Received: 17 October 2017 | Accepted: 3 December 2018

DOI: 10.1111/mec.15023

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Conservation genomics in the fight to help the recovery of the critically endangered Siamese crocodile *Crocodylus siamensis*

Balaji Chattopadhyay[1] | Kritika M. Garg[1] | Yun Jing Soo[1] | Gabriel W. Low[1] | Jackson L. Frechette[2] | Frank E. Rheindt[1]

[1]Department of Biological Sciences, National University of Singapore, Singapore, Singapore

[2]Fauna & Flora International (FFI) Cambodia Programme, Phnom Penh, Cambodia

Correspondence
Frank E. Rheindt, Department of Biological Sciences, National University of Singapore, Singapore, Singapore.
Email: dbsrfe@nus.edu.sg

Funding information
WCS-internal funding; SEABIG, Grant/Award Number: WBS R-154-000-648-646 and WBS R-154-000-648-733; Singapore Ministry of Education Tier 1 grant, Grant/Award Number: WBS R-154-000-691-112

Abstract

Endangered species are often characterized by low genetic diversity and it is imperative for conservation efforts to incorporate the knowledge obtained from genetic studies for effective management. However, despite the promise of technological advances in sequencing, application of genome-wide data to endangered populations remains uncommon. In the present study we pursued a holistic conservation-genomic approach to inform a field-based management programme of a Critically Endangered species, the Siamese crocodile *Crocodylus siamensis*. Using thousands of single nucleotide polymorphisms from throughout the genome, we revealed signals of introgression from two other crocodile species within our sample of both wild and captive-bred Siamese crocodiles from Cambodia. Our genetic screening of the Siamese crocodiles resulted in the subsequent re-introduction of 12 individuals into the wild as well as the selection of four individuals for captive breeding programmes. Comparison of intraspecific genetic diversity revealed an alarmingly low contemporary effective population size in the wild (<50) with evidence of a recent bottleneck around Tonle Sap Lake. We also projected a probable future extinction in the wild (within fewer than five generations) in this population in the absence of re-introduction efforts. However, an increase in the number of potential breeders through re-introductions, including the one resulting from this project, could counter this trend. Our results have been implemented in ongoing re-introduction and captive breeding programmes, with major implications for the conservation management of Siamese crocodiles, and provide a blueprint for the rescue effort of other "terminally ill" populations of critically endangered species.

KEYWORDS
bottleneck, conservation genomics, ddRADseq, extinction, gene enrichment, Siamese crocodile

1 | INTRODUCTION

The Anthropocene epoch is characterized by our planet's sixth mass extinction event (Ceballos et al., 2015; Dirzo et al., 2014). While habitat loss is often the primary cause leading to the initial endangerment of species, genetic factors affect species' viability and susceptibility to extinction once genetic diversity has already become impoverished (Frankham, 2005; Hung et al., 2014; Keller & Waller, 2002; Rogers & Slatkin, 2017; Saccheri, Kuussaari, Kankare, & Vikman, 1998). Small isolated populations are specifically sensitive in this regard as they are more vulnerable to fragmentation

and genetic drift leading to increased inbreeding, concomitant loss of genetic diversity and heightened risk of extinction (Allendorf, Hohenlohe, & Luikart, 2010; Frankham, 2005; Keller & Waller, 2002). In many species that have become exceedingly rare, surviving individuals can no longer find mates of their own species and resort to hybridization, leading to a loss of genetic purity (Cabria et al., 2011; Cordingley et al., 2009; Gese et al., 2015; Lancaster, Gemmell, Negro, Goldsworthy, & Sunnucks, 2006; Pinto, Beja, Ferrand, & Godinho, 2016; Rheindt & Edwards, 2011; Schwartz et al., 2004; Todesco et al., 2016; Vuillaume, Valette, Lepais, Grandjean, & Breuil, 2015; Wayne & Shaffer, 2016).

The past decade has seen great advances in the potential for integrating genomic data with in situ and ex situ strategies to address important issues concerning biodiversity conservation (Allendorf et al., 2010; Funk, McKay, Hohenlohe, & Allendorf, 2012; Garner et al., 2016; McMahon, Teeling, & Höglund, 2014; Ouborg, Pertoldi, Loeschcke, Bijlsma, & Hedrick, 2010). Despite these advances, a disconnect remains between academic research in conservation genomics and applied conservation outcomes in the field (Corlett, 2017; Shafer et al., 2015). The few applied conservation studies that have made use of genome-wide markers have usually been restricted to addressing specific genetic factors affecting endangered populations, such as the loss of genetic diversity (Çilingir et al., 2017; Miller et al., 2011). Hence, there is a clear need for research efforts which take into account a whole range of genetic factors to understand the vulnerability of endangered populations, assess their future viability and help in conservation management efforts. In this study we pursued such a holistic conservation-genomic approach based on thousands of genome-wide markers to inform a field-based management programme of the Critically Endangered Siamese crocodile *Crocodylus siamensis*.

The Siamese crocodile is one of the most threatened species of crocodiles and is listed as critically endangered with fewer than 1,000 adults in the wild (Bezuijen, Simpson, Behler, Daltry, & Tempsiripong, 2012; Ihlow, 2008). Wild capture and poaching of Siamese crocodiles to stock farms for the skin trade have decimated wild populations over the past few decades (Daltry et al., 2016; Platt, Sovannara, Kheng, Stuart, & Walston, 2006; Ross, 1998; Siamese Crocodile Working Group, 2004; Simpson & Bezuijen, 2010). The global range of wild Siamese crocodiles has shrunk to a small fraction of its original (Figure 1) and consists of isolated populations (up to 50 individuals), only a few of which are reproductively active (Daltry et al., 2016; Ross, 1998; Sam et al., 2015; Simpson & Bezuijen, 2010). A decade-long survey has suggested the presence of five active nesting sites annually in Cambodia (Sam et al., 2015; Simpson & Bezuijen, 2010). At the same time, more than a million crocodile individuals of largely unknown provenance are thought to survive in captivity across farms in Southeast Asia (Daltry et al., 2016; Lapbenjakul et al., 2017; Simpson & Bezuijen, 2010). The small and fragmented remaining populations and resulting low levels of active breeders in the wild means that population recovery critically depends on captive breeding and re-introduction into the wild (Fitzsimmons et al., 2002; Lapbenjakul et al., 2017; Sam et al., 2015). Yet these efforts are complicated by the common practice of interbreeding Siamese crocodiles in captivity with other species, especially saltwater *C. porosus* and Cuban crocodile *C. rhombifer*, resulting in hybridization events and introgression threatening the genomic integrity of captive Siamese crocodile populations (Daltry et al., 2016; Fitzsimmons et al., 2002; Lapbenjakul et al., 2017; Sam et al., 2015).

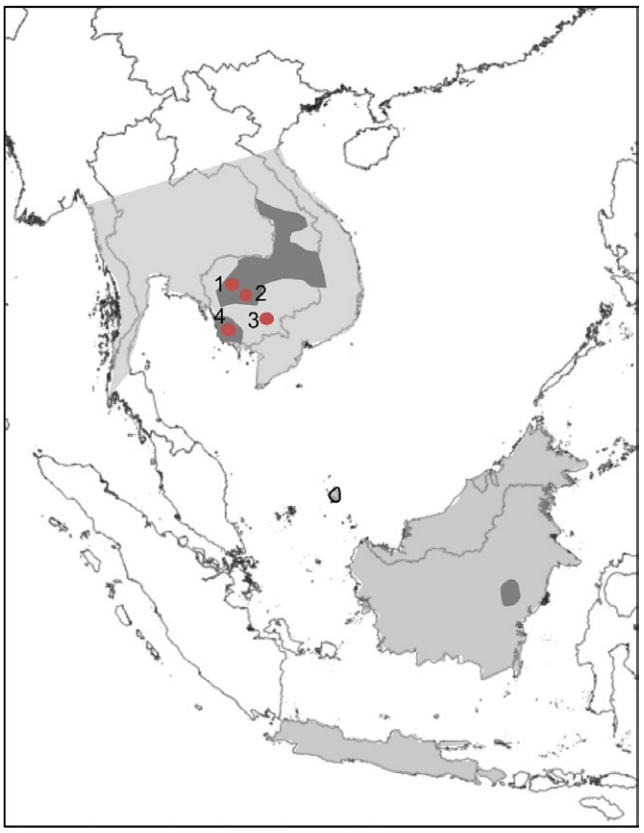

FIGURE 1 Map showing the distribution of the Siamese crocodile and sampling locations. The current distribution is depicted in dark grey. Light grey areas represent the historical distribution of the species. Map was adapted from Sam et al. (2015), Simpson and Bezuijen (2010) and Bezuijen et al. (2012). Sampling localities are indicated with red dots: (1) farms near Siem Reap; (2) farms near Tonle Sap; (3) Phnom Tamao Wildlife Rescue Centre; (4) Wildlife Conservation Society facility at Koh Kong [Colour figure can be viewed at wileyonlinelibrary.com]

Any efforts at Siamese crocodile conservation through captive breeding and re-introductions necessitate first a determination of the genomic purity of potential candidates for relocation (Fitzsimmons et al., 2002; Lapbenjakul et al., 2017; Sam et al., 2015), and second an assessment of the overall population-genomic diversity as well as genetic relatedness among them. In combination these measures can ensure a maximization of population-genomic diversity and a minimization of human-induced genetic introgression among individuals to be released into the wild (Lapbenjakul et al., 2017; Sam et al., 2015). In addition, an understanding of the genetic diversity and recent population history of wild populations of the

Siamese crocodile is required to assess its conservation status and viability in the wild.

