phylogenetic results are similar to that of Oaks (2011) and Meredith et al. (2011), where *C. acutus* appears as a sister species of *C. intermedius* and *C. moreletii* is the closest relative to *C. rhombifer*. Within each species group, a pectinate-like pattern was recovered with very short branches, suggesting a recent and continuous diversification. The topology is highly concordant with the STRUCTURE clustering results and hybridization tests (see below).

By combining the mitochondrial and nuclear genetic information, a degree of geographic concordance is evident, in which the phylogenetic tree (Bayesian inference; **Figure 2A**), the hybrid index (**Figure 2B**), and the admixture assignment ($K = 2$ and $K = 5$; **Figures 2C,D**, respectively) show that the

genetic groups are readily separated by the microsatellites genetic ancestry information (**Figure 2A**): non-admixed *C. acutus* from the Caribbean islands (QZ and BC) are grouped in one clade with the non-admixed *C. acutus* (8 of 11 individuals) from the Pacific population Ventanilla (VT). This clade is embedded within the *C. acutus* entire clade, encompassing the non-admixed individuals from the Pacific (purple and red groups) and hybrid individuals from the Yucatan peninsula and the Pacific. A separate clade encompasses the non-admixed *C. moreletii* and hybrid individuals, where all haplotypes are from the Gulf of Mexico and the Yucatan peninsula, and only eight haplotypes from the Pacific (GR and VT). This clade includes one clearly differentiated group with all hybrids from TM, TB, VZ, YC, CP, QR, and GR, sister to an unresolved remaining set of haplotypes, but where non-admixed *C. moreletii* from TM and SL are each grouped separately, among the other hybrid individuals from the Gulf of Mexico (YC, VZ, TB, CP, SL) and VT.

The phylogenetic ML analyses based on the subgroup of individuals with both SNPs and mtDNA loci, with the Australian saltwater crocodile *Crocodylus porosus* as outgroup, were highly concordant (**Figures 4B,C**). Samples from the Gulf of Mexico and northern Yucatan peninsula were grouped in a clade with high support (>75%), in agreement with the putative *C. moreletii* distribution. In addition, samples from eastern
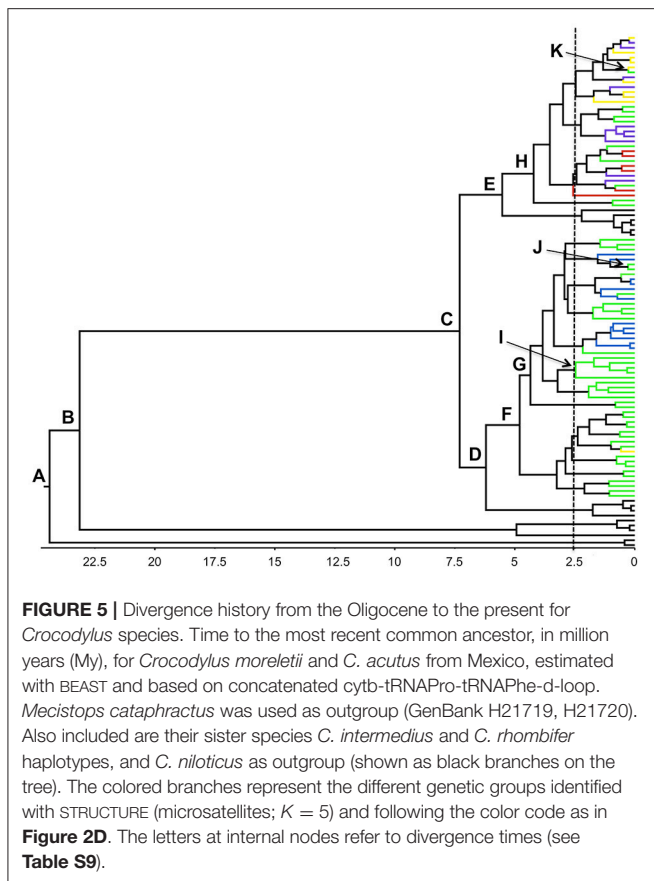
FIGURE 4 | Haplotype, mitochondrial, and genealogical relationships among *Crocodylus* species. **(A)** Minimum haplotype network based on the mtDNA haplotypes for *Crocodylus acutus* and *C. moreletii* from Mexico. Circles represent haplotypes and circle size is proportional to haplotype frequency. Color of circles represents the proportion of individuals based on the STRUCTURE (*K* = 5) and the phylogenetic results (see **Figure 2**). **(B)** Maximum-likelihood phylogenetic tree based on the biallelic SNPs and **(C)** on the concatenated mtDNA, with *Crocodylus porosus* as outgroup. Colors on the network and trees indicate the assignment of each individual into the five SNP-based genetic clusters: non-admixed *C. moreletii* (blue), hybrids (green), Yucatan peninsula hybrids (purple), non-admixed Caribbean islands *C. acutus* (yellow), and non-admixed Pacific *C. acutus* (red). Gray-filled circles indicate high bootstrap support (>75%). Numbers (1, 2, 3) highlight key "switches" of individuals between SNPs and mtDNA topologies (see Results for explanation).

