SUPPORTING INFORMATION

SI Methods

DATA COLLECTION

I obtained tissue samples from 80 individuals representing all 23 crocodylian species (Table S2). I sequenced DNA from four regions of the mitochondrial genome: (1) *cytochrome b* (cytb) and portions of flanking tRNA genes for glutamic acid ($tRNA_{Glu}$) and threonine ($tRNA_{Thr}$), (2) nicotinamide adenine dinucleotide dehydrogenase subunit 2 (ND2) and portions of flanking tRNA genes for methionine ($tRNA_{Met}$) and tryptophan ($tRNA_{Trp}$), (3) nicotinamide adenine dinucleotide dehydrogenase subunit 3 (ND3) and portions of flanking tRNA genes for glycine ($tRNA_{Gly}$) and arginine ($tRNA_{Arg}$), and (4) the 5′ end of the control region (Dloop) and a portion of the adjacent tRNA gene for phenylalanine ($tRNA_{Phe}$). I also sequenced DNA from nine nuclear loci, including a portion of the entirely exonic oocyte maturation factor (*c-mos*), and eight exon primed, intron-crossing loci (EPIC; Palumbi and Baker, 1994; Palumbi, 1996): α-cardiac actin (ACTC) Exon 4–5, α-tropomyosin (aTROP) Exon 5–6, β-actin (ACTB) Exon 3–4, acetocholine receptor γ-subunit (AChR) Exon 7–8, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) Exon 11–12, lactate dehydrogenase b (LDH-B) Exon 6–7, lactate dehydrogenase a (LDH-A) Exon 7–8, and rhodopsin (RHO) Exon 2–3.

To minimize the risk of amplifying nuclear translocated copies of mitochondrial genes, I sequenced entire reading frames of protein-coding genes and portions of their flanking tRNA genes to allow any potential indicators of pseudogenes to be identified. For Dloop, I sequenced a portion of the adjacent $tRNA_{Phe}$ to help identification, and verified phylogenetic congruence with the protein-coding mitochondrial regions with analyses (see below). I took the following steps to minimize the chances of amplifying paralogs of the nuclear loci: (1) When appropriate comparative sequences were available from GenBank, I modified or designed primers to be long and highly specific to regions conserved across archosaurian orthologs, but variable across paralogs; (2) I selected or designed primers to amplify portions of the flanking exons long enough to aid

identification; and (3) I designed polymerase chain reaction (PCR) thermocycle programs to maximize fidelity (i.e. high annealing temperature or "touchdown" temperature methods).

I extracted DNA from tissues using guanidine thiocyanate salt extractions (Sambrook and Russell, 2001) or DNeasy kits (Qiagen, Valencia, CA). I amplified all loci via PCR in PTC-200 Peltier Thermal Cyclers (MJ Research, Waltham, MA), and purified the product using ExoSAP-IT (USB Corporation, Cleveland, OH) or polyethylene glycol (PEG) precipitation. I sequenced the cleaned PCR products using ABI Prism cycle sequencing chemistry (Applied Biosystems, Foster City, CA), and purified the products via filtration through G-50 fine Sephadex (GE Healthcare, Uppsala, Sweden) columns set in 96-well filter plates (Phenix Research Products, Hayward, CA) prior to vizualization on an ABI 3100 Genetic Analyzer.

All PCR amplifications were performed on total DNA in volumes of 25 μL, with 0.1 μL Taq DNA polymerase (New England BioLabs), 1 X ThermoPol Reaction Buffer (New England BioLabs), dNTPs (0.2 mM of each), 0.2 μM of each primer, and 1–2 μL (20–50 ng) of template. Unless otherwise stated, the following thermocycle protocol was used in all PCR amplifications: 1) 95°C for 2 min, 2) 45 cycles of 94°C for 0:45 min, the annealing temperature for 0:45 min, and 72°C for 1 min, and 3) ending with a 6 min extension at 72°C. In the following paragraph, the annealing temperature is given in parentheses following primer combinations that adhere to this thermocycle protocol. All PCR and cycle sequencing primers are summarized in Table S4.

For all individuals, I amplified cytb with two PCR reactions. Primer combinations for amplification of the 5′ end were L14198/H14653 (48°C) for alligatorids and L14174/H16543 (48°C) for all crocodylids except *Osteolaemus tetraspis*, for which L14086/H14638 (48°C) was used. The 3′ end of cytb was amplified using L14547/H15443 (52°C) for alligatorids and L14508/H15443 (52°C) for crocodylids. For all individuals, I used internal sequencing primers L14900 and H15046 for the 3′ end. I amplified the entire ND2 gene using the primer combination L3854/H4972 (56.6°C) for all individuals, except the three *Tomistoma schlegelii*, for which I used

L3856/H4972long (56.6°C). I used the following internal sequencing primers for ND2: L4234 (all individuals), L4451 (all alligatorids), L4453 (all crocodylids except *Mecistops cataphractus*—L4454cat), H4432 (all individuals except *M. cataphractus*—H4433cat and *Melanosuchus niger*—H4431melano), H4815 (all alligatorids), and H4758 (all crocodylids). I amplified the entire ND3 gene with primer combination L9453/H9884 (48°C) for all individuals. I amplified the 5′ end of Dloop using L15637/CR2H (57°C) for all individuals except *M. cataphractus*, for which I used L15637/H16258 (48°C). Primer combinations for nuclear loci were ACTCexon4F/ACTCexon5R (48°C) for ACTC, aTROPexon5F/aTROPexon6R (52°C) for aTROP, cmosF/cmosR (65.5°C) for *c-mos*, and GAPDHexon11F/GapdH950 (64°C) for GAPDH. For LDH-B, I used primer combination LDHBexon6F/LDHBexon7R (56.6°C) for all individuals except LSUMZ H-21741, LSUMZ H-21727, LSUMZ H-21729, LSUMZ H-21755, LSUMZ H-21756, LSUMZ H-21766, LSUMZ H-21768, LSUMZ H-21769, LSUMZ H-21771, LSUMZ H-21831, LSUMZ H-6420, LSUMZ H-6903, LSUMZ H-6976, LSUMZ H-6985, LSUMZ H-6990, LSUMZ H-6998, and LSUMZ H-7873 , for which I used LDHBexon6intF/LDHBexon7intR (48°C). I used primer combination LAI7_F1/LAI7_R1 to amplify LDH-A following the PCR thermocycle program described by (Gatesy et al., 2004). I used primer combinations ACTBexon3F/ACTBexon4R, AChRexon7F/AChRexon8R, and RHOexon2F/RHOexon3R to amplify ACTB, AChR, and RHO, respectively, under the following "touchdown" thermocycle conditions: 1) 95°C for 2 min, 2) 17 cycles of 94°C for 0:45 min, the annealing temperature for 0:45 min, and 72°C for 1 min, starting with an annealing temperature of 65°C and decreasing by 1°C per cycle, 3) 28 cycles of 94°C for 0:45 min, 48°C for 0:45 min, and 72°C for 1 min, and 4) ending with a 6 min 72°C extension.