In the present study we obtained samples of wild caught adult Siamese crocodiles that had been transferred to a farm near Cambodia's Tonle Sap Lake as well as numerous long-held captive individuals (wild caught as well as captive-bred) from breeding centres and farms for a total of 60 individuals. We used genome-wide data amounting to thousands of loci to genetically screen these individuals for captive breeding and re-introduction programmes. In the process, we ascertained the genomic purity of these individuals and the extent of introgression of non-Siamese crocodile DNA in their genomes. Rampant hybridization and back crosses are quite common in farm individuals (Fitzsimmons et al., 2002; Lapbenjakul et al., 2017; Sam et al., 2015) and hence one of the most important prerequisites for captive breeding and re-introduction programmes is to screen for human-induced introgression. Next, from a panel of single nucleotide polymorphisms (SNPs) mined specifically from genetically pure Siamese crocodiles, we studied overall genomic diversity to identify potential candidates for re-introduction and captive breeding. Finally, for the wild caught Tonle Sap population, we compared historical demographic models and investigated whether the population has retained genetic signatures of drastic population decline as reported for this species, while assessing its contemporary genomic diversity. To analyse the prospect of future survivability of the wild population we performed forward genetic simulations. Considering that our wild caught samples represent the diversity of the Tonle Sap population, we investigated whether re-introduction of individuals selected from our study would improve its survivability. Our study has had a direct impact on ongoing field-based conservation efforts as our genetic analyses were used to screen and select individuals for captive breeding and re-introduction efforts. At the same time, our forward genetic simulations can now be used to inform the scope and extent of future re-introduction activities required to generate viable populations.

2 | METHODS

2.1 | Sample acquisition

We obtained a total of 91 blood and muscle samples of Siamese, saltwater and mugger crocodile *Crocodylus palustris*, as well as the false gharial *Tomistoma schlegelii* as an outgroup (Supporting information Table S1). The Siamese crocodile samples (blood from 60 individuals) were collected from farms around Tonle Sap Lake (TS) and Siem Reap (FSR), Phnom Tamao Wildlife Rescue Centre (PT) and a Wildlife Conservation Society (WCS) facility at Koh Kong (KK). Samples from TS were identified as wild caught individuals, whereas samples from the other localities were obtained from a mixture of both wild caught and farm bred individuals (frequently the origin was unknown). We obtained 28 saltwater crocodile muscle tissue samples from the Lee Kong Chian Natural History Museum Cryogenic Collection in Singapore, the Singapore Zoo and a local Singaporean crocodile farm (Supporting information Table S1). Additionally, blood samples from two mugger crocodiles and one false gharial were provided by the Lee Kong Chian Natural History Museum Cryogenic Collection (Supporting information Table S1).

2.2 | Ethics statement

This study complied with all ethical regulations, and protocols were approved by National University of Singapore Institutional Animal Care and Use Committee (IACUC, Protocol Number: B16-0132).

2.3 | DNA extraction and ddRAD-Seq library preparation

We used GeneAll DNA Extraction kits and Qiagen DNeasy Blood and Tissue kits to extract DNA following the manufacturers' instructions. We used either QUBIT 2.0 Broad Range or High Sensitivity Assays (Invitrogen) to estimate the concentration of double stranded DNA.

We prepared double digest restriction enzyme-associated DNA sequence (ddRAD-Seq) libraries using a modification of Peterson, Weber, Kay, Fisher, and Hoekstra (2012) and Chattopadhyay et al. (2016). In brief, we chose *Eco*RI and *Msp*I for restriction digestion using the conditions given by Chattopadhyay et al. (2016) for restriction and ligation. We used Sera-Mag magnetic beads (Thermo Scientific) for size selection (300–500 bp fragments) and performed 12 PCR (polymerase chain reaction) cycles for final library preparation. We pooled samples in equimolar concentration and performed a quality check for optimal fragment size using a Fragment Analyser (Advanced Analytical). We obtained 150 bp paired-end reads from a single lane of the Illumina HiSeq 2500 (Singapore Centre for Environmental Life Sciences Engineering) for all the Siamese and saltwater crocodile samples. The mugger and false gharial samples were run on a separate HiSeq 4000 (150 bp paired-end) lane at the Genome Institute of Singapore. The two runs were spiked with 5% and 15% phiX, respectively, to avoid problems associated with low diversity.

2.4 | SNP calling and data filtering

We visually checked the quality of raw reads in FASTQC (Andrews,). We used two different pipelines, STACKS 1.44 (reference based; Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013) and PYRAD 3.0.66 (de novo; Eaton, 2014) to identify ddRAD loci. STACKS is better suited for population genomic analyses (Catchen et al., 2013) whereas PYRAD performs better in phylogenomic analyses (Eaton, 2014). For phylogenomic and introgression analyses we used the sequence data obtained from the PYRAD pipeline while all other analyses were performed with SNPs obtained from STACKS. We truncated the data to 135 bp and demultiplexed reads using process_radtags in STACKS. We removed low-quality reads (Phred score <10 at default settings) and reads with uncalled bases. For reference-based identification of ddRAD loci, we first aligned the demultiplexed reads to the saltwater crocodile

genome (GCF_001723895) sequenced by St John et al. (2012) using the bwa-mem algorithm in the BURROWS–WHEELER ALIGNER 0.7.12 (Li & Durbin, 2009). The aligned reads were then sorted according to coordinates using samtools sort in SAMTOOLS 0.1.19 (Li et al., 2009). We further used the ref_map.pl pipeline in STACKS to call SNPs and the populations program to filter SNPs. We set the stack depth to five (based on preliminary runs this stack depth gave us a high number of SNPs; see Supporting information Appendix S1) in the ref_map.pl pipeline to generate loci for SNP calling. We generated two data sets in STACKS, one including all samples across four species and the other with Siamese crocodile samples only. For the SNP set including all four species, we filtered loci such that they were present in all four species allowing for a maximum of 10% missing data within each species (see Supporting information Appendix S1). We did not allow for any missing data in the Siamese crocodile SNP set.

We used the demultiplexed data obtained from STACKS to identify phylogenomically informative loci using PYRAD. We used all samples and set the minimum locus coverage at 10 reads for each sample. We used the clustering threshold within and between samples (0.88) and allowed for a maximum of 10% missing data while generating the concatenated sequence matrix. We generated three different data sets (one using only read 1, another using only read 2 and a third using both reads) but retained only the read 1 data set containing the highest number of loci for final comparison between pipelines as well as for phylogenomic and introgression analyses (see below and also Supporting information Appendix S1).

To assess potential pipeline biases, we compared data sets from both the STACKS and the PYRAD pipeline generated for the purpose of interspecific analyses. We selected the STACKS refmap pipeline using both reads and the PYRAD de novo pipeline using read 1 only for all pipeline comparisons (see Supporting information Appendix S1).

For both SNP data sets obtained from STACKS and the read 1 only data obtained from PYRAD, we removed SNPs under potential selection using default settings in BAYESCAN 2.1 (Foll & Gaggiotti, 2008). In brief, BAYESCAN divides F_{ST} into a population-specific (beta) and a locus-specific component (alpha). It tests for significant deviation of the locus-specific component from the population-specific component. A significant alpha value would suggest that the locus is under selection. We used two different values of prior odds (10 and 100), assuming that the neutral model is 10/100 times more likely than the selection model at a locus, and used a 5% cutoff value for the false discovery rate to identify outlier loci. We further removed potentially linked SNP loci using PLINK 1.9 (Purcell et al., 2007), applying the indep-pairwise algorithm with a sliding window size of 25 SNPs, a step size of 10 and an r^2 correlation coefficient cut-off of 0.95.

2.5 | Genomic population structure and assignment across species

We used all samples of Siamese and saltwater crocodile and performed model-based Bayesian clustering in STRUCTURE 2.3.4 (Pritchard, Stephens, Rosenberg, & Donnelly, 2000) to ascertain species identities and infer putative admixed individuals. We ran STRUCTURE from $K = 1$ to $K = 8$ with 10 iterations per K, a 100,000 generation burnin and 500,000 Markov chain Monte Carlo steps per iteration. STRUCTURE runs were implemented without prior information about group membership. We obtained an estimate of the optimal number of genetic clusters (K) using Evanno's, Regnaut, and Goudet (2005) method while at the same time comparing results across various K values regardless of the optimal K in order to assess substructuring present within the data.

We further assessed genetic differentiation and introgression using a maximum likelihood-based clustering approach in ADMIXTURE 1.3 (Alexander, Novembre, & Lange, 2009). Similar to STRUCTURE, we obtained an estimate for the best possible number of clusters using cross-validation method (Alexander, Shringarpure, Novembre, & Lange, 2015). The fast implementation of the algorithm allowed us to test across a wide range of genetic clusters ($K = 1$–15) and choose the best K as the one with the lowest coefficient of error. Again, we compared across multiple K values to understand substructuring within the data. We also performed principal coordinate analysis (PCoA) on our data using GENALEX 6.5 (Peakall & Smouse, 2012) and discriminant analysis of principal components (DAPC) in ADEGENET 2.1.1 (Jombart, 2008) in the R package (R version 3.3.2, R Core Team, 2016; see Supporting information Appendix S1) as these analyses are less sensitive to deviations from population genetic assumptions.

Using simulations, we assessed if our SNP markers can successfully identify Siamese crocodiles and saltwater crocodiles and can distinguish between hybrids and backcrosses by following the methodology adopted by Burgarella et al. (2009) and Chattopadhyay et al. (2016). In short, we used the RAD-Seq genotypes of genetically pure Siamese and saltwater crocodiles (as determined through genetic assignment in STRUCTURE and ADMIXTURE as well as through Patterson's D-statistic (see next section), which searches for signals of introgression through allele sharing given a four-taxon topology) to simulate an SNP data set (number of loci identical to the observed data) for 100 pure individuals of both species in HYBRIDLAB 1.0 (Nielsen, Bach, & Kotlicki, 2006). Using the simulated Siamese and saltwater crocodile genotypes, we further simulated SNP data sets for 100 individuals each of various admixed categories (first-generation hybrids, backcross between first-generation hybrids and pure Siamese crocodile, backcross between first-generation hybrids and pure saltwater crocodile and cross between first-generation hybrids). HYBRIDLAB obtains allele frequencies from observed data and assumes Hardy–Weinberg equilibrium to generate simulated genotypes. Due to computational limitations, we were able to simulate only 100 genotypes of each admixture category. From the simulated data we randomly selected 20 pure individuals and five individuals from each of the four admixture categories ($n = 20$) without replacement and performed genetic assignment in STRUCTURE and ADMIXTURE using methods detailed in the previous sections. We generated five such random data sets of simulated individuals. We considered two genetic clusters for assignment analyses and performed separate runs, one only for simulated purebreds and the other consisting of both purebreds as well as different admixture categories. We used

a general cutoff for the ancestry coefficient (q) of $0.90 < q < 0.10$ for pure individuals and $0.90 \geq q \geq 0.10$ for admixed individuals, as these values are considered a standard in genetic assignments of wild populations (Burgarella et al., 2009). We then obtained estimates of (a) hybrid proportion (proportion of individuals in the sample identified as hybrids), (b) efficiency (power to detect true hybrid and purebred individuals calculated as the proportion of correctly identified individuals within each category), (c) accuracy (number of correctly identified individuals within a category divided by total number of individuals assigned to that category), and (d) type I error (number of purebreds incorrectly identified as hybrids divided by the total number of simulated purebreds).