Yucatan peninsula, Mexican Caribbean islands and Pacific coast are recovered as a different well-supported clade; the latter in agreement with the *C. acutus* identity. Interestingly, we found three key "switches" of individuals between SNPs and mtDNA topologies (**Figures 4B,C**), where switch 1 depicts individuals from a southern Pacific coast population (VT; **Figure 1**) and switches 2 and 3 are individuals from the Yucatan peninsula (YC; **Figure 1**). While switch 1 may be explained in terms of an incomplete lineage sorting within the *C. acutus*-like clade, individuals within switches 2 and 3 may be associated with the highly-mixed ancestry of their genome: *C. acutus* paternal ancestry and *C. moreletii* maternal contribution. Furthermore, three distinct clades are consistent with results from assignment (SNPs; FastStructure; **Figure 3C**) and phylogenetic (both SNPs and mtDNA based trees; **Figures 4B,C**) analyses: non-admixed *C. moreletii*, non-admixed Caribbean islands *C. acutus*, and non-admixed Pacific *C. acutus*. Furthermore, results from the haplotype network showed a pattern in agreement with the SNPs phylogenetic topology (**Figure 4A**): *C. moreletti* with a star-like configuration and several abundant haplotypes at the center of the network, and a group encompassing most hybrid individuals with higher association to *C. moreletii*; *C. acutus* exhibited a similar pattern, although here the dominant haplotypes are

at the center (Pacific haplotypes) and the haplotypes from the Caribbean islands are on the edge, showing some unique and some widespread haplotypes with few mutations among them. The networks for each species are deeply separated (i.e., connected by 66 mutational steps), while the haplotypes in the middle joining the connection are hybrids from the Yucatan peninsula (mainly from YC and QR).

Divergence Time Estimation

The dataset used for divergence time estimation showed 89 unique haplotypes, 51 of which were assigned to *C. moreletii* and 38 to *C. acutus*. The topology obtained with BEAST was concordant with the *C. moreletti* and *C. acutus* different lineages, which diverged around 7.3 million years ago (My) (95% HPD: 6.09-8.79; node C in **Figure 5**). In accordance with our estimations, the divergence between *C. rhombifer* and the Mexican *C. moreletii* occurred 6.2 My (95% HPD: 5.21–7.33; node D), while that between *C. intermedius* and *C. acutus* happened 5.52 My (95% HPD: 4.02–7.23; node E). The TMRCA for *C. moreletii* was calculated at 4.8 My (95% HPD: 3.91–5.76; node F) and for *C. acutus* was dated about 4.21 My (95% HPD: 2.77–5.73; node H) (**Table S9**). Based on these mitochondrial results that depict the history of females, the

FIGURE 5 | Divergence history from the Oligocene to the present for *Crocodylus* species. Time to the most recent common ancestor, in million years (My), for *Crocodylus moreletii* and *C. acutus* from Mexico, estimated with BEAST and based on concatenated cytb-tRNAPro-tRNAPhe-d-loop. *Mecistops cataphractus* was used as outgroup (GenBank H21719, H21720). Also included are their sister species *C. intermedius* and *C. rhombifer* haplotypes, and *C. niloticus* as outgroup (shown as black branches on the tree). The colored branches represent the different genetic groups identified with STRUCTURE (microsatellites; *K* = 5) and following the color code as in **Figure 2D**. The letters at internal nodes refer to divergence times (see **Table S9**).

TMRCA of the earliest clade that shows hybridization (hybrid haplotypes) between the two species in the tree can be traced back starting approximately 2.47 million years (node I) to as early as 230,000 years ago (node K).

Bayesian Species Delimitation

Our results from the marginal likelihood estimation (Ln = −14588.68) and the Bayes factor both showed strong statistical support for the hypothesis of five evolutionary units (EU's) (**Figure 6**). Moreover, the five EU's model is highly concordant with the nuclear (SNPs; **Figure 3C**) clustering and phylogenetic analyses (**Figure 4**), supporting the split of *Crocodylus acutus* into two lineages. The latter suggests the occurrence of a different species of *Crocodylus*, distributed on the Mexican Caribbean (Cozumel and Banco Chinchorro) and the coast of Quintana Roo, clearly distinct to that from *C. acutus* on the Pacific coast.

The position of the hybrids (Hyb) lineage within the Bayesian phylogenetic hypothesis shows that the entire group is closely related to the *C. moreletii* lineage (nACm), with only some genetic contribution from the *C. acutus* lineages (nAPCa and nACiCa), in agreement with the ML phylogenetic relationships, which also showed most hybrid individuals with strong admixture from *C. moreletii*. Interestingly, a second, differentiated hybrids group is supported (YpHyb), geographically localized on the Yucatan peninsula and with higher contribution from *C. acutus*.
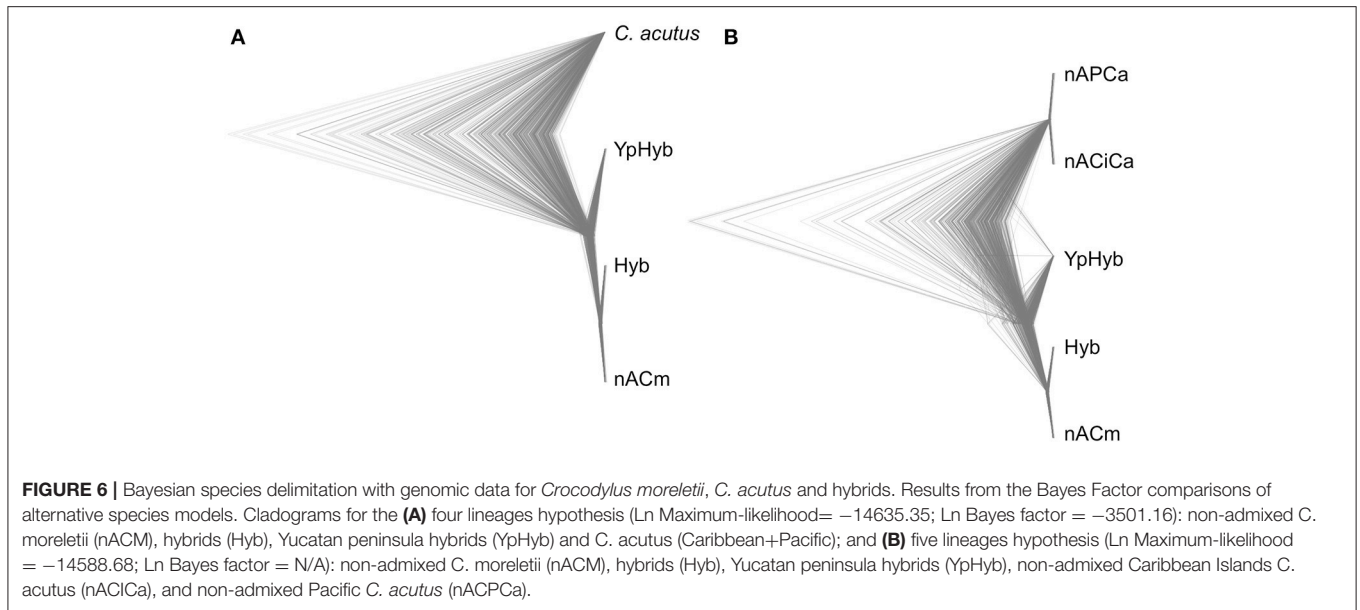
DISCUSSION