I edited and aligned sequences using Sequencher 4.7 (Gene Codes Corporation, Ann Arbor, MI). I identified the reading frames of all protein-coding regions to confirm the absence of stop codons. For non-protein-coding loci that contained indels, I produced alignments with Sequencher 4.7, ClustalW (Thompson et al., 1994), and

T-Coffee 4.85 (Notredame et al., 2000), and used them to guide a manual alignment. I used complete mitochondrial genomes of six crocodylian species (Janke and Arnason, 1997; Janke et al., 2001, 2005) from GenBank to aid alignments and identification of gene borders for all mitochondrial sequences. For nuclear EPIC loci, I obtained homologous cDNA sequences of *Gallus gallus* (and crocodylians when available) from GenBank and aligned with the collected sequences. These alignments, along with the"GT–AG rule", were used to identify intron splice sites and determine the reading frame of the flanking exons.

ASSESSING AMONG-LOCUS CONGRUENCE

In order to justify concatenation of the loci for non-coalescent-based phylogenetic analyses, I assessed the degree of inter-locus phylogenetic congruence using partition homogeneity tests and independent tree inferences. I performed two partition homogeneity tests (Farris et al., 1994) on the full dataset in PAUP* 4.0b10 (Swofford, 2003); the first tested congruence among all 13 separate gene regions, and the second tested congruence between the mitochondrial and nuclear data. I ran separate RAxML (v7.0.0; Stamatakis, 2006) and MrBayes (version 3.1.2; Huelsenbeck and Ronquist, 2001) analyses on cytb, ND2, ND3, Dloop, the concatenated portions of the mitochondrial tRNA genes, and each of the nine nuclear loci (see below for detailed analysis settings). I partitioned Cytb, ND2, and ND3 by codon position, and ran the MrBayes analyses for 5.0×10^6 generations. Also, I ran RAxML and MrBayes analyses on all the mitochondrial data with four subsets (protein-coding genes partitioned by codon position, and non-protein-coding sites) and all the nuclear data with four subsets (introns, and exons partitioned by codon position). The MrBayes analyses were run for 1.0×10^7 generations each. After confirming congruence, I concatenated the data into one alignment.

PHYLOGENETIC ANALYSES

I analyzed the alignment under a Bayesian Dirichlet process prior model (DPP; Huelsenbeck and Suchard, 2007) to infer substitution rate variation across sites. For the

analysis, I (1) applied a GTR model to all sites (default setting); (2) used a gamma distribution with a mean of $2S - 3$ (where S is the number of sequences in the alignment) and variance λ (the rate parameter of the exponential probability distribution) as the prior for tree length; (3) used Dirichlet probability distributions as priors for the relative substitution rates, nucleotide frequencies, and proportion of total tree length allocated to each branch; and (4) assumed all phylogenetic trees are equally probable *a priori*. The prior probability distributions for the tree and branch lengths are derived from an exponential branch length prior with mean $1/\lambda$, where λ was set to 10 (i.e. a prior mean branch length of 0.1) (Huelsenbeck and Suchard, 2007). Additionally, I set the concentration parameter (χ) to 1.068, which creates a Dirichlet process prior with a mean number of subsets of $E(K) = 10$ according to the prior probability

$$f(K|\chi, N) = \frac{|S_1(N, K)|\chi^K}{\prod_{i=1}^{N}(\chi + i - 1)} \tag{1}$$

where $S_1(\cdot, \cdot)$ is the Stirling numbers of the first kind. I ran the analysis for 2.0×10^6 generations, with the DPP rate-variation model implemented for the second 1.0×10^6 generations. I began the analysis under a standard GTR model to allow the chain to find reasonable values for other parameters (e.g., topology and branch lengths) before the DPP was introduced into the MCMC proposal mechanism. I selected the partition from the posterior sample of the DPP analysis that minimized the distance to all other partitions in the posterior sample to use in all subsequent phylogenetic analyses. I used the Bayesian information criterion (BIC; Schwarz, 1978) to select the optimal nucleotide substitution model for each subset (ModelTest 3.7; Posada and Crandall, 1998).

For maximum-likelihood (ML) phylogenetic analyses I used RAxML and GARLI v0.96 (Zwickl, 2006). For all GARLI analyses: (1) Starting trees were generated using random-order stepwise addition with 100 attachments evaluated per taxon, (2) the BIC-selected model of nucleotide substitution was applied to each subset, (3) the overall substitution rate was allowed to vary among subsets using rate multipliers, (4) ten independent search replicates were performed, and (5) each replicate was

automatically terminated when there were no improvements in lnL due to topology change greater than 0.01 for more than 10,000 generations and no total improvement in lnL greater than 0.001 for more than 500 generations. To assess support, I ran 1000 bootstrap replicates using the same settings described above, except with one tree search per bootstrap replicate. For RAxML analyses, I ran 100 search replicates, applied GTR+Γ models of nucleotide substitution to all subsets, used random starting trees, and allowed the initial rearrangement setting to be determined automatically during the beginning of the search.