2.6 | Phylogenetic reconstruction and introgression analyses

We reconstructed the phylogeny of our samples using the maximum likelihood-based concatenation method in RAXML GUI 1.5 (Silvestro & Michalak, 2012). We performed a full likelihood tree search assuming a GTR + gamma model of sequence evolution and using a rapid bootstrapping (1,000 replicates) approach. The final tree was midpoint rooted and viewed in FIGTREE 1.4.2 (Rambaut, 2015).

Based on the observed phylogenetic relationships (see below), we calculated Patterson's *D*-statistic (=ABBA-BABA test) as implemented in PYRAD to identify Siamese crocodiles that show signs of introgression from saltwater crocodiles. In a four-taxon test we considered wild Siamese crocodiles as pure, all captive and farm Siamese crocodiles as potentially introgressed, saltwater crocodiles as the source of introgression, and the false gharial as an outgroup (Supporting information Figure S1). We performed individual-based ABBA-BABA tests on each sample in PYRAD with Bonferroni corrections to account for multiple comparisons.

2.7 | Genomic diversity and population differentiation within Siamese crocodiles

We calculated observed heterozygosity for sample source (KK, TS, FSR and PT) along with the effective number of alleles and inbreeding coefficient using GENODIVE 2.0b27 (Meirmans & Van Tienderen, 2004). Furthermore, we compared the genetic diversity of wild-caught (Tonle Sap population) and farm animals (Koh Kong, Siem Reap, and Phnom Tamao Wildlife Rescue Centre). We obtained estimates of homozygosity by loci (Aparicio, Ortego, & Cordero, 2006), proportion of heterozygous loci and internal relatedness (Amos et al., 2001) using the GENHET (Coulon, 2010) function in R. We estimated the inbreeding coefficient for each individual in PLINK based on homozygosity. To test the efficiency of our markers in estimating relatedness, we simulated 100 dyads each of the following five relationships assuming the same error rate as the estimation step: full-sibs, half-sibs, first cousins, second cousins and unrelated. In the end, we explored correlations between true values and the seven estimates of relatedness computed by the COANCESTRY program (Wang, 2011; Supporting information Table S2). Given that all seven parameters emerged as equally good estimators of relatedness, we chose the widely used Queller and Goodnight statistic (Queller & Goodnight, 1989).

We further tested for evidence of population subdivision within our data set of Siamese crocodiles. Three out of four populations consisted of both wild caught animals of unknown geographical origin as well as farm bred individuals, so it is important to test for population-genetic structure that may point to different sources of captive stock. The SNP data set consisting only of Siamese crocodile samples was subjected to analyses of population differentiation and admixture using the same STRUCTURE, ADMIXTURE and PCoA approaches as outlined in previous sections. In addition, we also obtained networks to understand genetic differentiation using the k-nearest neighbour approach as implemented in the R package NETVIEW (Neuditschko, Khatkar, & Raadsma, 2012; Steinig, Neuditschko, Khatkar, Raadsma, & Zenger, 2016; see Supporting information Appendix S1).

2.8 | Historical demography and extinction probability

We reconstructed the historical demography of the wild-caught samples from the Tonle Sap population using the composite likelihood approach implemented in FASTSIMCOAL 2.6.03 (Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013). We used the folded site frequency spectrum (SFS) estimated using ARLEQUIN 3.5 (Excoffier & Lischer, 2010) to compare various demographic models. Using available information of the general decline of Siamese crocodiles across their range, we compared four models: constant population size model (null model), a scenario of a bottleneck, a continuous decline model and a scenario of a bottleneck followed by continuous population decline from the next generation onwards. We used log uniform prior distributions for all parameters (Supporting information Table S3), and for each model performed 50 independent runs in FASTSIMCOAL2. Furthermore, for each run we performed 100,000 simulations to estimate the expected SFS and likelihood of the given demographic model. To avoid issues of local maxima, we performed 40 cycles using a conditional maximization algorithm (ECM) for parameter estimations. For each model, from the 50 independent runs we chose the run in which the estimated maximum likelihood was closest to the observed maximum likelihood for model selection following Johnson and Omland (2004). We estimated the Akaike information criterion (AIC), ΔAIC and the weight of the model to determine the best-fit model for our data. The model with the lowest AIC value and highest weight was chosen as the best fit of the data.

For parameter estimation from the best model, we performed another set of 50 runs for that model excluding the monomorphic loci (as we did not have an accurate estimate of the number of monomorphic loci). Hence, as suggested in the FASTSIMCOAL manual, we fixed the ancestral N_e for these runs. Ancestral N_e was estimated based on nucleotide diversity (θ_π) and $\theta_\pi = 2N_e\mu$ for haploid populations, where N_e is the effective population size and μ is the mutation rate per generation, a commonly used estimate for ancestral N_e

based on RAD-Seq data (González-Serna, Cordero, & Ortego, 2018; Lanier, Massatti, He, Olson, & Knowles, 2015; Massatti et al., 2018; Papadopoulou & Knowles, 2015, 2017). This long-term harmonic mean of N_e corresponds to a time of $2N_e$ generations with more weight on recent generations (for further details see Hare et al., 2011). We estimated nucleotide diversity in STACKS based on the sequence data for the filtered SNPs. We assumed a generation time of 25 years (Bezuijen et al., 2012), a mutation rate of 7.9×10^{-9} per base pair per generation (Green et al., 2014) and obtained an ancestral haploid N_e estimate of 247,000 at around 500,000 generations ago. All final N_e estimates reported from our analyses were converted to diploid N_e estimates. The mutation rate had been estimated based on the pairwise divergence between genomes of wild caught alligator and saltwater crocodile (Green et al., 2014). We chose the run with the estimated maximum likelihood closest to the observed likelihood, and performed 100 parametric bootstrap replicates (SFS simulated based on the number of SNPs) to estimate the confidence in parameter estimation.

We performed forward genetic simulations in QUANTINEMO (Neuenschwander, Hospital, Guillaume, & Goudet, 2008), a simulation program to simulate complex population genetic and demographic scenarios from an input of indvidual-based genotypic data (see Supporting information Appendix S1). The program obtained estimates of future genetic variation and diversity, which are representative of the predicted population fitness of future generations and can be correlated with the possibility of extinction in the wild (Frankham, 2005; Saccheri et al., 1998). We used the genotypes of the wild caught population of Tonle Sap as input into the forward genetic simulations. We performed individual-based forward genetic simulations, all of which had population size corresponding to census size. We performed two sets of simulations, one in which the carrying capacity was set to be equal to the census population size and another in which the carrying capacity was fixed to 50 (based on census data from throughout Cambodia; Sam et al., 2015; Starr, Han, & Daltry, 2010). The second set of simulations allowed for population growth whenever the carrying capacity was higher than the census population size. For every set of simulations, we further simulated four different census sizes using long-term census data on Siamese crocodile populations in Cambodia (Bezuijen et al., 2012; Ihlow et al., 2008; Platt et al., 2006; Sam et al., 2015; Simpson & Bezuijen, 2010; Starr et al., 2010) (see Supporting information Appendix S1 for more details). For all simulations we considered a promiscuous mating system with equal sex ratio. We simulated a closed population, as Siamese crocodile populations across Cambodia show infrequent dispersal and high roost fidelity (Eam, Sam, Hor, Mizrahi, & Frechette, 2017; Sam et al., 2015). We ran each simulation for 1,000 generations and obtained statistics every five generations. Each simulation was further replicated 100 times.

We also modelled the effect of re-introductions into the wild by including released individuals (n = 11) for allele frequency estimates and performing the above-mentioned simulations again. These simulations would provide information about the beneficial effect of releasing individuals. Hence our final data set consisted of information from 16 forward genetic simulations (eight with input data from the wild caught population of Tonle Sap and another eight with input data from the wild caught population of Tonle Sap plus re-introduced individuals).

3 | RESULTS

3.1 | Sampling and SNP calling

We obtained a total of ~254 million reads across all 91 samples, 232 million of which passed quality control. The average number of reads per sample was 2.5 million ± 1 million (*SD*). We removed four saltwater crocodile samples from further downstream analyses based on their low number of reads, as their inclusion would have drastically reduced SNP counts, and proceeded with analyses of the remaining 87 individuals. We obtained 8,057 SNPs using STACKS for the complete data set consisting of all four species and retained 5,325 SNPs after removing SNPs under selection and linkage disequilibrium (Supporting information Table S4).

For the data set consisting of only Siamese crocodiles, we removed two individuals with potential Cuban crocodile ancestry prior to SNP calling (see below). In addition, we also removed two individuals from the WCS facility at Koh Kong owing to high levels of missing data. In the end, we obtained a total of 2,067 neutral and unlinked SNPs for a data set of 56 Siamese crocodiles using STACKS (Supporting information Table S4).

With the PYRAD pipeline, we obtained 1,372 loci and 1,151 SNPs (total concatenated sequence length = 180,264 bp) using read 1 only. No locus was under selection using both values of prior odds.

3.2 | Genetic assignment

K = 2 was the best K for both STRUCTURE (based on the Δk method, Evanno et al., 2005; Supporting information Figure S2a) and ADMIXTURE (Supporting information Figure S2b). Both STRUCTURE and ADMIXTURE clearly differentiated between Siamese and saltwater crocodiles (Figure 2a; Supporting information Figure S3a). We also observed two admixed individuals only using STRUCTURE (Figure 2a; Supporting information Figure S3a). When we explored higher K values in both STRUCTURE (Figure 2b) and ADMIXTURE (Supporting information Figure S3b), the two admixed Siamese individuals were consistently found to exhibit a signal of human-induced introgression from a ghost genetic lineage not present in our samples (indicated with blue arrows in Figure 2b; Supporting information Figure S3b). As it is known that Cuban crocodiles were widely introduced into Southeast Asian crocodile farms in the 1950s and were crossed and backcrossed with Siamese crocodiles (Fitzsimmons et al., 2002; Sam et al., 2015), we believe that these two individuals potentially bear the evidence of human-induced introgression between Siamese and Cuban crocodiles in captivity.