Phylogeography, Ancient Hybridization, and Crocodiles Beyond Borders

We were able to describe the diversification of these crocodiles' lineages, where their initial divergence occurred 7.3 million years ago. We expected a timeframe supporting that hybridization between these crocodiles was an ancient process (Pacheco-Sierra et al., 2016); in fact, the time to the most recent ancestor of the earliest clade showing hybridization (hybrid haplotypes) can be traced back starting approximately 2.47 million years to as early as 230,000 years ago. Furthermore, the historical signature of hybridization and asymmetric introgression between the two species is supported by the mtDNA haplotype network showing that *C. acutus* and *C. moreletii* are deeply separated, i.e., connected by 66 mutational steps and joined by hybrid individuals from the historical zone of sympatry on the Yucatan peninsula.

Intriguingly, despite the expectation that geographic barriers would prevent the genetic exchange between populations from the Gulf of Mexico and those along the Pacific—where only *C. moreletii* and *C. acutus* are respectively naturally distributed (Thorbjarnarson, 1989; Pacheco-Sierra et al., 2016)—, we found "crocodiles beyond borders." Our evidence shows an historic genetic exchange between individuals from both sides of the continent: hybrids have admixture proportions not only from their neighbors *C. acutus* (Caribbean islands), but also from *C. acutus* inhabiting Pacific localities. Another key finding is the presence of specific *C. acutus* alleles (**Figures 2D**, **3B,C**) and haplotypes (**Figures 3A**, **4**) from the Caribbean islands on individuals along the southern Pacific, together with evidence of *C. moreletii* haplotypes on individuals from two Pacific localities in Guerrero and Oaxaca. Serrano-Gómez et al. (2016) found molecular evidence of the presence of *C. moreletii* in *C. acutus* populations from Oaxaca and Guerrero (Pacific), the origin of which is now here resolved.

The Quaternary climatic oscillations of the last 3 million years that triggered major warm and cold periods, including change of sea levels and floods (Hewitt, 2011), could have facilitated dispersal of crocodile individuals between Atlantic and Pacific coasts, as observed for other aquatic species (Razo-Mendivil et al., 2013). This is concordant with the geological history of the Yucatan peninsula, which slowly emerged from south to north, coinciding with historical connectivity of rivers across the southern narrowest part of Mexico, near the Isthmus of Tehuantepec (Vázquez-Domínguez and Arita, 2010). Along this Pacific coast region is that we found hybrids and haplotypes from the Caribbean islands. Accordingly, a likely biogeographic scenario (see **Figure 7**) for these two crocodile species, based on their geographic distribution, estimated divergence times (**Figure 5**), and genomic-based clustering results, involved first the split between *Crocodylus moreletii* (Mexico) and its sister species *C. rhombifer* (Cuba) along the Caribbean region, occuring around 6.2 million years ago (My); later, the dispersal of an ancestral *Crocodylus acutus* through Central America, reaching the Mexican region ca. 4.7–3.5 My and the arrival of *C. acutus* to the Yucatan peninsula about 3 My, when the oldest hybridization

FIGURE 6 | Bayesian species delimitation with genomic data for *Crocodylus moreletii*, *C. acutus* and hybrids. Results from the Bayes Factor comparisons of alternative species models. Cladograms for the **(A)** four lineages hypothesis (Ln Maximum-likelihood= −14635.35; Ln Bayes factor = −3501.16): non-admixed C. moreletii (nACM), hybrids (Hyb), Yucatan peninsula hybrids (YpHyb) and C. acutus (Caribbean+Pacific); and **(B)** five lineages hypothesis (Ln Maximum-likelihood = −14588.68; Ln Bayes factor = N/A): non-admixed C. moreletii (nACM), hybrids (Hyb), Yucatan peninsula hybrids (YpHyb), non-admixed Caribbean Islands C. acutus (nACICa), and non-admixed Pacific *C. acutus* (nACPCa).

events could have initiated; then the first secondary contact zone on the northern Yucatan peninsula was formed nearly 2.5 My, while *C. acutus* dispersed through the Isthmus of Tehuantepec reaching the Pacific; around the same time, 3–2.5 My, *C. acutus* also dispersed into de Mexican Caribbean region and likely to other Central American islands and cays; meanwhile, the hybridization process continued toward both the north and the south developing the geographic cline we observe in the present (**Figure 7**).