I estimated a rooted, ultrametric, time-calibrated phylogeny with BEAST v1.5.4 (Drummond and Rambaut, 2007), applying the BIC-selected models and separate lognormal relaxed-clocks to each subset (Drummond et al., 2006). I used a uniform prior ($U(0, 10)$) for the uncorrelated lognormal relaxed clock mean for each subset, with an initial value of 0.005 substitutions/site/my. This initial value was obtained by dividing the average divergence (substitutions/site) across the basal node of the crocodylian phylogeny by 157 my. The 157 my denominator is based on a 78.5 mya divergence between Alligatoridae and Crocodylidae (Brochu, 2004a,b); 0.005 is only an initial value and does not limit exploration of the parameter space set by the uniform prior. I used a normal distribution—$N(67.5, 3.188774)$—as a prior on the age of the node between Alligatorinae and Caimaninae that places 95% of the prior probability density between 71 and 64 mya. The Alligatorinae-Caimaninae split between 71 and 64 mya is considered among the best vertebrate fossil calibration points (Muller and Reisz, 2005). I placed an upper bound of 90 mya on the root of Crocodylia to conservatively extend the likely Campanian origin of Crocodylia (Brochu, 2003; Salisbury et al., 2006) by 6.5 my. Separate analyses were also done with the upper bound set to 100my to examine the effect of this prior on the age of *Crocodylus*. All other priors and MCMC operators were left at their default settings, and the MCMC operators were allowed to automatically optimize over the run. I ran four independent analyses for 5.0×10^7 generations, sampling every 20,000 generations.

SPECIES TREE ESTIMATION OF CROCODYLIA

To relax the assumption of congruence among the gene trees of each locus, I used the multi-species coalescent model of *BEAST (BEAST v1.5.4; Heled and Drummond, 2010) to estimate the species tree of Crocodylia. I used the same model of nucleotide substitution as in the standard BEAST analysis, and applied the same age constraints to nodes of the species tree. Gene trees were estimated independently (conditional on the species tree) for the 10 loci. Initial gene trees were randomly generated and the ploidy level of the autosomal and mitochondrial loci was set accordingly. I assigned a Yule process prior to the species tree, and constrained the effective population size along each branch to be constant. Six independent analyses were run for 2.0×10^8 generations, while sampling from the Markov chain every 1.0×10^5 generations.

ASSESSING MCMC STATIONARITY

I used several criteria to assess stationarity of the cold Markov chain for all Bayesian phylogenetic analyses: (1) All MCMC sample parameters were plotted versus generation time and visualized using Tracer (Rambaut and Drummond, 2005), (2) the cumulative and non-overlapping posterior probabilities of the 20 most variable nodes (the *cumulative* and *slide* commands, respectively) were plotted in Are We There Yet? (AWTY; Wilgenbusch et al., 2004), (3) node posterior probabilities were compared between independent runs using the compare command in AWTY, and (4) consensus trees from the independent runs were compared to ensure congruence. I assumed a run reached stationarity when all of these criteria yielded patterns congruent with stationarity, and discarded all samples of a run prior to this point.

TESTING A PRIORI PHYLOGENETIC HYPOTHESES

I tested the monophyly of (1) *Crocodylus* + *Mecistops*, (2) the Australasian species, (3) *C. niloticus*, (4) *C. siamensis* + *C. porosus*, and (5) *C. novaeguineae* using an approximately unbiased (AU) test (CONSEL; Shimodaira and Hasegawa, 2001)). I ran three ML heuristic searches in RAxML (100 search replicates each) using the same settings as the aforementioned unconstrained searches, except the topology was

constrained to be congruent with each hypothesis. I optimized the site-wise lnL scores on the set of unique ML trees found from the unconstrained and constrained searches using a modified version of RAxML (provided by Alexandros Stamatakis); RAxML was modified to allow site-wise likelihoods to be optimized on sets of topologies under partitioned phylogenetic models. The per-site likelihoods from all unique topologies were compared in CONSEL (Shimodaira and Hasegawa, 2001) using the approximately unbiased (AU) test (Shimodaira, 2002; Shimodaira and Hasegawa, 1999) with 100,000 bootstrap replicates. The P-value I report for a hypothesis is the largest P-value of all the unique topologies inferred under that constraint.

BIOGEOGRAPHIC ANALYSES OF CROCODYLUS

To obtain trees for ancestral-area reconstructions, I pruned the species trees from the posterior sample of the *BEAST analysis to contain only Crocodylinae. I used the majority rule consensus tree with mean branch lengths for all ML character-state reconstructions; I used the whole posterior sample of trees for all Bayesian reconstructions.

BayesTraits (Pagel, 1999; Pagel and Meade, 2007) uses a continuous-time Markov model of discrete character evolution with an arbitrary number of parameters for the instantaneous rates of transition between character-states (geographic areas in this case). The probability of the data (phylogenetic tree with branch lengths and geographic area states at the tips) is maximized by integrating over the conditional likelihoods of all possible states at all the internal nodes. The relative contribution to the likelihood of a given state at a node is its marginal probability. For the ML reconstructions in BayesTraits, I ran each analysis three times (1000 replicates each) to ensure consistent results. I selected the best-fit model of character evolution using the Akaike information criterion (AIC; Akaike, 1974) and a series of nested likelihood ratio tests (LRT). For the LRTs, I followed the following steps. (1) I ran an analysis with the fully-parameterized model (six transition rates), (2) I set the transition rates with the most similar estimates from the previous run to be equal. If there were multiple rates with identical values (e.g. 0), all of these rates were set to be equal. (3) I ran the next

analysis invoking these new constraints, and compared the resulting lnL score to that of the previous, less-constrained model using a LRT test statistic of:

$$\delta_{ij} = 2(\ln L_i - \ln L_j) \tag{2}$$

and assuming a χ^2 null distribution with the degrees of freedom equal to the difference in the number of free parameters between the models. (3) I repeated steps 2 & 3 until either the new model was rejected by the LRT, or there was only one transition rate parameter (i.e., all rates were set equal).

For Bayesian reconstructions, I used the optimal model of character evolution as determined via ML above, and used Exponential hyper-parameters for the instantaneous rates of transition among states. I placed uniform priors on the mean of the exponential distributions ($1/\lambda$) with lower bounds of zero and upper bounds of twice the rate inferred from the ML analyses. Furthermore, to accommodate uncertainty in model selection, I used a reversible jump MCMC model with exponential hyper-parameters for all transition rates with a uniform prior—$U(0, 0.1)$—on $1/\lambda$. For all Bayesian reconstructions, I adjusted the ratedev parameter so that the acceptance rates of proposed changes was 20%. I ran analyses for 1.0×10^8 generations, sampled every 25,000 generations, and discarded samples from the first 5.0×10^7 generations as burn-in. All analyses were run three times independently to ensure consistency, and post burn-in stationarity was confirmed by plotting all model parameters over generations using Tracer.