Similar to STRUCTURE and ADMIXTURE, PCoA and DAPC analyses separated Siamese and saltwater crocodiles as well as the two mugger crocodiles and the single gharial sample (Figure 2c; Supporting information Figure S4b). In the PCoA, the first axis mainly distinguished

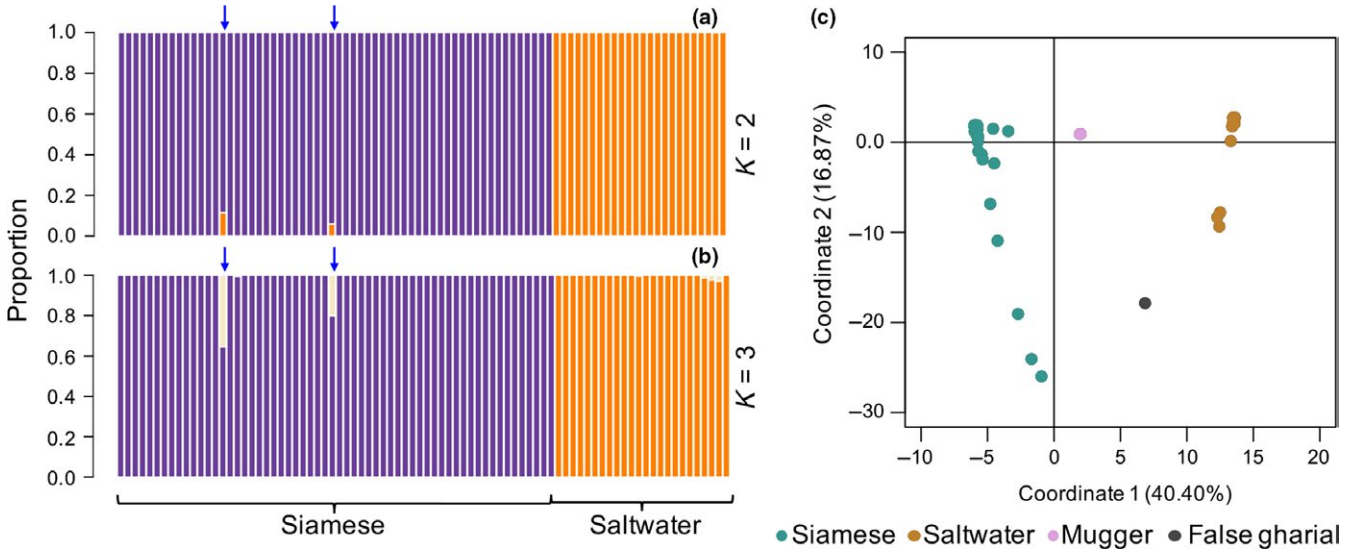

FIGURE 2 Genetic assignment for Siamese and saltwater crocodile samples as implemented in Structure for (a) $K = 2$ and (b) $K = 3$. Each bar represents an individual and the *y*-axis represents proportions of co-ancestry. The blue arrows indicate Siamese crocodiles introgressed with a genomic ghost lineage. (c) Two-dimensional clustering of individuals obtained from principal coordinate analysis. The first two coordinates explained ~57% of the variation in the data [Colour figure can be viewed at wileyonlinelibrary.com]

across the four species whereas the second axis, which carried much less overall information, revealed considerable intraspecific differentiation within both Siamese and saltwater crocodiles (Figure 2c). Saltwater individuals seemed to form two distinct clusters whereas Siamese crocodiles were arranged into a main cluster with outliers mainly from the WCS facility at Koh Kong (but also one captive farm individual from Siem Reap; Figure 2c).

Simulated genotypes suggest that our SNPs are powerful in differentiating between various purebred and admixed classes (Supporting information Table S5). All simulated individuals are classified into correct categories using both STRUCTURE and ADMIXTURE. We observed 100% accuracy and efficiency in identifying purebreds and hybrids using our SNP set (Supporting information Table S5).

3.3 | Phylogenomic reconstruction

Phylogenomic reconstruction returned species identities similar to genetic assignment (Figure 3). One of the Siamese crocodile samples that had shown signs of admixture in the co-ancestry plots (Figure 2b; Supporting information Figure S3b) emerged as basal to all other Siamese crocodiles in the phylogenetic tree (indicated with blue arrow, Figure 3). Furthermore, the Indian mugger emerged as sister to the Siamese crocodile in agreement with Oaks (2011).

3.4 | Introgression analyses

As the number of loci drastically differed among the three PYRAD data sets (read 1 only generated 1,372 loci, both reads generated 465 loci and read 2 only generated 468 loci; see Supporting information Appendix S1), we used only the read 1 data set consisting of 1,372 loci for ABBA-BABA analyses. Multiple individuals from all sampled captive populations (number of ABBA sites greater than BABA sites) as well as from the wild caught panel from Tonle Sap (number of BABA sites greater than ABBA sites) showed signs of ancestral introgression with the saltwater crocodile genome according to our ABBA-BABA tests (Supporting information Figure S1, Table S1). Following Bonferroni correction, four farm individuals and two wild-caught individuals from Tonle Sap revealed statistically significant ($p = 0.001$; corrected p value $<2.8 \times 10^{-8}$) introgression (Supporting information Table S1).

3.5 | Genomic diversity and screening for breeding and re-introduction

Considering that introgression with saltwater crocodiles may be an ancient natural phenomenon, we retained individuals identified as introgressed in our SNP data set consisting of Siamese crocodiles only. We also included introgressed individuals in all analyses other than simulations of historical demography and prediction of extinction. The average heterozygosity and inbreeding coefficient were similar across all four populations (Table 1). The observed heterozygosity and effective number of alleles for each population were low (Table 1). Our markers had high power in estimating true relatedness values (Supporting information Table S2). Overall, there was no significant difference in heterozygosity and inbreeding coefficients between wild caught and farm individuals (Supporting information Table S6). Wild caught individuals from Tonle Sap were significantly related compared to the farm individuals (Supporting information Table S6).

During the course of this research, a subset of our study individuals was released into the wild ($n = 12$) or selected for captive breeding ($n = 4$). We also ascertained the genetic diversity of these individuals. The individuals chosen for captive breeding and re-introduction had a higher genetic diversity when compared to the

remaining individuals (Supporting information Table S7). Individuals chosen for release and for captive breeding had lower internal relatedness, pairwise relatedness, inbreeding coefficients and homozygosity by loci compared to the remaining samples (Supporting information Table S7). Similarly, the proportion of heterozygous loci was higher for released and captive breeding individuals as compared to others (Supporting information Table S7).

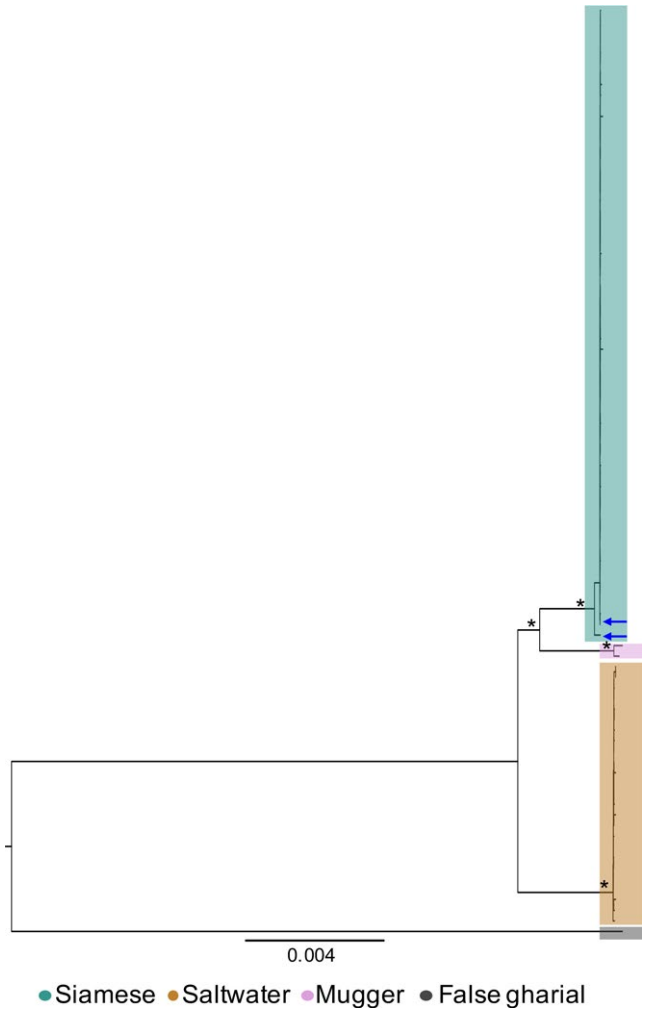

FIGURE 3 Midpoint rooted phylogenetic tree obtained from maximum likelihood-based phylogenomic reconstruction in RAXML using 1,372 concatenated sequence loci (180,264 bp in total) obtained from PYRAD using read 1 only. Species are colour-coded and nodal bootstrap support of 100 is marked with an asterisk. The blue arrows indicate Siamese crocodiles which were observed to be introgressed with a ghost lineage based on assignment analyses [Colour figure can be viewed at wileyonlinelibrary.com]

3.6 | Population-genetic differentiation within Cambodian populations of the Siamese crocodile

For both STRUCTURE and ADMIXTURE analyses using the data set consisting of Siamese crocodiles only, $K = 2$ was the best K (Supporting information Figure S5). At $K = 2$, individuals from the WCS facility at Koh Kong showed weak signs of separation from other populations (Figure 4a,b). At $K = 3$ and $K = 4$, both the Koh Kong and the Tonle Sap population showed weak levels of differentiation from the other two populations (Supporting information Figure S6). Similarly, in PCoAs, we observed three loose clusters, including a main Siamese crocodile cluster with individuals from all four farms, a second cluster composed of a few individuals from the WCS facility at Koh Kong and a third cluster consisting of several Tonle Sap individuals (Figure 4c). In agreement with these weak overall results for population differentiation within Siamese crocodiles, our network analyses did not reveal any meaningful evidence of population subdivision (see Supporting information Figure S7). Siamese crocodile samples for this study were obtained from different farms, and location information for individuals other than the Tonle Sap population remains unknown (see Section 2). Overall, genetic assignment tests and network analyses revealed no evidence of population structure.