A more recent history is depicted by the contemporary high genetic distance and significant differentiation between the non-admixed populations of both species, while adjacent groups along the Gulf have lower genetic distance with some migration. The lowest genetic diversity values and no private alleles exhibited by the parental *C. moreletii* populations clearly correlate with their extreme isolation and differentiation. Hence, the latest outcome of the hybridization process suggested by our results is a pattern where hybrids dispersed from the sympatry zone (Yucatan peninsula) along two directions, northwards isolating the current non-admixed *C. moreletii* populations (SL, TM; **Figure 1**), and southwards where non-admixed *C. acutus* individuals now remain in the Caribbean islands.

Species Boundaries and Hybrids: Taxonomic and Conservation Challenges

Despite hybridization between crocodiles has long been known, even as far as involving Egyptian Nile *Crocodylus niloticus* mummies and modern individuals (Hekkala et al., 2015), it is not until very recently that its recognition and conservation-related concerns have been actually raised. Specifically, involving the Cuban crocodile *Crocodylus rhombifer*, a critically endangered island endemic that hybridizes with *C. acutus* (Milián-García et al., 2015), a case that hit the spotlights in 2016, described as a conservation conundrum (Reardon, 2016). Certainly, our study

proves that *C. acutus* and *C. moreletii* pose a few conservation conundrums of their own.

The identification of hybrid species and of species boundaries in these natural systems has proven rather difficult (Gompert and Buerkle, 2016; Elgvin et al., 2017; Gompert et al., 2017). Different characteristics of the hybridization process and the hybrid zone can hinder the identification of species boundaries, especially where hybrid lineages get swamped by the homogenizing effect of gene flow from either parental species (Elgvin et al., 2017), or where boundaries are genetically fuzzy (Gompert and Buerkle, 2016; Pacheco-Sierra et al., 2016) as in these crocodile species. Nonetheless, our species delimitation analyses showed remarkable results, with strong statistical support for the hypothesis of five evolutionary units (EU's; **Figure 6**), supporting the split of *Crocodylus acutus* into two lineages, indicating the occurrence of a different species, one as isolated populations distributed on the Mexican Caribbean (Cozumel and Banco Chinchorro), clearly distinct from another lineage along the Pacific coast. Notably, phylogenetic studies thus far (Meredith et al., 2011; Oaks, 2011) have not evaluated the taxonomy of Crocodylia at finer levels, which has hindered having a truly updated phylogeny. There are more examples that have also identified significantly differentiated *C. acutus* populations (i.e., distinct lineages), for instance in Cuba and other localities in the Antilles (Milián-García et al., 2015). Hence, such findings evidence that a full taxonomic revision of *Crocodylus acutus* is urgently needed, and that it should be split, at least along its Mexican distribution, into two species, from the Pacific and from the Caribbean; the populations on the islands have indeed been identified as the last refuge with genetically pure American crocodiles in the Mexican Caribbean (Machkour-M'Rabet et al., 2009; Pacheco-Sierra et al., 2016).

The role of hybridization and admixture in species conservation is still amply debated, mainly for its potential
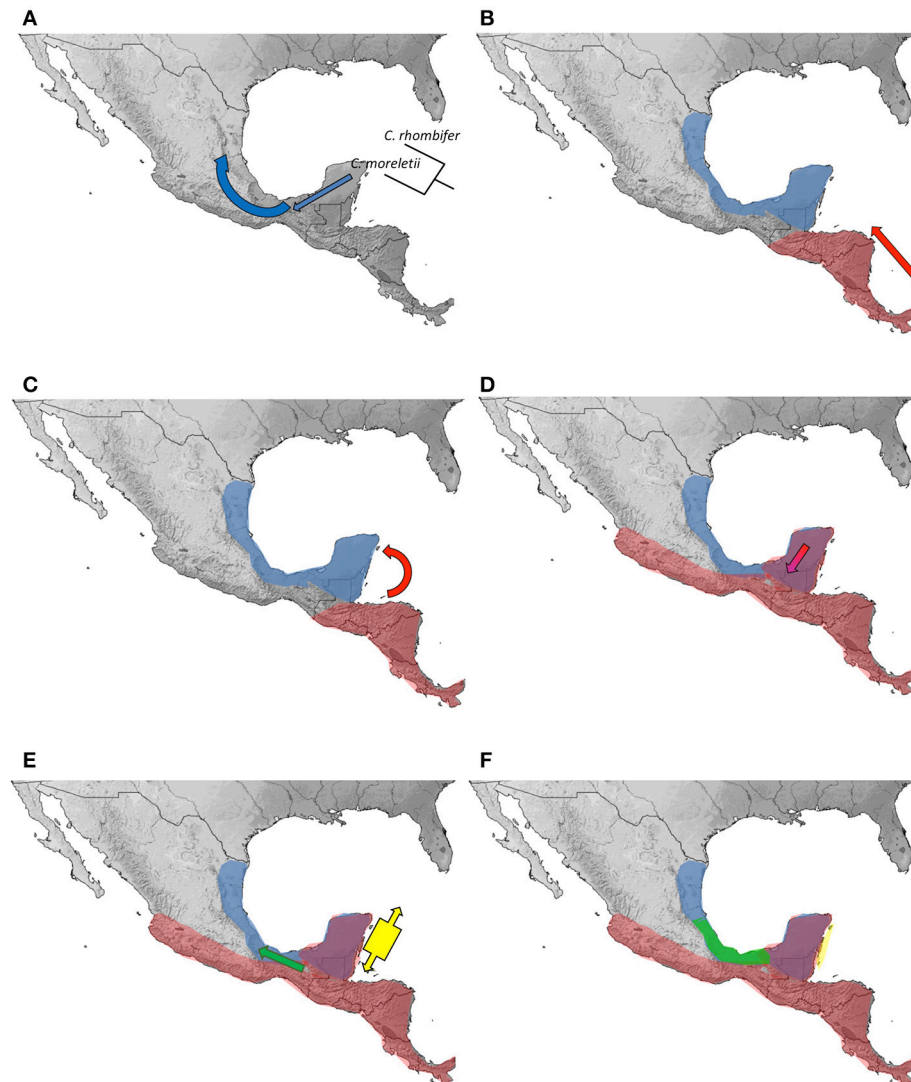
FIGURE 7 | Biogeographical hypothesis about the secondary contact and hybridization process between *Crocodylus acutus* and *C. moreletii*, based on their geographic distribution, estimated divergence times, and genomic-based clustering results: **(A)** split between *Crocodylus moreletii* (Mexico) and its sister species *C. rhombifer* (Cuba), occuring around 6.2 million years ago (My) and dispersal to the Yucatan peninsula, **(B)** ancestral *Crocodylus acutus* dispersed through Central America, reaching the Mexican region ca. 4.7–3.5 My, **(C)** *C. acutus* arrival to the Yucatan peninsula about 3 My, when the oldest hybridization events could have initiated, **(D)** the first secondary contact zone on the northern Yucatan peninsula was formed nearly 2.5 My; *C. acutus* dispersed through the Isthmus of Tehuantepec reaching the Pacific and followed along the coast, **(E)** about 3–2.5 My ago *C. acutus* also dispersed into the Mexican Caribbean region and likely to other Central American islands and cays, **(F)** the hybridization process continued toward both the north and the south, reaching the northernmost distribution along the Gulf of Mexico, and forming the cline we observe in the present.