Lagrange (v2.0.1) (Ree and Smith, 2008) uses a continuous-time Markov model of range evolution with two instantaneous rate parameters (dispersal among areas and local extinction). The probability of the data (phylogeny with branch lengths and range data at the tips) is maximized under this model by integrating over the conditional likelihoods of all possible range inheritance scenarios at all internal nodes. The primary difference between this model and that of BayesTraits is that it integrates over range inheritance scenarios rather than character-states at each node (Ree et al., 2005; Ree and Smith, 2008). Another difference in the current implementation of Lagrange is that

it does not allow multiple free dispersal parameters, so I used a model with one dispersal parameter.

To test the out-of-Africa hypothesis, I ran three additional analyses with the basal node of *Crocodylus* constrained to each of the three possible character states; this was done under ML in Bayestraits and Lagrange and under a Bayesian framework in BayesTraits. Because these models are not nested, a decrease of 2 $\ln L$ units or more in comparison with the unconstrained model was considered strong support against a basal character-state (Pagel, 1999) for the ML analyses. For the Bayesian results, I used approximate Bayes factors (BF) (Jeffreys, 1935; Newton and Raftery, 1994) to assess the support for the basal state. The log-transformed BF was calculated as:

$$2\ln(BF_{ij}) = 2\big(\ln(HL_i) - \ln(HL_j)\big) \tag{3}$$

where HL_i and HL_j is the harmonic mean likelihood for models i (unconstrained) and j (basal node constrained to an area), respectively. Following (Kass and Raftery, 1995), I consider $2\ln(BF_{ij})$ greater than six as strong support against hypothesis j.

SI Results

DATA COLLECTION

In the following discussion of the collected sequence data, all numbers referring to codon positions correspond to the 3$^{\text{rd}}$ nucleotide. For cytb, 1197 bp (corresponding to bases 14,283–15,479 of the *Crocodylus niloticus* mitochondrial genome; GenBank accession no. AJ810452) of open reading frame were collected and aligned across all crocodylids. *Paleosuchus palpebrosus* has a stop codon at Site 1137; beyond this point, alligatorids could not be aligned with the crocodylids. Thus the last 60 bp of cytb were coded as missing data for alligatorids. The cytb alignment has no insertions or deletions (indels). For ND2, 1056 bp (corresponding to bases 3898–4953 of the *C. niloticus* mitochondrial genome) of open reading frame were sequenced and aligned across crocodylids. *Paleosuchus* has a stop codon at site 1044, beyond which alligatorids could

not be aligned with crocodylids. Thus, the last 12 bp were coded as missing data for alligatorids. The ND2 alignment has no indels. For ND3, 348 bp (corresponding to bases 9488–9835 of the *C. niloticus* mitochondrial genome) of open reading frame were sequenced and aligned for all individuals with one anomaly; all three individuals of *Melanosuchus niger* have an insertion of a cytosine at the 87^{th} position of the reading frame, causing a frameshift and premature stop codons at positions 90, 96, 249, 309, 336, and 345. This does not seem to be an artifact resulting from the amplification of a nuclear pseudogene, because the sequences of each of the individuals possess no other anomalies diagnostic of nuclear translocated copies (e.g., no heterozygous sites, no other indels, no stop codons if the 87th base is removed, and both flanking tRNAs are identifiable to other alligatorids). Additionally, sequences of the three individuals are identical, except for a single synonymous substitution at a 3rd codon position. When these three *Melanosuchus* are aligned with all individuals of the three species of *Caiman* (the sister clade of *Melanosuchus*), 7, 2, and 41 of the substitutions that differentiate *Melanosuchus* from any of the *Caiman* species occur at the 1^{st}, 2^{nd}, and 3^{rd} codon positions, respectively. Furthermore, of these 50 substitutions, only eight of them are nonsynonymous, yielding a K_A:K_S ratio of 0.19. These numbers are consistent with a protein-coding gene under purifying selection, and not a nuclear translocated pseudogene. The hypothesis that this insertion represents a real frameshift mutation warrants further investigation, but for the purposes of this study, the cytosine at the 87^{th} position in these three individuals was removed and the remaining alignment used in subsequent analyses. Given the possibility that the ND3 sequences for *M. niger* are nuclear pseudogenes, separate phylogenetic analyses were performed with ND3 coded as missing data for these three individuals; the inferred position of *Melanosuchus* did not change, thereby ensuring the placement of this genus is not driven by this locus.

Five hundred and forty-four bp of Dloop and 20 bp of the adjacent $tRNA_{Phe}$ (a total of 564 bp, corresponding to bases 15707–16268 of the *C. niloticus* mitochondrial genome) were sequenced and aligned only for Crocodylinae; remaining individuals were coded as missing data. Fifty-nine bp of $tRNA_{Glu}$, 24 bp of $tRNA_{Met}$, 20 bp of $tRNA_{Trp}$,

28 bp of $tRNA_{Gly}$, and 39 bp $tRNA_{Arg}$ were aligned for all crocodylids, with alligatorids coded as missing data. Dloop and the tRNAs contained indels, but all were easily aligned.