3.7 | Demographic history and extinction probability

Historical demographic reconstruction suggests that the bottleneck model followed by a continuous decline was the most well-supported model (Figure 5 and Supporting information Table S8). Based on this model, the bottleneck occurred 100 years ago (assuming a generation time of 25 years; Bezuijen et al., 2012) resulting in a coalescent diploid N_e of 89 in the Tonle Sap wild population, which continued to decline to 36 (diploid N_e) in the present generation (Supporting information Table S9).

We performed forward genetic simulations for four different values of census population size (five, 10, 20 and 50) to understand the loss of genetic diversity in wild populations of Siamese crocodiles. For a closed population with a carrying capacity of 50 individuals, the heterozygosity was reduced to below 0.1 within 200 generations

TABLE 1 Summary statistics for each population of Siamese crocodile based on the SNP data set for Siamese crocodiles only

Species	Population code	Number of samples	Effective number of alleles	Observed heterozygosity	Inbreeding coefficient
Siamese crocodile	FSR	12	1.278	0.205	−0.148
	PT	15	1.297	0.216	−0.12
	KK	11	1.236	0.186	−0.27
	TS	18	1.266	0.198	−0.181

Notes. FSR: crocodiles from farm around Siem Reap; KK: crocodiles from Wildlife Conservation Society facility at Koh Kong; PT: Phnom Tamao Wildlife Rescue Centre; TS: wild caught crocodiles from a farm next to Tonle Sap Lake.

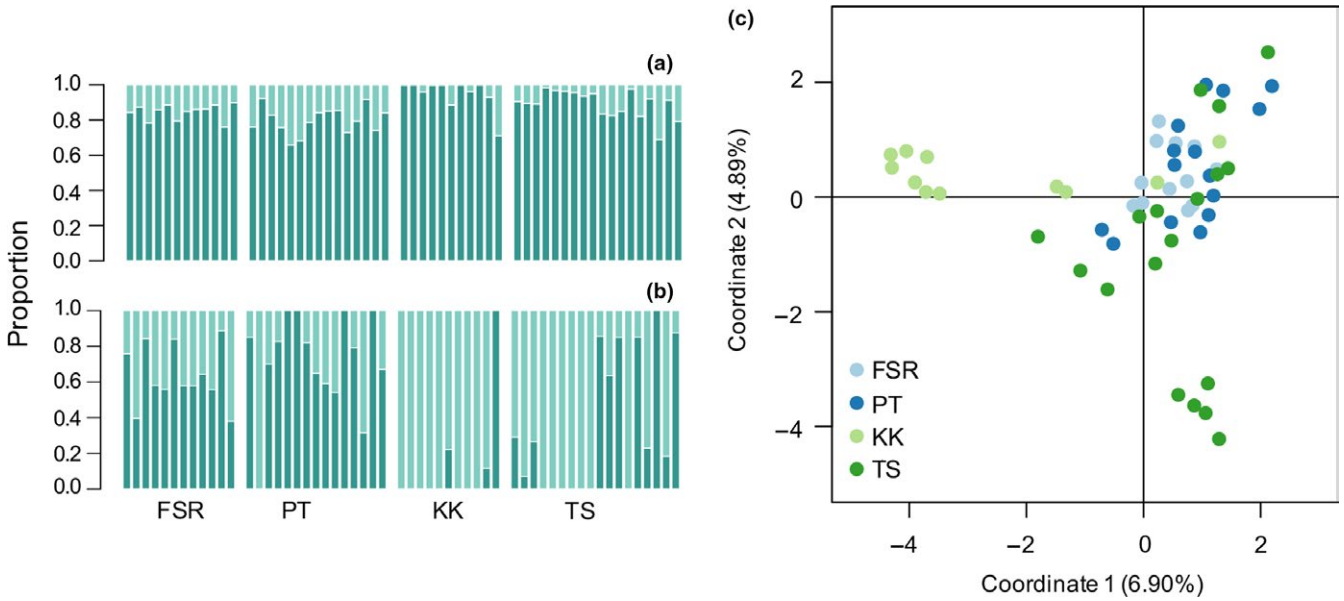

FIGURE 4 Population subdivision within Siamese crocodiles at $K = 2$ based on (a) Bayesian clustering and genetic assignment in Structure and (b) maximum likelihood-based assignment in ADMIXTURE. (c) Principal coordinate analysis plot showing segregation of Siamese crocodile individuals in two-dimensional space. The first two coordinates explained ~12% of the variation in the data. Population names: FSR: crocodiles from farms around Siem Reap; PT: Phnom Tamao Wildlife Rescue Centre; KK: crocodiles from farms at Wildlife Conservation Society facility at Koh Kong; and TS: wild caught crocodiles from a farm next to Tonle Sap Lake [Colour figure can be viewed at wileyonlinelibrary.com]

for all four census population size estimates and almost all loci became fixed within the population by 400 generations (Figure 6a,b). For all these simulations we observed a decrease in variability.

However, when the carrying capacity was assumed to equal the census population size, then drastic declines in heterozygosity and a simultaneous increase in the number of fixed loci was detected within few generations relative to the census population size (Figure 6c,d). For census population sizes of 5–10, extinction ensued within the first five generations, whereas for populations with census sizes of 20 and 50, all loci were fixed within 100–200 generations (Figure 6c,d). It seems that when we allow for growth (high carrying capacity), even a small population can recover and slow down the loss of variability.

We observed a moderately beneficial effect of introducing captive individuals (Supporting information Figure S8; Table S10): for each subsequent generation, the increase in the number of fixed loci and the decrease in expected heterozygosity were less pronounced when accounting for re-introductions (Supporting information Figure S8, Table S10). This result demonstrates that each re-introduction event considerably slows the decline of population-genetic diversity.

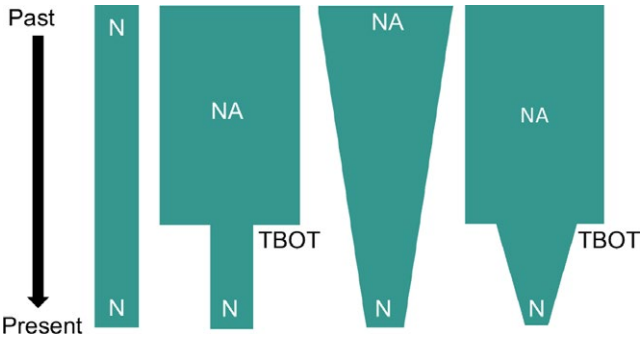

FIGURE 5 Models of demographic history compared to test for declines in the wild-caught Tonle Sap population of Siamese crocodiles. The first model is a null model assuming no change in population size over time, while the next three models assume a bottleneck, continuous decline and a bottleneck followed by continuous decline respectively. N, present population size; NA, the predecline population size. The time of population bottleneck is represented by TBOT [Colour figure can be viewed at wileyonlinelibrary.com]

4 | DISCUSSION

4.1 | Siamese crocodiles are introgressed with DNA from other crocodile species

Our SNP markers allow for a robust identification of Siamese and closely related crocodile species and additionally contain information on intraspecific variability within Siamese crocodiles (Figures 2 and 4; Supporting information Table S5), rendering them particularly useful for genetic investigation of these animals. Our assignment tests revealed admixture from a ghost lineage into two captive individuals of Siamese crocodile, which we hypothesize is from the Cuban crocodile genome (Figure 2b; Supporting information Figure S3b). Cuban crocodiles have been widely interbred historically with Siamese crocodiles across Southeast Asian crocodile farms to improve leather quality (Fitzsimmons et al., 2002; Sam et al., 2015).

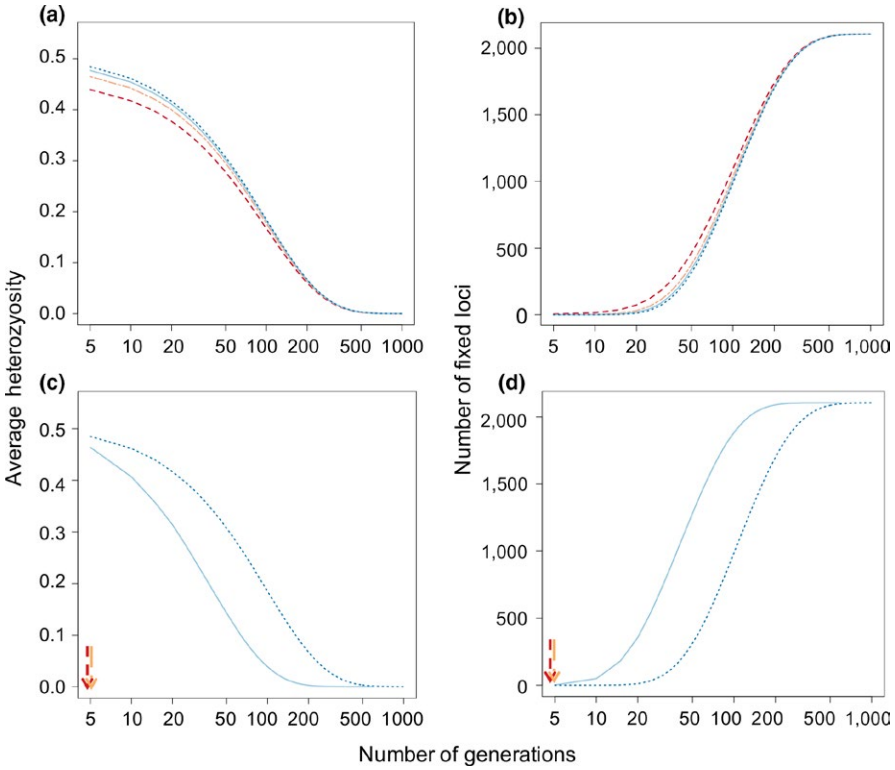

FIGURE 6 Forward genetic simulation-based predictions of decline in average heterozygosity (a, c) and increase in number of fixed loci (b, d) under different carrying capacities (50 for a and b, and carrying capacity equalling the census population size for c, d). Colours and line type represent different census population sizes (red dashed line = 5, orange double dashed line = 10, light blue solid line = 20 and dark blue dotted line = 50). Coloured arrows indicate extinction (reduction of the value of the respective summary statistic to zero). This occurred when the census population size was five or 10. The simulations were carried out for Tonle Sap individuals only [Colour figure can be viewed at wileyonlinelibrary.com]

While Cuban crocodile samples were not available for this study to conclusively test the identity of the ghost lineage, historical references of introgression strongly point in this direction (Fitzsimmons et al., 2002; Sam et al., 2015).