effect on wild species extinction, whether via human-mediated hybridization or by the interbreeding of non-native or domestic species with native ones (Allendorf et al., 2001; Vaz Pinto et al., 2016; vonHoldt et al., 2018). At this juncture with these crocodiles, it is crucial to consider that the remaining isolated non-admixed *C. moreletii* populations have very low population numbers, facing extremely low genetic diversity, high homozygosity and null gene flow. Furthermore, the fact that they are so scarce along its historical natural distribution highlights that *C. moreletii* might be in the process of disappearing as a

species or, in our perspective, likely evolving via hybridization. Specifically, although introgressive hybridization is now increasingly viewed as a driving force in speciation, the overall pattern observed in these crocodiles resembles that of marine iguanas (MacLeod et al., 2015), likely going through a process of despeciation, also described for Darwin's finches (Grant and Grant, 2014), where one species is genetically absorbed into another via hybridization. Alternatively, since both *C. moreletii* and *C. acutus* (Caribbean islands) non-admixed populations are isolated, hybrids are becoming more and

more reproductively isolated from their parentals, potentially leading them toward a process of incipient speciation (Nolte and Tautz, 2010). The fact that all genetic groups showed some level of inbreeding (based on microsatellites loci) is in accordance with these scenarios, where parentals and hybrids exhibit a pattern of isolation by distance, and also by the presence of small and isolated populations that reproduce among close relatives. Indeed, many possible genetic effects can occur as a consequence of hybridization, one key factor being if it is the early or late stage of the process (Landry et al., 2007).

In addition, still prevalent is the controversy surrounding the setting of adequate conservation policies dealing with hybrids. Notably, we here demonstrate that the Gulf of Mexico hybrids lineage is closely related to the *C. moreletii* lineage, with only some genetic contribution from *C. acutus*. A similar biased unidirectional hybridization was found between *C. moreletii* and *C. acutus* from Belize populations (Hekkala et al., 2015). The second differentiated hybrid lineage, geographically localized on the Yucatan peninsula and radiating more recently, has higher contribution from *C. acutus*. Worryingly, few international and government conservation guidelines address hybrids (Jackiw et al., 2015), the International Union for Conservation of Nature (IUCN) only includes "apomictic plant hybrids" (IUCN, 2017), while hybrids with "recent lineage" (previous four generations of a lineage) of at least one species included in Appendices I or II are protected by the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES Resolution Conf 10.17 [Rev. CoP14]). This is of the utmost importance for these crocodiles because, despite that *C. moreletii* and *C. acutus* are on the Appendix II and I, respectively, their hybrids do not classify as recent and are, therefore, not protected. It is crucial to highlight that our results evidence that the hybridization between *C. moreletii* and *C. acutus* is not the result of human mediated movement as it has been suggested, since the admixture proportions and haplotypes are ancient, encompassing an extended, mosaic hybridization pattern, with no hybrids distributed on specific patches.

Despite the prevailing consensus that nearly every hybridization system is different and general conservation rules are not quite effective (Gompert and Buerkle, 2016), a very recent call for redefining the role of admixture in species conservation has been made, emphasizing that protection measures are required for taxa that have experienced gene flow and introgression over the course of their evolutionary histories (vonHoldt et al., 2018); *C. moreletii* and *C. acutus* certainly fit such criteria. Hence, adequate protocols to identify introgression and hybridization are urgently needed. The genetic approach we followed proved that combining information from different genetic markers significantly complements each other, yielding more robust results not easily discernable with limited genetic information. For instance, we show that the SNPs had a higher resolution for identifying the degree of hybridization, i.e., the proportion that each hybrid individual has from either parental (**Figures 3B,C**). Also, by combining the mitochondrial and nuclear genetic information we were able to more accurately establish the phylogeographic patterns followed by both species throughout their evolutionary history. Like Rosauer et al. (2016), we believe that molecular approaches based on phylogenetic and phylogeographic rigorous analyses use the best available data to inform conservation priorities, independent of taxonomy, and can play a key role in informing conservation prioritization. Indeed, our findings reinforce several urgent needs for the conservation of these species, where of great concern is the report from the latest monitoring program for *C. moreletii* (Sánchez Herrera et al., 2015), which specifies that "… the species is in good shape and with potential for developing sustainable production projects" and ranching program (Barrios-Quiroz and Cremieux-Grimaldi, 2018). Nevertheless, as we evidence, most populations considered as *C. moreletii* are in fact hybrids and, furthermore, the remaining non-admixed (parental) populations are actually in crucial need of protection as endangered with extinction. Importantly also is the fact that hybrids are not morphologically discernable (see Pacheco-Sierra et al., 2016), hence they cannot be identified in the field. Furthermore, laws like the USA Endangered Species Act indicates that hybrid individuals or populations merit protection if one or both of the parent species is considered to be endangered (vonHoldt et al., 2018). Outstandingly, we identified two divergent, geographically differentiated, crocodiles hybrids lineages, for which one parental species is critically endangered.