The following nuclear sequence data were collected and aligned for all individuals: **ACTC**: 8 bp of Exon 4, 120 bp of Intron 4, and 56 bp of Exon 5; **aTROP**: 60 bp of Exon 5, 168 bp of Intron 5, and 79 bp of Exon 6; **AChR**: 74 bp of Exon 7, 412 bp of Intron 7, and 36 bp of Exon 8; ***c-mos***: 579 bp; **GAPDH**: 33 bp of Exon 11, 408 bp of Intron 11, and 19 bp of Exon 12; **LDH-A**: 35 bp of Exon 7, 550 bp of Intron 7, and 122 bp of Exon 8; **LDH-B**: 47 bp of Exon 6, 552 bp of Intron 6, and 26 bp of Exon 7; **RHO**: 91 bp of Exon 2, 132 bp of Intron 2, and 40 bp of Exon 3. For LDH-B, some individuals lack the first 25 bp, whereas two individuals, LSUMZ H-21755 and LSUMZ H-21756, lack the first 46 and last 26 bp, respectively; these regions were coded as missing data. For ACTB, 32 bp of Exon 3, 134 bp of Intron 3, and 134 bp of Exon 4 were obtained and aligned for all crocodylids and five alligatorids (LSUMZ H-18733—*Alligator mississippiensis*, LSUMZ H-7868—*A. sinensis*, LSUMZ H-21699—*A. sinensis*, LSUMZ H-21700—*A. sinensis*, and LSUMZ H-6997—*P. palpebrosus*); remaining individuals were coded as missing data. Most intron alignments possessed some indels, but all were easily aligned. All nuclear exons were easily aligned across all individuals, with no indels except one three bp deletion of a codon for methionine at the 510^{th} position in *c-mos* for all *Crocodylus*.

Sequence data were obtained for LSUMZ H-6998 (*P. palpebrosus*) for all loci except ACTB, for which it would not amplify. However, sequence data were gathered for ACTB from a conspecific individual (LSUMZ H-6997). Thus, the ACTB sequence of H-6997 was concatenated with the rest of the sequence data of H-6998, to form a chimerical sequence that was used in all subsequent analyses. To ensure this action was justified, I compared loci for which both individuals had sequence data (AChR, aTROP, *c-mos*, and RHO; 1608 bp). Across 1608 bp of nuclear data, these two individuals share an identical sequence that is unique from all other individuals in the dataset.

The resulting dataset is an alignment of 3335 bp of mitochondrial data and 3947

bp of nuclear data, for a total of 7282 bp for 79 individuals (including indels and missing data). All indels were treated as missing data in analyses.

AMONG-LOCUS CONGRUENCE

There is consistent disagreement between the mitochondrial and nuclear data regarding the placement of one individual of *C. moreletii* (H-21729) and two individuals of *C. acutus* (H-21711 and H-21713). The mitochondrial data place the two *C. acutus* with *C. rhombifer*, and nest the *C. moreletii* within the remaining *C. acutus*, both with strong support. The nuclear data place these three individuals within their respective conspecific clades with strong support. This pattern suggests these three individuals are hybrids or possess introgressed mitochondria. Ancestral polymorphisms are another potential explanation, but seem much less likely because the nuclear data place the individuals within their respective species, whereas the mitochondrial data do not. Because the mitochondrial genome is effectively haploid and uniparentally inherited, its effective population size is approximately one quarter of that of nuclear loci; thus, according to coalescent theory should complete lineage sorting approximately four times faster (assuming neutrality and constant population size) following reproductive isolation (Birky et al., 1989; Palumbi et al., 2001). Thus, if incomplete lineage sorting caused incongruence, the ancestral polymorphisms should appear in the nuclear data, but not the mitochondrial data; however, the converse situation is seen in the data. A selective sweep could cause rapid lineage sorting in a nuclear locus; however, support for the placement of these putative hybrids into their conspecific clades comes from site patterns across multiple, independent nuclear loci. Thus, introgression is the more likely explanation in this case.

If these three individuals are introgressed, the nature of the hybridization is unknown. All three tissue samples in question came from captive animals and lack vouchers and locality information (Table S2). Thus, hybridization might have occurred in captivity (hybridization is rampant in farms; Fitzsimmons et al., 2002). Because of the ambiguity associated with these three individuals (H-21729, H-21711, and H-21713), they were excluded from all subsequent analyses, after which the partition homogeneity

tests were non-significant, and there was no significant incongruence among genes or between genomes.

CONVERGENCE OF BEAST RUNS

For the BEAST analyses, all four independent Markov chains rapidly converged to stationarity in the same region of parameter space. To be conservative, I removed the first 1000 of the 2500 samples as burn-in from each chain, and the remaining 1500 posterior samples were combined across the four independent chains. The effective sample sizes (ESS) for all model parameters from all post-burnin chains were greater than 584 and 1387 for the 90 and 100 my maximum root age analysis, respectively.

For the *BEAST analysis with an upper limit on the root age of 90 my, four of the six independent chains rapidly converged to stationarity in the same parameter space, whereas the other two chains did not converge and sampled from unrealistic regions of parameter space for branch lengths and population sizes throughout the run. The *BEAST analysis with an upper root age of 100 my had five of the six chains converge rapidly and one that sampled unrealistic branch lengths and population sizes throughout. I excluded the chains that failed to converge from both *BEAST analyses. I conservatively discarded the first 1000 of the 2000 samples from each converged chain as burn-in, and combined the remaining posterior samples across the chains. The effective sample sizes (ESS) for all model parameters from all post-burnin chains were greater than 153 and 288 for the 90 and 100 my maximum root age analysis, respectively; most parameters had ESS values of greater than 500. The parameters with the smallest ESS values were a few of the split population size parameters.

PRECISION OF REPLICATE BIOGEOGRAPHIC ANALYSES

The inferred parameter values and state probabilities are identical among all independent, ML replicate analyses performed in BayesTraits, and among replicate runs of Lagrange. Posterior mean values of parameters and state probabilities are identical to at least two decimal places for Bayesian replicate analyses in BayesTraits.

References

Akaike, H., 1974. A new look at the statistical model identification. Ieee T Automat Contr 19:716–723.

Birky, C. W., P. Fuerst, and T. Maruyama, 1989. Organelle gene diversity under migration, mutation and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. Genetics 121:613–627.

Brochu, C. A., 2003. Phylogenetic approaches toward crocodylian history. Annu Rev Earth Pl Sc 31:357–397.

———, 2004a. Calibration age and quartet divergence date estimation. Evolution 58:1375–1382.

———, 2004b. Patterns of calibration age sensitivity with quartet dating methods. J Paleontol 78:7–30.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut, 2006. Relaxed phylogenetics and dating with confidence. Plos Biol 4:e88.

Drummond, A. J. and A. Rambaut, 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214.

Farris, J. S., M. Källersjö, A. G. Kluge, and C. Bult, 1994. Testing significance of incongruence. Cladistics 10:315–319.