Our genome-wide data additionally reveal a high incidence of genetic introgression from saltwater crocodiles into Siamese crocodiles, affecting roughly 10% of individuals across both wild-caught as well as captive stocks (Supporting information Table S1). However, in contrast to the hypothesized admixture from Cuban crocodiles, the signature of introgression in Siamese crocodiles from the saltwater crocodile genome was not detectable through genetic assignment tests (Figure 2a,b; Supporting information Figure S3) but was found only by using Patterson's *D* statistic (Durand, Patterson, Reich, & Slatkin, 2011; Green et al., 2010), which tests for shared variability and has proven useful in the detection of subtle signatures of admixture across various genomes (Chattopadhyay et al., 2016; Durand et al., 2011; Green et al., 2010; Rheindt, Fujita, Wilton, & Edwards, 2013; Supporting information Table S1, Figure S1). Historically, both Siamese and saltwater crocodiles have shared a wide sympatric range across Southeast Asia, thereby strengthening the possibility of ancient admixture in the wild (Fitzsimmons et al., 2002; Sam et al., 2015). Simulations confirmed that our SNP markers have the power to successfully identify hybrids and backcrosses, products of recent admixture events occurring in the past two or three generations (Supporting information Table S5). However, a lack of a signature for contemporary admixture in the sampled individuals (Figure 2a,b; Supporting information Figure S3) led us to conclude that admixture events picked up by the ABBA-BABA analyses are likely to be of an origin preceding the last ~50–75 years, and possibly reflect ancient introgression that may have trickled into the Siamese crocodile genome across evolutionary timescales.

The wild population of Siamese crocodiles is thought to be less than 1,000 (Bezuijen et al., 2012; Ihlow et al., 2008). However, although over 1 million Siamese-like crocodiles are known to reside in crocodile farms throughout Southeast Asia, their potential use for conservation and re-introduction purposes has remained controversial, with their genomic purity in doubt (Daltry et al., 2016; Lapbenjakul et al., 2017; Sam et al., 2015; Simpson & Bezuijen, 2010). Decades of mixed pairings in captivity may have eroded the genomic purity of captive individuals to an unknown degree (Fitzsimmons et al., 2002; Sam et al., 2015). Using a panel of 60 crocodiles which were specifically donated for genetic screening related to captive breeding and re-introduction purposes, we detected signs of obvious admixture with what we suspect to be Cuban crocodile in two individuals, and more subtle signs of introgression with saltwater crocodile in an additional six individuals. However, more importantly, we show that in our panel of Siamese crocodiles, a substantial proportion (~87%) of individuals lack any detectable sign of introgression, and are presumably good candidates for re-introduction purposes. Even if our panel of captive crocodiles is not representative across Southeast Asian crocodile farms and contains a surplus of nonintrogressed individuals, one would still expect an increase of the global world population of pure Siamese crocodiles by two to three orders of magnitude from hundreds of individuals to tens or hundreds of thousands of individuals. The results of this study have already been implemented by the Wildlife Conservation Society in Cambodia, as the investigated panel has been screened

for nonintrogressed individuals that contain the highest genetic diversity and lowest inbreeding coefficients (Supporting information Table S7) to select for re-introduction and ex situ breeding.

One Siamese crocodile showing signs of historical, probably ancient introgression was among those released into the wild before the final results of this investigation were available. This finding emphasizes the importance of studies such as the present one in which genomic data and sophisticated analyses of gene flow are integrated within conservation action.

4.2 | Increased extinction risk following a bottleneck in the wild

There was no significant difference in inbreeding coefficients and heterozygosity between captive individuals and wild adults (Supporting information Table S6). However, wild adults showed higher average relatedness compared to captive individuals (Supporting information Table S6). Adults are regularly captured from the wild and used as breeders in crocodile farms. While this ongoing recruitment of wild individuals in farms may stall the decrease in genetic diversity in captivity, the wild population may suffer enormously as a consequence of continuous capture. The signal of the bottleneck for the Tonle Sap population dates back to four generations in the past (Figure 5; Supporting information Table S9), coinciding with reported times of decline of Siamese crocodile populations across their range (Bezuijen et al., 2012; Simpson & Bezuijen, 2010) and with our estimates of a low coalescent effective population size (Figure 5; Supporting information Table S9). This result is in agreement with reports of extensive overharvesting (Platt, Sovannara, Kheng, Thorbjarnarson, & Rainwater, 2004; Siamese Crocodile Working Group, 2004) in the Tonle Sap region, which has historically been considered one of the most suitable habitats for Siamese crocodiles (Kimura, 1969; Siamese Crocodile Working Group, 2004).

Our forward genetic simulations reveal that with current genetic diversity and population size, Siamese crocodiles might go extinct from many known nesting locations in less than 100 years (Figure 6c, d). This prediction is borne out by forward simulations showing that homozygosity and number of fixed alleles swiftly increase within a few generations under realistic assumptions about Cambodian population sizes as corroborated by field studies (Sam et al., 2015; Figure 6). These latter two parameters are known to be precursors of fitness loss and extinction (Charlesworth & Willis, 2009; Kardos, Taylor, Ellegren, Luikart, & Allendorf, 2016; Keller & Waller, 2002; Rogers & Slatkin, 2017; Saccheri et al., 1998). Even in studies addressing extinction in the past using ancient fossil samples, for example in the case of woolly mammoths, the loss of heterozygosity has been correlated to loss in fitness, adaptability and subsequent extinction (Rogers & Slatkin, 2017).

An effective population size of around 500 in the wild is often thought to be required to reduce the effects of genetic drift (Frankham, Bradshaw, & Brook, 2014). Siamese crocodiles are characterized by a long generation time, low breeding success and low growth rate (Bezuijen et al., 2012), all of which adversely impact effective population size (Frankham, 2005; Hamilton, 2011). The low effective population size of the contemporary population of Tonle Sap suggests a high risk of imminent extinction in the wild under assumptions of a small total carrying capacity. However, when we modelled for perhaps less realistic greater overall values for carrying capacity, our viability projections for wild Siamese crocodiles improved drastically. The true carrying capacity of Siamese crocodiles in Cambodian wetland habitats is poorly understood, but given our estimates of past effective population sizes (Figure 6; Supporting information Table S9) and considering their large body size, slow reproductive cycle and heavy persecution, we assume that the true carrying capacity of Siamese crocodiles (fertile adults) in predominantly isolated Cambodian wetlands would be near the lower end of our projected spectrum.

In the near future, conservation priorities should include genetically coordinated re-introduction programmes and the maintenance of genetically pure stock across crocodile farms (Simpson & Bezuijen, 2010). When we modelled the effect of re-introducing individuals (using the genotypes of actually released individuals, except the introgressed individual), there was a moderate decrease in the fixation of loci (Supporting information Figure S8, Table S10). This is a positive sign and multiple re-introductions should be planned to increase the genetic diversity of wild populations. However, re-introductions should be accompanied by genomic studies to avoid release of introgressed individuals (see Supporting information Table S1).

Overall, our forward genetic simulations provided two major insights which can help avert extinction risks and assist in the recovery of Siamese crocodile populations in the wild. First, we observed a positive effect of released individuals in delaying the fixation of loci as this can help in stalling the erosion of genetic diversity in the wild. Our second observation was that if populations are allowed to grow (in terms of fertile individuals), specifically when coupled with re-introductions, the deleterious effects of small population size on genetic diversity can be effectively thwarted. However, all our simulations showed a gradual decrease in genetic diversity indices (Supporting information Figure S8; Table S10), suggesting that present re-introduction efforts may not be sufficient and more diversity needs to be introduced into the wild to stall the erosion of genetic diversity. We expect that more carefully coordinated re-introductions harnessing information from approaches such as forward genetic simulations will be capable of improving the genetic diversity of this wild population and enhancing its long-term survival while minimizing introgression. For example, 11 released individuals were able to delay the drop in diversity (Supporting information Figure S8, Table S10), hence we suggest that for upcoming re-introductions higher numbers of released individuals should be considered. For example, our simulations demonstrate that re-introduction efforts that increase the number of fertile breeding individuals to ~20 should lead to viable populations (Figure 6). Similar genetic analyses should be performed with other surviving populations of Siamese crocodiles as well.

4.3 | Genome-wide data facilitate captive breeding and re-introduction efforts

Based on analyses including only nonintrogressed individuals, Siamese crocodile samples in our data set appear to have a shallow population-genetic structure, with only incipient signs of differentiation of the populations at Tonle Sap and at the WCS facility at Koh Kong, respectively (Figure 4). In general, levels of genetic diversity across localities did not vary significantly (Table 1) and the subset from farms chosen for re-introduction (n = 12) and captive breeding (n = 4) had a comparatively high heterozygosity and low inbreeding coefficients, suggesting that captive farm animals can be used effectively to revitalize breeding populations (Supporting information Table S7). We recommend that candidates for re-introduction efforts and captive breeding should be genetically screened to choose individuals with high heterozygosity, low relatedness and low inbreeding coefficients (as selected in this study, see Supporting information Table S7), whose release into the wild may help in improving genetic diversity (see the forward genetic simulations).