CONCLUDING REMARKS

We here reveal a complex evolutionary history regarding the *Crocodylus acutus* and *C. moreletii* hybridization system, which involves distinct lineages and evolutionary trajectories. Different lines of evidence from our study demonstrate two novel findings, first that what has been recognized as *Crocodylus acutus* is actually two distinct species, one from the Caribbean islands and one for the Pacific; secondly, that two evolutionary distinct hybrids lineages are identified, one encompassing a widespread Gulf of Mexico distribution and a second from the Yucatan peninsula. Importantly, the non-admixed parental populations of both *C. moreletii* and *C. acutus* are clearly genetically discernible. The phylogeographic patterns of the two species showed that the distinct lineages diversified throughout their hybridization history, a process that initiated from secondary contact on the Yucatan peninsula (Pacheco-Sierra et al., 2016), and which we evidence started around 2.5 million years ago. To fully protect the evolutionary potential of this extraordinary crocodiles system, it is essential to consider the evolutionary trajectory of their species and their hybrids, to adequately incorporate them into conservation programs.

DATA ACCESSIBILITY

Sanger sequences are available from GenBank (Accession numbers Cytb:MH749473 - MH749743; dloop: MH749744 - MH750014). https://doi.org/10.5061/dryad.rp257gt. Files are: (1)

microsatellites genotypes table that includes the microsatellite genotypes for 374 crocodile individuals, and (2) SNPs data, the vcf file containing 12879 SNP markers identified in *Crocodylus acutus*, *C. moreletii*, and hybrids from the study. The SNPs file is under an embargo until we finish a follow up genomic adaptation study and corresponding analyses we are currently performing with these data, and publish the results.

AUTHOR CONTRIBUTIONS

GP-S and EV-D conceived and designed the study. GP-S and JD-L performed fieldwork and sampling. GP-S, JP-A, MS-A, and EV-D analyzed and interpreted the data, and wrote the manuscript.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo. 2018.00138/full#supplementary-material

REFERENCES

Abbott, R. D., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N. et al. (2013). Hybridization and speciation. *J. Evol. Biol.* 26, 229–246. doi: 10.1111/j.1420-9101.2012.02599.x

Allendorf, F. W., Leary, R. F., Spruell, P., and Wenburg, J. K. (2001). The problems with hybrids: setting conservation guidelines. *Trends. Ecol. Evol.* 16, 613–622. doi: 10.1016/S0169-5347(01)02290-X

Bandelt, H. J., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol. Ecol. Biol.* 16, 37–48. doi: 10.1093/oxfordjournals.molbev.a026036

Barrios-Quiroz, G. and Cremieux-Grimaldi, J. C. (comp.). (2018). *Protocolo de Rancheo para el Cocodrilo de Pantano (Crocodylusmoreletii) en México.* Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (Conabio), México.

Beaupre, S. J., Jacobson, E. R., Lillywhite, H. B., and Zamudio, K. (2014). *Guidelines for the Use of Live Amphibians and Reptiles in Field and Laboratory Research.* American Society of Ichthyologists and Herpetologist. Available online at: http://www.asih.org/sites/default/files/documents/resources/guidelinesherpsresearch2004.pdf

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D. et al. (2014). BEAST 2: a software platform for bayesian evolutionary analysis. *PLoS Comp. Biol.* 10:e1003537. doi: 10.1371/journal.pcbi.1003537

Bouckaert, R. R. (2010). Densitree: making sense of sets of phylogenetic trees. *Bioinformatics* 26, 1372–1372. doi: 10.1093/bioinformatics/btq110

Brandley, M. C., Wang, Y., Guo, X., Nieto Montes de Oca, A., Feria-Ortíz, M., Hikida, T. et al. (2010). Bermuda as an evolutionary life raft for an ancient lineage of endangered lizards. *PLoS ONE* 5:e0011375. doi: 10.1371/journal.pone.0011375

Brennan, A. C., Barker, D., Hiscock, S. J., and Abbott, R. J. (2012). Molecular genetic and quantitative trait divergence associated with recent homoploid hybrid speciation: a study of *Senecio squalidus* (*Asteraceae*). *Heredity* 108, 87–95. doi: 10.1038/hdy.2011.46

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., and RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol. Ecol. Biol.* 29, 1917–1932. doi: 10.1093/molbev/mss086

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A. et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2215. doi: 10.1093%2Fbioinformatics%2Fbtr330

Drummond, A. J., Ho, S. Y., Phillips, M. J., and Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4:e0040088. doi: 10.1371/journal.pbio.0040088

Drummond, A. J., and Rambaut, A. (2007). BEAST: bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214. doi: 10.1186/1471-21 48-7-214

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075

Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Res.* 4, 359–361. doi: 10.1007/s12686-011-9548-7

Elgvin, T. O., Trier, C. N., Tørresen, O. K., Hagen, I. J., Lien, S., Nederbragt, A. J. et al. (2017). The genomic mosaicism of hybrid speciation. *Sci. Adv.* 3:e1602996. doi: 10.1126/sciadv.1602996

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S. et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:e19379. doi: 10.1371/journal.pone.0019379

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Fraley, C., and Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *J. Am. Statist. Assoc.* 97, 611–631. doi: 10.1198/016214502760047131