Fitzsimmons, N. N., J. C. Buchan, P. V. Lam, G. Polet, T. T. Hung, N. Q. Thang, and J. Gratten, 2002. Identification of purebred *Crocodylus siamensis* for reintroduction in Vietnam. J Exp Zool 294:373–381.

Friesen, V., B. Congdon, M. Kidd, and T. Birt, 1999. Polymerase chain reaction (PCR) primers for the amplification of five nuclear introns in vertebrates. Mol Ecol 8:2147–2149.

Friesen, V., B. Congdon, H. Walsh, and T. Birt, 1997. Intron variation in marbled murrelets detected using analyses of single-stranded conformational polymorphisms. Mol Ecol 6:1047–1058.

Gatesy, J., R. H. Baker, and C. Hayashi, 2004. Inconsistencies in arguments for the supertree approach: supermatrices versus supertrees of Crocodylia. Syst Biol 53:342–355.

Gratten, J., 2003. The molecular systematics phylogeography and population genetics of Indo-Pacific *Crocodylus*. Ph.D. thesis, University of Queensland, Brisbane St. Lucia, Queensland.

Heled, J. and A. J. Drummond, 2010. Bayesian inference of species trees from multilocus data. Mol Biol Evol 27:570–580.

Huelsenbeck, J. P. and F. Ronquist, 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.

Huelsenbeck, J. P. and M. A. Suchard, 2007. A nonparametric method for accomodating and testing across-site rate variation. Syst Biol 56:975–987.

Janke, A. and U. Arnason, 1997. The complete mitochondrial genome of *Alligator mississippiensis* and the separation between recent Archosauria (birds and crocodiles). Mol Biol Evol 14:1266–1272.

Janke, A., D. Erpenbeck, M. Nilsson, and U. Arnason, 2001. The mitochondrial genomes of the iguana (*Iguana iguana*) and the caiman (*Caiman crocodylus*): Implications for amniote phylogeny. P Roy Soc Lond B Bio 268:623–631.

Janke, A., A. Gullberg, S. Hughes, R. K. Aggarwal, and U. Arnason, 2005. Mitogenomic analyses place the gharial (*Gavialis gangeticus*) on the crocodile tree and provide pre-K/T divergence times for most crocodilians. J Mol Evol 61:620–626.

Jeffreys, H., 1935. Some tests of significance, treated by the theory of probability. Proc Camb Philol Soc 31:203–222.

Kass, R. E. and A. E. Raftery, 1995. Bayes factors. J Am Stat Assoc 90:773–795.

Muller, J. and R. R. Reisz, 2005. Four well-constrained calibration points from the vertebrate fossil record for molecular clock estimates. Bioessays 27:1069–1075.

Newton, M. A. and A. E. Raftery, 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. J Roy Stat Soc B 56:3–48.

Notredame, C., D. G. Higgins, and J. Heringa, 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 301:205–217.

Pagel, M., 1999. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. Syst Biol 48:612–622.

Pagel, M. and A. Meade, 2007. BayesTraits: Available for download at www.evolution.rdg.ac.uk.

Palumbi, S. R., 1996. Nucleic acids ii: The polymerase chain reaction. Pp. 205–247, *in* D. M. Hillis, C. Moritz, and B. K. Mable, eds. Molecular Systematics, second ed. Sinauer Associates, Inc., Sunderland, Massachusetts, U.S.A.

Palumbi, S. R. and C. S. Baker, 1994. Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. Mol Biol Evol 11:426–435.

Palumbi, S. R., F. Cipriano, and M. P. Hare, 2001. Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. Evolution 55:859–868.

Posada, D. and K. A. Crandall, 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818.

Rambaut, A. and A. Drummond, 2005. Tracer v1.3, available from www.evolve.zoo.ox.ac.uk/software.

Ray, D. A. and L. Densmore, 2002. The crocodilian mitochondrial control region: General structure, conserved sequences, and evolutionary implications. J Exp Zool 294:334–345.

Ree, R. H., B. R. Moore, C. O. Webb, and M. J. Donoghue, 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Evolution 59:2299–2311.

Ree, R. H. and S. A. Smith, 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst Biol 57:4–14.

Salisbury, S. W., R. E. Molnar, E. Frey, and P. M. A. Willis, 2006. The origin of modern crocodyliforms: new evidence from the Cretaceous of Australia. P Roy Soc Lond B Bio 273:2439–2448.

Sambrook, J. and D. W. Russell, 2001. Molecular Cloning: A Laboratory Manual, vol. 1. 3rd ed. Cold Spring Harbor, New York.

Schwarz, G., 1978. Estimating the dimension of a model. Ann Stat 6:461–464.

Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst Biol 51:492–508.

Shimodaira, H. and M. Hasegawa, 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 16:1114–1116.

———, 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17:1246–1247.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Swofford, D. L., 2003. PAUP*, phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, MA.

Thompson, J. D., D. G. Higgins, and T. J. Gibson, 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680.

Waltari, E. and S. V. Edwards, 2002. Evolutionary dynamics of intron size, genome size, and physiological correlates in archosaurs. Am Nat 160:539–552.

Wilgenbusch, J. C., D. L. Warren, and D. L. Swofford, 2004. AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. Http://ceb.csit.fsu.edu/awty .

Zwickl, D. J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. thesis, University of Texas at Austin, Austin, Texas, USA.

Table S1. The taxonomy used throughout this work.