Our efforts yielded four major outcomes of direct relevance to the conservation and management of this critically endangered crocodile:

1. Purchasing farm individuals for re-introduction programmes: following our study, we have strong indications that there are certainly tens of thousands, but perhaps hundreds of thousands of pure Siamese Crocodiles alive (mostly in farms), which can potentially be used for re-introductions.
2. Detection of hybrids and introgression remains one of the major goals for the conservation management of Siamese crocodiles. Going far beyond standard assignment tests employing specific analyses such as the ABBA-BABA approach revealed signatures of introgression not visible using more conventional methodologies. This approach requires thousands of SNPs and hence our data (the first genome-wide data for this group of organisms) provide a framework to effectively screen animals even for subtle signatures of introgression. Our study shows that we may be underestimating hybridization by employing only standard population genetic methods and should use methods such as ABBA-BABA tests to identify introgression.
3. Individuals from the test panel used in this study have been selected by managers for re-introduction into the wild (n = 12) and also for captive breeding (n = 4) by WCS directly on the basis of genetic diversity indices computed in our study.
4. Our forward genetic simulations provide critical information about the potential beneficial impact of released individuals, arguing in favour of "rescue re-introductions" into extremely small wild populations to improve the chances of halting the decline of population genetic diversity. These simulations are part of an arsenal of tools that can help formulate future conservation policy on Siamese crocodiles.

ACKNOWLEDGMENTS

Funding for parts of this project was provided by two SEABIG grants (WBS R-154–000–648–646 and WBS R-154–000–648–733), a Singapore Ministry of Education Tier 1 grant (WBS R-154–000–691–112) and WCS-internal funding. We thank Emilie Cros for help with lab work.

CONFLICT OF INTEREST

The authors declare no competing financial interests.

AUTHOR CONTRIBUTIONS

B.C., K.M.G. and F.E.R. conceived and designed the study, S.Y.J. and J.F. contributed samples, S.Y.J., K.M.G., G.W.L. and B.C. generated the data, B.C., K.M.G. and S.Y.J. performed all analyses with input from FER, B.C., K.M.G. and F.E.R. wrote the manuscript with contributions from other co-authors.

DATA ACCESSIBILITY

Sequence data generated in this study have been submitted to the NCBI Sequence Read Archive (SRP117802).

ORCID

Balaji Chattopadhyay https://orcid.org/0000-0002-4423-3127
Kritika M. Garg https://orcid.org/0000-0003-3510-3408
Frank E. Rheindt https://orcid.org/0000-0001-8946-7085

REFERENCES

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*, 1655–1664. https://doi.org/10.1101/gr.094052.109

Alexander, D. H., Shringarpure, S. S., Novembre, J., & Lange, K. (2015). ADMIXTURE 1.3 Software Manual. http://software.genetics.ucla.edu/admixture/download.html

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of Conservation genetics. *Nature Reviews Genetics*, *11*, 697–709. https://doi.org/10.1038/nrg2844

Amos, W., Wilmer, J. W., Fullard, K., Burg, T. M., Croxall, J. P., Bloch, D., & Coulson, T. (2001). The influence of parental relatedness on reproductive success. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1480), 2021–2027. https://doi.org/10.1098/rspb.2001.1751.

Andrews, S.. (2010) FastQC: a quality control tool for high throughput sequence data. Retrieved from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Aparicio, J., Ortego, J., & Cordero, P. (2006). What should we weigh to estimate heterozygosity, alleles or loci? *Molecular Ecology*, *15*, 4659–4665.

Bezuijen, M., Simpson, B., Behler, N., Daltry, J., & Tempsiripong, Y. (2012). *Crocodylus siamensis*. The IUCN Red List of Threatened Species, e.T5671A3048087.

Burgarella, C., Lorenzo, Z., Jabbour-Zahab, R., Lumaret, R., Guichoux, E., Petit, R. J., … Gil, L. (2009). Detection of hybrids in nature: Application to oaks (*Quercus suber* and *Q. ilex*). *Heredity*, *102*, 442–452. https://doi.org/10.1038/hdy.2009.8

Cabria, M., Michaux, J., Gómez-Moliner, B., Skumatov, D., Maran, T., Fournier, P., & J. López de luzuriaga, Zardoya, R., (2011). Bayesian analysis of hybridization and introgression between the endangered european mink (*Mustela lutreola*) and the polecat (*Mustela putorius*). *Molecular Ecology*, *20*, 1176–1190. https://doi.org/10.1111/j.1365-294X.2010.04988.x

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*, 3124–3140. https://doi.org/10.1111/mec.12354

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., Gracía, A., Pringle, R. M., & Palmer, T. M. (2015). Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Science Advances*, *1*, e1400253. https://doi.org/10.1126/sciadv.1400253

Charlesworth, D., & Willis, J. H. (2009). The genetics of inbreeding depression. *Nature Reviews Genetics*, *10*, 783–796. https://doi.org/10.1038/nrg2664

Chattopadhyay, B., Garg, K. M., Kumar, A. V., Swami Doss, D. P., Rheindt, F. E., Kandula, S., & Ramakrishnan, U. (2016). Genome-wide data reveal cryptic diversity and genetic introgression in an Oriental cynopterine fruit bat radiation. *BMC Evolutionary Biology*, *16*, 41. https://doi.org/10.1186/s12862-016-0599-y

Çilingir, F. G., Rheindt, F. E., Garg, K. M., Platt, K., Platt, S. G., & Bickford, D. P. (2017). Conservation genomics of the endangered Burmese roofed turtle. *Conservation Biology*, *31*, 1469–1476. https://doi.org/10.1111/cobi.12921

Cordingley, J., Sundaresan, S., Fischhoff, I., Shapiro, B., Ruskey, J., & Rubenstein, D. (2009). Is the endangered Grevy's zebra threatened by hybridization? *Animal Conservation*, *12*, 505–513.

Corlett, R. T. (2017). A bigger toolbox: Biotechnology in biodiversity conservation. *Trends in Biotechnology*, *35*, 55–65. https://doi.org/10.1016/j.tibtech.2016.06.009

Coulon, A. (2010). GENHET: An easy-to-use R function to estimate individual heterozygosity. *Molecular Ecology Resources*, *10*, 167–169. https://doi.org/10.1111/j.1755-0998.2009.02731.x

Daltry, J. C., Langelet, E., Solmu, G. C., van der Ploeg, J., van Weerd, M., & Whitaker, R. (2016). Successes and failures of Crocodile harvesting strategies in the Asia pacific region. In A. A. Aguirre, & R. Sukumar (Eds.), *Tropical conservation: Perspectives on local and global priorities* (pp. 345–362). New York, NY: Oxford University Press.

Dirzo, R., Young, H. S., Galetti, M., Ceballos, G., Isaac, N. J. B., & Collen, B. (2014). Defaunation in the Anthropocene. *Science*, *345*, 401–406. https://doi.org/10.1126/science.1251817

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, *28*, 2239–2252. https://doi.org/10.1093/molbev/msr048

Eam, S., Sam, H., Hor, L., Mizrahi, M., & Frechette, J. (2017). Movement of captive-reared Siamese crocodiles *Crocodylus siamensis* released in the Southern Cardamom National Park, Cambodia. *Cambodian Journal of Natural History*, 102–108.

Eaton, D. A. (2014). PYRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, *30*(13), 1844–1849. https://doi.org/10.1093/bioinformatics/btu121

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, *9*(10), e1003905. https://doi.org/10.1371/journal.pgen.1003905

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*, 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fitzsimmons, N. N., Buchan, J. C., Lam, P. V., Polet, G., Hung, T. T., Thang, N. Q., & Gratten, J. (2002). Identification of purebred *Crocodylus siamensis* for reintroduction in Vietnam. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, *294*, 373–381. https://doi.org/10.1002/jez.10201

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, *180*, 977–993. https://doi.org/10.1534/genetics.108.092221

Frankham, R. (2005). Genetics and extinction. *Biological Conservation*, *126*, 131–140. https://doi.org/10.1016/j.biocon.2005.05.002

Frankham, R., Bradshaw, C. J., & Brook, B. W. (2014). Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation*, *170*, 56–63. https://doi.org/10.1016/j.biocon.2013.12.036

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, *27*, 489–496. https://doi.org/10.1016/j.tree.2012.05.012

Garner, B. A., Hand, B. K., Amish, S. J., Bernatchez, L., Foster, J. T., Miller, K. M., … Luikart, G. (2016). Genomics in conservation: Case studies and bridging the gap between data and application. *Trends in Ecology & Evolution*, *31*, 81–83. https://doi.org/10.1016/j.tree.2015.10.009

Gese, E. M., Knowlton, F. F., Adams, J. R., Beck, K., Fuller, T. K., Murray, D. L., … Waits, L. P. (2015). Managing hybridization of a recovering endangered species: The red wolf *Canis rufus* as a case study. *Current Zoology*, *61*, 191–205.

González-Serna, M. J., Cordero, P. J., & Ortego, J. (2018). Using high-throughput sequencing to investigate the factors structuring genomic variation of a Mediterranean grasshopper of great conservation concern. *Scientific Reports*, *8*, 13436. https://doi.org/10.1038/s41598-018-31775-x

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

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., … Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science*, *328*, 710–722. https://doi.org/10.1126/science.1188021

Hamilton, M. (2011). *Population genetics*. Chichester, UK: John Wiley & Sons.

Hare, M. P., Nunney, L., Schwartz, M. K., Ruzzante, D. E., Burford, M., Waples, R. S., … Palstra, F. (2011). Understanding and estimating effective population size for practical application in marine species management. *Conservation Biology*, *25*, 438–449. https://doi.org/10.1111/j.1523-1739.2010.01637.x

Hung, C.-M., Shaner, P.-J.-L., Zink, R. M., Liu, W.-C., Chu, T.-C., Huang, W.-S., & Li, S.-H. (2014). Drastic population fluctuations explain the rapid extinction of the passenger pigeon. *Proceedings of the National Academy of Sciences*, *111*(29), 10636–10641. https://doi.org/10.1073/pnas.1401526111

Ihlow, F., Bonke, R., Hartmann, T., Geissler, P., Behler, N., & Rödder, D. (2015). Habitat suitability, coverage by protected areas and population connectivity for the Siamese crocodile *Crocodylus siamensis* Schneider, 1801. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *25*, 544–554.

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

Kardos, M., Taylor, H. R., Ellegren, H., Luikart, G., & Allendorf, F. W. (2016). Genomics advances the study of inbreeding depression in the wild. *Evolutionary Applications*, *9*, 1205–1218. https://doi.org/10.1111/eva.12414

Keller, L. F., & Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, *17*, 230–241. https://doi.org/10.1016/S0169-5347(02)02489-8

Kimura, W. (1969). *Crocodiles in Cambodia. Research Report No. 3*. Atagawa Higashi Izu Town, Japan: Atagawa Tropical Garden and Alligator Farm.