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q. et al. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS ONE* 9:e90346. doi: 10.1371/journal.pone.00 90346

Glenn, T. C., Staton, J. L., Vu, A. T., Davis, L. M., Bremer, J. R., Rhodes, W. E. et al. (2002). Low mitochondrial DNA variation among American alligators and a novel non-coding region in crocodilians. *J. Experim. Zool.* 294, 312–324. doi: 10.1002/jez.10206

Gompert, Z., and Buerkle, C. A. (2010). Introgress: a Software package for mapping components of isolation in hybrids. *Mol. Ecol. Res.* 10, 378–384. doi: 10.1111/j.1755-0998.2009.02733.x

Gompert, Z., and Buerkle, C. A. (2012). bgc: Software for Bayesian estimation of genomic clines. *Mol. Ecol. Res.* 12, 1168–1176. doi: 10.1111/1755-0998.12009.x

Gompert, Z., and Buerkle, C. A. (2016). What, if anything, are hybrids: enduring truths and challenges associated with population structure and gene flow. *Evol. Appl.* 9, 909–923. doi: 10.1111/eva.12380

Gompert, Z., Fordyce, J. A., Forister, M. L., Shapiro, A. M., and Nice, C. C. (2006). Homoploid hybrid speciation in an extreme habitat. *Science* 314, 1923–1925. doi: 10.1126/science.1135875

Gompert, Z., Mandeville, E. G., and Buerkle, C. A. (2017). Analysis of population genomic data from hybrid zones. *Annu. Rev. Ecol. Syst.* 48, 207–229. doi: 10.1146/annurev-ecolsys-110316-022652

Grant, P. R., and Grant, B. R. (2014). *40 Years of Evolution: Darwin's Finches on Daphne Major Island.* Princeton, NJ: Princeton University Press.

Green, R. E., Braun, E. L., Armstrong, J., Earl, D., Nguyen, N., Hickey, G. et al. (2014). Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science* 346:1254449. doi: 10.1126/science.1254449

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Hekkala, E. R., Platt, S. G., Thorbjarnarson, J. B., Rainwater, T. R., Tessler, M., Cunningham, S. W. et al. (2015). Integrating molecular, phenotypic and environmental data to elucidate patterns of crocodile hybridization in Belize. *R. Soc. Open Sci.* 2:150409. doi: 10.1098/rsos.150409

Hewitt, G. M. (2011). Quaternary phylogeography: the roots of hybrid zones. *Genetica* 139, 617–638. doi: 10.1007/s10709-011-9547-3

Huelsenbeck, J. P., and Ronquist, F. (2001). MrBayes: bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755. doi: 10.1093/bioinformatics/17.8.754

IUCN (2017). *The IUCN Red List of Threatened Species.* Version 2017-1. Available online at: http://www.iucnredlist.org

Jackiw, R. N., Mandil, G., and Hager, H. A. (2015). A framework to guide the conservation of species hybrids based on ethical and ecological considerations. *Conserv. Biol.* 29, 1040–1051. doi: 10.1111/cobi.12526

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

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S. et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., and Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Res.* 15, 1179–1191. doi: 10.1111/1755-0998.12387

Lamichhaney, S., Han, F., Webster, M. T., Andersson, L., Grant, B. R., and Grant, P. R. (2017). Rapid hybrid speciation in Darwin's finches. *Science* 359:224–228. doi: 10.1126/science.aao4593

Landry, C. R., Hartl, D. L., and Ranz, J. M. (2007). Genome clashes in hybrids: insights from gene expression. *Heredity* 99, 483–493. doi: 10.1038/sj.hdy.6801045

Leaché, A. D., Fujita, M. K., Minin, V. N., and Bouckaert, R. R. (2014). Species delimitation using genome-wide SNP data. *Syst. Biol.* 63, 534–542. doi: 10.1093/sysbio/syu018

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509

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

Machkour-M'Rabet, S., Hénaut, Y., Charruau, P., Gevery, M., Winterton, P., and Legal, L. (2009). Between introgression events and fragmentation, islands are the last refuge for the American crocodile in Caribbean Mexico. *Mar. Biol.* 156, 1321–1333. doi: 10.1007/s00227-009-1174-5

MacLeod, A., Rodríguez, A., Vences, M., Orozco-terWengel, P., García, C., Trillmich, F. et al. (2015). Hybridization masks speciation in the evolutionary history of the Galápagos marine iguana. *Proc. R. Soc. B* 282:20150425. doi: 10.1098/rspb.2015.0425

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends Ecol. Evol.* 20, 229–237. doi: 10.1016/j.tree.2005.02.010

Mallet, J. (2007). Hybrid speciation. *Nature* 446, 279–283. doi: 10.1038/nature05706

Meredith, R. W., Hekkala, E. R., Amato, G., and Gatesy, J. (2011). A phylogenetic hypothesis for *Crocodylus* (*Crocodylia*) based on mitochondrial DNA: evidence for a trans-Atlantic voyage from Africa to the New World. *Mol. Phyl. Evol.* 60, 183–191. doi: 10.1016/j.ympev.2011.03.026

Milián-García, Y., Ramos-Targarona, R., Pérez-Fleitas, E., Espinosa-López, G., and Russello, M. A. (2015). Genetic evidence of hybridization between the critically endangered Cuban crocodile and the American crocodile: implications for population history and in situ/ex situ conservation. *Heredity* 114, 272–280. doi: 10.1038/hdy.2014.96

Nei, M., Tajima, F., and Tateno, Y. (1983). Accuracy of estimated phylogenetic trees from molecular data II. Gene frequency data. *J. Mol. Evol.* 19, 153–170. doi: 10.1007/BF02300753