- Class Reptilia
  - Order Crocodylia
    - Family Alligatoridae
      - Subfamily Alligatorinae—the alligators
        - Genus *Alligator*
          - *A. mississippiensis*—American alligator
          - *A. sinensis*—Chinese alligator
      - Subfamily Caimaninae—the caimans
        - Genus *Caiman*—*the true caimans*
          - *C. crocodilus*—spectacled or common caiman
          - *C. yacare*—Yacaré caiman
          - *C. latirostris*—broad-snouted caiman
        - Genus *Melanosuchus*—the black caimans
          - *M. niger*—black caiman
        - Genus *Paleosuchus*—the dwar‘f caimans
          - *P. palpebrosus*—Cuvier's dwarf, or dwarf, caiman
          - *P. trigonatus*—Schneider's dwarf, or smooth-fronted, caiman
    - Family Crocodylidae
      - Subfamily Crocodylinae—the crocodiles
        - Genus *Crocodylus*—the true crocodiles
          - *C. acutus*—American crocodile
          - *C. intermedius*—Orinoco crocodile
          - *C. rhombifer*—Cuban crocodile
          - *C. moreletii*—Morelet's crocodile
          - *C. niloticus*—Nile crocodile
          - *C. siamensis*—Siamese crocodile
          - *C. palustris*—mugger crocodile
          - *C. porosus*—estuarine or saltwater crocodile
          - *C. mindorensis*—Philippine crocodile
          - *C. novaeguineae*—New Guinea crocodile
          - *C. johnstoni*—Johnston's, or Australian freshwater, crocodile
        - Genus *Mecistops*—the African slender-snouted crocodiles
          - *M. cataphractus*—African slender-snouted crocodile
        - Genus *Osteolaemus*—the African dwarf crocodiles
          - *O. tetraspis*—African dwarf crocodile
      - Subfamily Gavialinae—the gharials
        - Genus *Gavialis*—the true gharials
          - *G. gangeticus*—true or Indian gharial
        - Genus *Tomistoma*—the false gharials
          - *T. schlegelii*—false gharial

Table S2. List of all tissue samples used in this study.

Species/LSUMZ#	Locality	Collector
Alligator mississippiensis		
H-18733	USA: Louisiana; Rockefeller NWR	T. Bryan
Alligator sinensis		
H-7868	Captive	M. Brown
H-21699–700	Captive	K. Vliet
Caiman crocodilus		
H-13961–2, 13964	Brazil: Amazonas; Rio Ituxi at the Madeirera Scheffer	L. Vitt
H-21701–2	Captive	K. Vliet
Caiman latirostris		
H-21705–6	Captive	K. Vliet
Caiman yacare		
H-21707	Captive	K. Vliet
Melanosuchus niger		
H-21751–3	Captive	K. Vliet
Paleosuchus palpebrosus		
H-6997–8	Captive	L. Densmore
H-21761	Captive	K. Vliet
Paleosuchus trigonatus		
H-6420	Captive	J. Behler
H-7873	Captive	H. Dessauer
Gavialis gangeticus		
H-21748	Captive	K. Vliet
Tomistoma schlegelii		
H-21763–5	Captive	K. Vliet
Crocodylus acutus		
H-6760	Captive	H. Dessauer
H-6982	Captive	L. Densmore
H-21708–15	Captive	K. Vliet
Mecistops cataphractus		
H-6976	Captive	L. Densmore
H-21718–20	Captive	K. Vliet
Crocodylus intermedius		
H-20683–6*	Captive	J. Boundy
H-21724	Captive	K. Vliet
Crocodylus johnstoni		
H-7070	Captive	L. Densmore
H-21725–6	Captive	K. Vliet
Crocodylus moreletii		
H-6903	Mexico	H. Dessauer
H-21727, 21729–30	Captive	K. Vliet
Crocodylus niloticus		
H-21731, 21733–9	Captive	K.Vliet
Crocodylus mindorensis		
H-21766	Philippines: Maridagao Carmen North Cotabato	F. Pontillas
H-21768	Philippines: Dalican Dinaig Maguindanao	F. Pontillas
H-21769, 21831	Philippines: Davao	F. Pontillas
H-21771	Philippines: Rio Grande Cotabato Maguindanao	F. Pontillas
H-21815	Philippines: Busuanga	F. Pontillas
H-21872	Philippines: Zamboanga	F. Pontillas
Crocodylus novaeguineae		
H-6995, 7071	Captive	L. Densmore
Crocodylus palustris		
H-21741–2	Captive	K. Vliet
Crocodylus porosus		
H-6758	Solomon Islands: Guadalcanal Province; Guadalcanal	H. Dessauer
H-6984	Captive	L. Densmore
Crocodylus rhombifer		
H-21745–7	Captive	K. Vliet
Crocodylus siamensis		
H-6978, 6985	Captive	L. Densmore
Osteolaemus tetraspis		
H-21755–6	Captive	K. Vliet
H-6990, 6992	Captive	L. Densmore

* Vouchered specimens.

Table S3. Biogeographic model selection in BayesTraits

Model parameters						**ln L**	**AIC**	**δ_{ij} (P-val)**
$q_{A\to N}$	$q_{A\to U}$	$q_{N\to A}$	$q_{N\to U}$	$q_{U\to A}$	$q_{U\to N}$			
0	1	2	3	4	5	-9.08	30.17	NA
0	1	2	0	0	3	-9.08	26.17	0.00 (1.00)
0	1	2	0	0	1	-9.14	24.29	0.12 (0.73)
0	0	1	0	0	0	-9.81	23.62	1.34 (0.25)
0	0	0	0	0	0	-11.37	24.73	3.11 (0.08)

Character-state abbreviations are A = Africa, N = Neotropics, and U = Australasia. The first six columns indicate which transition rates are free parameters. $\ln L$ is the maximum log likelihood score for the model, AIC is the Akaike information criterion score, and the last column is the likelihood ratio test statistic, δ_{ij} (Equation S2), where i is the model in the previous row and j is the current row, followed by the associated P-value from the χ^2 distribution in parantheses.

Table S4. Primers used in PCR and cycle sequence reactions. The numbers used in all mitochondrial primer names refer to the position of the 3′ base in the *Alligator mississippiensis* mitochondrial genome (Janke and Arnason, 1997). [1]This work; [2](Ray and Densmore, 2002); [3](Gratten, 2003); [4](Waltari and Edwards, 2002); [5](Friesen et al., 1999); [6](Friesen et al., 1997); [6*]modified from (Friesen et al., 1999); [7](Gatesy et al., 2004).