Lancaster, M. L., Gemmell, N., Negro, S., Goldsworthy, S., & Sunnucks, P. (2006). Ménage à trois on Macquarie Island: Hybridization among three species of fur seal (*Arctocephalus* spp.) following historical population extinction. *Molecular Ecology*, *15*, 3681–3692. https://doi.org/10.1111/j.1365-294X.2006.03041.x

Lanier, H. C., Massatti, R., He, Q., Olson, L. E., & Knowles, L. L. (2015). Colonization from divergent ancestors: Glaciation signatures on contemporary patterns of genomic variation in Collared Pikas (*Ochotona collaris*). *Molecular Ecology*, *24*, 3688–3705.

Lapbenjakul, S., Thapana, W., Twilprawat, P., Muangmai, N., Kanchanaketu, T., Temsiripong, Y., … Srikulnath, K. (2017). High genetic diversity and demographic history of captive Siamese and Saltwater crocodiles suggest the first step toward the establishment of a breeding and reintroduction program in Thailand. *PLoS ONE*, *12*, e0184526. https://doi.org/10.1371/journal.pone.0184526

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. … 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Massatti, R., Prendeville, H. R., Larson, S., Richardson, B. A., Waldron, B., & Kilkenny, F. F. (2018). Population history provides foundational knowledge for utilizing and developing native plant restoration materials. *Evolutionary Applications*, *11*, 2025–2039. https://doi.org/10.1111/eva.12704

McMahon, B. J., Teeling, E. C., & Höglund, J. (2014). How and why should we implement genomicsAinto conservation? *Evolutionary Applications*, *7*, 999–1007.

Meirmans, P. G., & Van Tienderen, P. H. (2004). GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, *4*, 792–794. https://doi.org/10.1111/j.1471-8286.2004.00770.x

Miller, W., Hayes, V. M., Ratan, A., Petersen, D. C., Wittekindt, N. E., Miller, J., … Schuster, S. C. (2011). Genetic diversity and population structure of the endangered marsupial *Sarcophilus harrisii* (Tasmanian devil). *Proceedings of the National Academy of Sciences*, *108*(30), 12348–12353. https://doi.org/10.1073/pnas.1102838108.

Neuditschko, M., Khatkar, M. S., & Raadsma, H. W. (2012). NETVIEW: A high-definition network visualization approach to detect fine-scale population structures from genome-wide patterns of variation. *PLoS ONE*, *7*, e48375. https://doi.org/10.1371/journal.pone.0048375

Neuenschwander, S., Hospital, F., Guillaume, F., & Goudet, J. (2008). QUANTINEMO: An individual based program to simulate quantitative traits with explicit genetic architecture in a dynamic metapopulation. *Bioinformatics*, *24*, 1552–1553. https://doi.org/10.1093/bioinformatics/btn219

Nielsen, E. E., Bach, L. A., & Kotlicki, P. (2006). HYBRIDLAB (version 1.0): A program for generating simulated hybrids from population samples. *Molecular Ecology Resources*, *6*, 971–973.

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

Ouborg, N. J., Pertoldi, C., Loeschcke, V., Bijlsma, R. K., & Hedrick, P. W. (2010). Conservation genetics in transition to conservation genomics. *Trends in Genetics*, *26*, 177–187. https://doi.org/10.1016/j.tig.2010.01.001

Papadopoulou, A., & Knowles, L. L. (2015). Species-specific responses to island connectivity cycles: Refined models for testing phylogeographic concordance across a Mediterranean Pleistocene Aggregate Island Complex. *Molecular Ecology*, *24*, 4252–4268. https://doi.org/10.1111/mec.13305

Papadopoulou, A., & Knowles, L. L. (2017). Linking micro-and macroevolutionary perspectives to evaluate the role of Quaternary sea-level oscillations in island diversification. *Evolution*, *71*, 2901–2917. https://doi.org/10.1111/evo.13384

Peakall, R., & Smouse, P. E. (2012). GENALEX 6.5: Genetic analysis in Excel. Population genetic software for teaching and research - an update. *Bioinformatics*, *28*, 2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, *7*, e37135. https://doi.org/10.1371/journal.pone.0037135

Pinto, P. V., Beja, P., Ferrand, N., & Godinho, R. (2016). Hybridization following population collapse in a critically endangered antelope. *Scientific Reports*, *6*, 18788. https://doi.org/10.1038/srep18788

Platt, S. G., Sovannara, H., Kheng, L., Stuart, B. L., & Walston, J. (2006). *Crocodylus siamensis* along the Sre Ambel River, southern Cambodia: Habitat, nesting, and conservation. *Herpetological Natural History*, *9*, 183–188.

Platt, S., Sovannara, H., Kheng, L., Thorbjarnarson, J., & Rainwater, T. (2004). Population status and conservation of wild Siamese crocodiles (*Crocodylus siamensis*) in the Tonle Sap Biosphere Reserve, Cambodia. *Natural History Bulletin of the Siam Society*, *52*, 133–149.

Pritchard, J. K., Stephens, M., Rosenberg, N. A., & Donnelly, P. (2000). Association mapping in structured populations. *The American Journal of Human Genetics*, *67*, 170–181. https://doi.org/10.1086/302959

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., … Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*, 559–575. https://doi.org/10.1086/519795

Queller, D. C., & Goodnight, K. F. (1989). Estimating relatedness using genetic markers. *Evolution*, *43*, 258–275. https://doi.org/10.1111/j.1558-5646.1989.tb04226.x

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rambaut, A. (2015) FigTree v1. 4.2: Tree figure drawing tool. Retrieved from http://tree.bio.ed.ac.uk/software/figtree/

Rheindt, F. E., & Edwards, S. V. (2011). Genetic introgression: An integral but neglected component of speciation in birds. *The Auk*, *128*, 620–632. https://doi.org/10.1525/auk.2011.128.4.620

Rheindt, F. E., Fujita, M. K., Wilton, P. R., & Edwards, S. V. (2013). Introgression and phenotypic assimilation in *Zimmerius* flycatchers (Tyrannidae): Population genetic and phylogenetic inferences from genome-wide SNPs. *Systematic Biology*, *63*, 134–152. https://doi.org/10.1093/sysbio/syt070

Rogers, R. L., & Slatkin, M. (2017). Excess of genomic defects in a woolly mammoth on Wrangel island. *PLoS Genetics*, *13*, e1006601. https://doi.org/10.1371/journal.pgen.1006601

Ross, J. P. (1998). *Crocodiles: Status survey and conservation action plan (IUCN, Gland Suiza)*. SSC Crocodile Specialist Group.

Saccheri, I., Kuussaari, M., Kankare, M., & Vikman, P. (1998). Inbreeding and extinction in a butterfly metapopulation. *Nature*, *392*, 491–494. https://doi.org/10.1038/33136

Sam, H., Leng, H., Ratanapich, N., Piseth, S., Sovannara, H., Simpson, B., … Daltry, J. C. (2015). Status, distribution and ecology of the Siamese

crocodile *Crocodylus siamensis* in Cambodia. *Cambodian Journal of Natural History*, *2*, 153–164.

Schwartz, M. K., Pilgrim, K. L., McKelvey, K. S., Lindquist, E. L., Claar, J. J., Loch, S., & Ruggiero, L. F. (2004). Hybridization between Canada lynx and bobcats: Genetic results and management implications. *Conservation Genetics*, *5*, 349–355. https://doi.org/10.1023/B:COGE.0000031141.47148.8b

Shafer, A. B., Wolf, J. B., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., … Zieliński, P. (2015). Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution*, *30*, 78–87. https://doi.org/10.1016/j.tree.2014.11.009

Siamese Crocodile Working Group (2004). Siamese Crocodile Working Group Meeting, 24–28 May 2004. *Crocodile Specialist Group Newsletter*, *23*, 18–20.

Silvestro, D., & Michalak, I. (2012). raxmlGUI: A graphical front-end for RAxML. *Organisms Diversity & Evolution*, *12*, 335–337. https://doi.org/10.1007/s13127-011-0056-0

Simpson, B. K., & Bezuijen, M. R. (2010). Siamese crocodile *Crocodylus siamensis*. *Crocodiles Status Survey and Conservation Action Plan*, 120–126.

St John, J. A., Braun, E. L., Isberg, S. R., Miles, L. G., Chong, A. Y., Gongora, J., … Ray, D. A. (2012). Sequencing three crocodilian genomes to illuminate the evolution of archosaurs and amniotes. *Genome Biology*, *13*, 415. https://doi.org/10.1186/gb-2012-13-1-415

Starr, A., Han, S., & Daltry, J. (2010). Monitoring and nest surveys reveal status and threats of community-protected *Crocodylus siamensis* sub-populations in Cambodia. *Crocodile Specialist Group Newsletter*, *29*, 7–9.

Steinig, E. J., Neuditschko, M., Khatkar, M. S., Raadsma, H. W., & Zenger, K. R. (2016). netview p: A network visualization tool to unravel complex population structure using genome-wide SNPs. *Molecular Ecology Resources*, *16*, 216–227.

Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., … Rieseberg, L. (2016). Hybridization and extinction. *Evolutionary Applications*, *9*, 892–908. https://doi.org/10.1111/eva.12367

Vuillaume, B., Valette, V., Lepais, O., Grandjean, F., & Breuil, M. (2015). Genetic evidence of hybridization between the endangered native species *Iguana delicatissima* and the invasive *Iguana iguana* (Reptilia, Iguanidae) in the Lesser Antilles: Management implications. *PLoS ONE*, *10*, e0127575. https://doi.org/10.1371/journal.pone.0127575

Wang, J. (2011). COANCESTRY: A program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology Resources*, *11*, 141–145. https://doi.org/10.1111/j.1755-0998.2010.02885.x

Wayne, R. K., & Shaffer, H. B. (2016). Hybridization and endangered species protection in the molecular era. *Molecular Ecology*, *25*, 2680–2689.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Chattopadhyay B, Garg KM, Soo YJ, Low GW, Frechette JL, Rheindt FE. Conservation genomics in the fight to help the recovery of the critically endangered Siamese crocodile *Crocodylus siamensis*. *Mol Ecol*. 2019;28:936–950. https://doi.org/10.1111/mec.15023