Nolte, A. W., and Tautz, D. (2010). Understanding the onset of hybrid speciation. *Trends Genet.* 26, 54–58. doi: 10.1016/j.tig.2009.12.001

Oaks, J. R. (2011). A time-calibrated species tree of crocodylia reveals a recent radiation of the true crocodiles. *Evolution* 65, 3285–3297. doi: 10.1111/j.1558-5646.2011.01373.x

Pacheco-Sierra, G., Gompert, Z., Domínguez-Laso, J., and Vázquez-Domínguez, E. (2016). Genetic and morphological evidence of a geographically widespread hybrid zone between two crocodile species, *Crocodylus acutus* and *Crocodylus moreletii. Mol. Ecol.* 25, 3484–3498. doi: 10.1111/mec.13694

Peakall, R., and Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in excel. Population genetic software for teaching and research–an update. *Bioinformatics* 28, 2537–2539. doi: 10.1093/bioinformatics/bts460

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25, 1253–1256. doi: 10.1093/molbev/msn083

Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26,1641–1650. doi: 10.1093/molbev/msp077

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

R Development Core Team (2016). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/

Raj, A., Stephens, M., and Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197, 573–589. doi: 10.1534/genetics.114.164350

Ray, D. A., and Densmore, L. (2002). The crocodilian mitochondrial control region: general structure, conserved sequences, and evolutionary implications. *J. Experim. Zool.* 294, 334–345. doi: 10.1002/jez.10198

Razo-Mendivil, U., Vázquez-Domínguez, E., and Pérez-Ponce de León, G. (2013). Discordant genetic diversity and geographic patterns between *Crassicutis cichlasomae* (*Digenea: Apocreadiidae*) and its cichlid host, "*Cichlasoma*" urophthalmus (*Osteichthyes: Cichlidae*), in Middle-America. *J. Parasit.* 99, 978–988. doi: 10.1645/13-225.1

Reardon, S. (2016). Cuban crocodiles pose a conservation conundrum. *Nature* 537, 596–597. doi: 10.1038/537596a

Roberts, H. F. (1919). A darwinian statement of the mendelian theory. *Nature* 103, 463–464. doi: 10.1038/103463b0

Rosauer, D. F., Blom, M. P. K., Bourke, G., Catalano, S., Donnellan, S., Gillespie, G. et al. (2016). Phylogeography, hotspots and conservation priorities: an example from the top end of Australia. *Biol. Conserv.* 204, 83–93. doi: 10.1016/j.biocon.2016.05.002

Rousset, F. (2017). *Genepop 4.6.9.* Available online at: http://genepop.curtin.edu.au/

Sánchez Herrera, O., Rivera-Téllez, E., López-Segurajáuregui, G., García Naranjo Ortiz de la Huerta, A., and Benítez Díaz, H. (2015). *Informe del Programa de Monitoreo del Cocodrilo de Pantano en México, Temporadas 2011 a 2013.* México: Comisión Nacional para el Conocimiento y Uso de la Biodiversidad.

Sardell, J. M., and Uy, J. A. (2016). Hybridization following recent secondary contact results in asymmetric genotypic and phenotypic introgression between island species of *Myzomela honeyeaters. Evolution* 70, 257–269. doi: 10.1111/evo.12864

Serrano-Gómez, S. S., Guevara-Chumacero, L. M., Barriga-Sosa, I. D. L. A., Ulloa-Arvizu, R., González-Guzmán, S., and Vázquez-Peláez, R. (2016). Low levels of

genetic diversity in *Crocodylus acutus* in *Oaxaca and Guerrero*, Mexico, and molecular-morphological evidence of the presence of *C. moreletii*. *Biochem. Syst. Ecol.* 69, 51–59. doi: 10.1016/j.bse.2016.08.005

Slatkin, M. (1991). Inbreeding coefficients and coalescence times. *Genet. Res.* 58, 167–175. doi: 10.1017/S0016672300029827

Stanley, R. R., Jeffery, N. W., Wringe, B. F., DiBacco, C., and Bradbury, I. R. (2017). genepopedit: a simple and flexible tool for manipulating multilocus molecular data in R. *Mol. Ecol. Res.* 17, 12–18. doi: 10.1111/1755-0998.12569

Thorbjarnarson, J. (1989). *Ecology of the American crocodile, Crocodylus acutus. Pp. 228-258 in Crocodiles. Their Ecology, Management and Conservation. A Special Publication of the Crocodile Specialist Group.* IUCN, Gland, Switzerland.

Vaz Pinto, P., Beja, P., Ferrand, N., and Godinho, R. (2016). Hybridization following population collapse in a critically endangered antelope. *Sci. Rep.* 6:18788. doi: 10.1038/srep18788

Vázquez-Domínguez, E., and Arita, H. T. (2010). The *Yucatan peninsula*: Biogeographical history 65 million years in the making. *Ecography* 33, 212–219. doi: 10.1111/j.1600-0587.2009.06293.x

vonHoldt, B. M., Brzeski, K. E., Wilcove, D. S., and Rutledge, L. Y. (2018). Redefining the role of admixture and genomics in species conservation. *Conserv. Lett.* 11, 1–6. doi: 10.1111/conl.12371

Weaver, J. P., Rodriguez, D., Venegas-Anaya, M., Cedeño-Vázquez, J. R., Forstner, M. R., and Densmore, L. D. (2008). Genetic characterization of captive *Cuban crocodiles* (*Crocodylus rhombifer*) and evidence of hybridization with the *American crocodile* (*Crocodylus acutus*). *J. Experim. Zool.* 309, 649–660. doi: 10.1002/jez.471

Wright, S. (1951). The genetical structure of populations. *Ann. Eugenet.* 15, 323–354. doi: 10.1111/j.1469-1809.1949.tb02451.x