Locus	**Location**	**Primer**	**Sequence (5′ ⇒ 3′)**
Cytb	ND6	L14086[1]	GCA AAR AGC ARA CTW AYY ACC CCA TA
	$tRNA_{Glu}$	L14174[1]	AAW GYM ATT YCC ATT ATT YTC ACT TGG
	$tRNA_{Glu}$	L14198[1]	TTC AAC CAA AAC CTG AGG YCT G
	cytb	L14508[1]	GCA AAC GGA GCY TCY CTA TTC TTC
	cytb	L14547[1]	ATC GGA CGA GGC CTA TAC TAC
	cytb	L14900[1]	CYG ACA AAR TYC CRT TYC ACC C
	cytb	H14638[1]	CCC TCA GAA TGA TAT TTG TCC TCA
	cytb	H14653[1]	GTR ATY ACG GTT GCC CCT CAG AA
	cytb	H15046[1]	TAG GCR AAT AGG AAR TAT CAT TC
	$tRNA_{Thr}$	H15443[1]	YTC TGT CTT ACA AGG CCA GYG CTT
ND2	$tRNA_{Met}$	L3854[1]	AAA RCT ATT GGG CCC ATA CCC C
	$tRNA_{Met}$	L3856[1]	AAR CTW TTG GGY CCA TRC CCC AA
	ND2	L4234[1]	CCA TTY CAC TTC TGA GTR CCA G
	ND2	L4451[1]	TCC ATY GCC CAA ATR GCA TG
	ND2	L4453[1]	TCV ATT GCC CAA ATA GCH TGA A
	ND2	L4454cat[1]	TCA ATC GCT CAG ATA GCT TGA AC
	ND2	L4454siam[1]	TCA ATT GCC CAA ATA TCT TGA AC
	ND2	H4431melano[1]	TTC ATG CTA TTT GGG CGA CTG AG
	ND2	H4432[1]	TTC ADG CTA TTT GGG CAA TBG A
	ND2	H4433cat[1]	GTT CAA GCT ATC TGA GCG ATT G
	ND2	H4758[1]	GAG TTG TAT CAT AGT CGD AGG TAR AAG
	ND2	H4815[1]	TTT TCG TCA RAG GCG GGT TRT G
	$tRNA_{Trp}$	H4972[1]	GGC TTT GAA GGC CCT CGG YTT
	$tRNA_{Trp}$	H4972long[1]	TAG GGC TTT GAA GGC CCT YGG CTT
ND3	$tRNA_{Gly}$	L9453[1]	CAA RTG ACT TCC AAT CAY TAR ACC C
	$tRNA_{Arg}$	H9884[1]	TCR TGA TTT TCT ARG YCG AAR YTA G
Dloop	$tRNA_{Phe}$	L15637[1]	GCA TAA CAC TGA AAA TGT TAA YAT GG
	Dloop	$CR2H^2$ (16179)	GGG GCC ACT AAA AAC TGG GGG
	Dloop	H16258[3]	CTA AAA TTA CAG AAA AGC CGA CCC
ACTC	Exon 4	$ACTCexon4F^4$	GAG CGT GGC TAY TCC TTT GT
	Exon 5	$ACTCexon5R^4$	GTG GCC ATT TCA TTC TCA AA
aTROP	Exon 5	$aTROPexon5F^5$	GAG TTG GAT CGG GCT CAG GAG CG
	Exon 6	$aTROPexon6R^5$	CGG TCA GCC TCT TCA GCA ATG TGC TT
ACTB	Exon 3	ACTBexon3F[1]	CAT CGG CAA TGA GCG GTT CAG GTG
	Exon 4	ACTBexon4R[1]	GCC AGG GCT GTG ATT TCC TTC TGC AT
AChR	Exon 7	$AChRexon7F^4$	CGC AAG CCG CTC TTC TA
	Exon 8	$AChRexon8R^4$	GAC AGT CTG GGC CAG GA
GAPDH	Exon 11	$GAPDHexon11F^{6*}$	ACC TTT GAT GCG GGT GCT GGC ATT GC
	Exon 12	GapdH950[6]	CAT CAA GTC CAC AAC ACG GTT GCT GTA
LDH-A	Exon 7	$LAI7\_F1^7$	TGG CTG AAA CTG TTA TGA AGA ACC
	Exon 8	$LAI7\_R1^7$	TGG ATT CCC CAA AGT GTA TCT G
LDH-B	Exon 6	LDHBexon6F[1]	GGA GTT GAA TCC TGC TAT GGG TAC TGA C
	Exon 6	LDHBexon6intF[1]	GAG AAM TGG AAA GAA GTC CAC AAG
	Exon 7	LDHBexon7R[1]	GGT CTC AAG TAG ATC AGC AAC ACT AAR G
	Exon 7	LDHBexon7intR[1]	CCA ATG GCC CAG TTA GTG TAT C
RHO	Exon 2	RHOexon2F[1]	GTG GTC TGC AAG CCC ATG AGC AAT TTC C
	Exon 3	RHOexon3R[1]	CRT TGT TGA CCT CAG GCT TCA GNG TGT AGT A
c-mos	Internal	cmosF[1]	AYT GGG ATC AAG TGT GCC TAC TG
	Internal	cmosR[1]	AGT AGA TGT CTG CTT TGG GGG TGA C

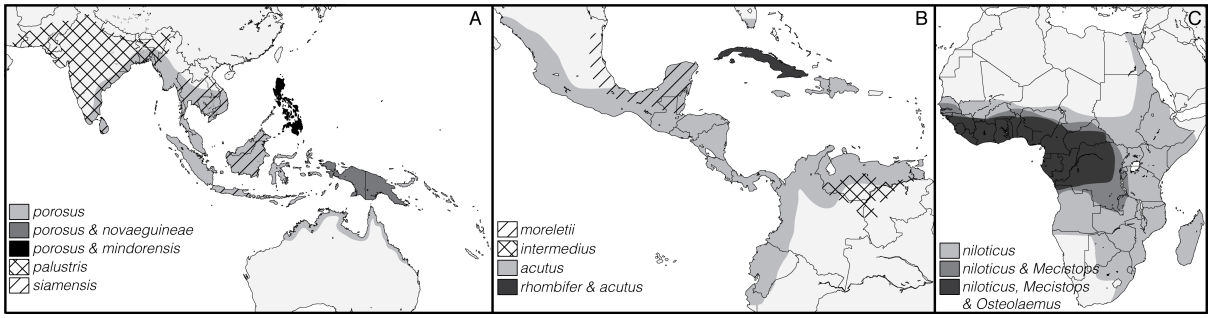

Figure S1. Approximate geographic distributions of all Crocodylinae (*Crocodylus* with specific epithet only) separated into A) Australasian, B) Neotropical, and C) African.