**Supplemental Information**

Discovery of Facultative Parthenogenesis in a New World Crocodile

Warren Booth1, 2,\*, Brenna A. Levine2,3, Joel B. Corush4, Mark A. Davis4, Quetzal Dwyer5,

Roel De Plecker6, and Gordon W. Schuett2,7

1 Department of Entomology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA.

2 Chiricahua Desert Museum, Rodeo, New Mexico, USA

3 Department of Biology, Kean University, Union, New Jersey USA

4 Illinois natural History Survey, Prairie Research Institute, University of Illinois Urbana-Champaign, Illinois USA

5 Reptilandia Reptile Lagoon, Johnson City, Texas, USA

6 Parque Reptilandia, Dominical 5000, Puntarenas Province, Costa Rica

7 Department of Biology ǀ Neuroscience Institute, Georgia State University, Atlanta, Georgia, USA

\* E-mail: warrenbooth@vt.edu

**ORCID**: WB: **0000-0003-2355-0702**; BAL: 0000-0002-4326-591X; JBC: 0000-0001-8192-4691 MAD: 0000-0001-9034-9430; GWS: 0000-0002-2133-3723

**Supplemental Experimental Procedures**

**Background on the mother crocodile**

Parque Reptilandia in Costa Rica houses a wide variety of reptiles from all over the world, most of the species represented are found in Costa Rica. One of the park’s residents is an American crocodile (*Crocodylus acutus*) that joined the collection in December 2002. It was donated by Alejandro Solórzano after the closure of the National Serpentarium in San José, Costa Rica. Since its transition to Parque Reptilandia, the animal has been housed in isolation. The female crocodile at that time was still a young individual, about 2 years old. The enclosure measured 17 m by 8 m (Figure S1), with approximately two thirds comprised of a lagoon (deepest point ~1.2 m). Most of the land area had full sun throughout the day and one corner had a low tree growing to provide shade when necessary. On 17 January 2018, the female was observed exhibiting defensive nesting behavior and was not eating on a regular basis. Fourteen eggs were excavated, of which 5 were damaged. The average weight of the eggs was 101.1 grams and they measured 8.08 x 4.23 cm. Nine eggs were artificially incubated at a temperature of 29-30°C and a humidity of 90-95%. Signs of egg viability were made through candling, which involved shining a light through the egg. If the eggs were banded (Figure S2), eggs were considered viable. Seven eggs exhibited banding. After 3 months, no eggs hatched and 5 of the eggs smelled bad and were removed. On 10 July 2018, the remaining eggs were open and inspected. Three contained only white putrefied liquid, 1 contained a white and gray substance, and 1 contained a stillborn crocodile fetus (Figure S3, S4).

**Genomic Parthenogenesis Assessment**

Following the parameters used in Card et al. [1] as a guideline, variants were filtered using VCFtools v. 0.1.16 [2], the R (R core team 2022) package vcfR [3], and bedtools v2.30.0 [4], with the following criteria: (1) indels were excluded; (2) individuals with a read depth of less than 5 were excluded; (3) variants with a Phred quality score below 30 were excluded; (4) non-biallelic SNPs were excluded; (5) SNPs with significant statistical biases were removed using the hard filter ‘MQ < 40.0’; (6) SNPs were thinned to avoid the potential effects of linkage by randomly selecting one variant per 10 kb, 25 kb, and 50 kb region of the saltwater crocodile (*Crocodylus* *porosus*) genome [5] (Ghosh et al. 2020), resulting in three sets of filtered VCFs. Prior to thinning, SNPs that were not found in both individuals were removed using the bedtools v2.30.0 ‘intersect’ function [4].

For each set of VCFs, files were prepared for input to program ParthenoGenius [6]. First, maternal and offspring VCFs were merged into a single VCF with BCFtools [7,8]. PGDspider [9] was then used to convert each merged VCF into Structure file format. Structure files were modified to remove column two which contained Structure PopData values, such that the resulting files contained only the sample IDs (mother = M2, offspring = P1) and genotypes at each SNP retained after filtering. These files were then converted into csv format.

ParthenoGenius [6] was used to test for evidence and mode of parthenogenesis for each of the three thinned SNP data sets. Briefly, ParthenoGenius is a python program that first compares the number of the mother’s homozygous loci for which the offspring does not have identical genotypes to the mother against the number expected due to genotyping error alone based on a per-base genotyping error rate to determine whether discordance between maternal and offspring genotypes at maternal homozygous loci is more likely due to sources of genotyping error or the presence of paternal alleles. If the number of maternal homozygous loci for which the offspring has non-identical genotypes to the mother is less than the number expected due to genotyping error alone (i.e., the offspring is homozygous at all or nearly all of the mother’s homozygous loci), the offspring is called as a parthenogen and the proportion of maternal homozygous loci for which the offspring’s genotypes differ is recorded by ParthenoGenius as an updated estimated per-base error rate for the following heterozygosity scan. This assumes that genotyping error rate is consistent across the genome. If parthenogenesis is supported, ParthenoGenius then scans maternal heterozygous loci to identify those at which the offspring has retained heterozygosity for maternal alleles. If the number of maternal heterozygous loci at which the offspring is heterozygous is less than the number expected based on genotyping error alone assuming a null hypothesis of gametic duplication (i.e., the offspring is homozygous for maternal alleles at all or nearly all of the maternal heterozygous loci), the mode of parthenogenesis is called as gametic duplication. Alternatively, if the number exceeds that expected due to genotyping error alone (i.e., the offspring has retained heterozygosity at some or all maternal heterozygous loci), the mode of parthenogenesis is called as automixis (although there is no test to parse terminal from central fusion automixis). A per-base error rate of 0.1% was used for the initial homozygosity scan of the three thinned data sets, representing a conservative estimate of genotyping error rate on par with that of a filtered SNP array [10].

**Supplemental References:**

1. Card DC, Vonk FJ, Smalbrugge S, Casewell NR, Wüster W, Castoe TA, Schuett GW, Booth W. 2021. Genome-wide data implicate terminal fusion automixis in king cobra facultative parthenogenesis. *Sci. Rep*. **11,**7271.
2. Danece P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker R, Lunter G, Marth G, Sherry ST, McVean G, Durbin R and 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinform*. **27,** 2156-2158.
3. Knaus BJ, Grunwald NJ. 2017. VCFR: a package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour*. **17,** 44-53.
4. Quinlan AR. 2014. BEDTools: The swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinform.* **47,** 11.12.1-11.12.34.
5. Ghosh A, Johnson MG, Osmanski AB, Louha S, Bayona-Vásquez NJ, Glenn TJ, Gongora J, Green RE, Isberg S, Stevens RD, Ray DA. 2019. A high-quality reference genome assembly of the saltwater crocodile, *Crocodylus porosus*, reveals patterns of selection in Crocodylidae. *Genome Biol. Evol*. **12,** 3635-3646.
6. Levine BA, Booth W. (In prep) ParthenoGenius: A user-friendly heuristic for inferring presence and mode of facultative parthenogenesis using genomic data sets. https://github.com/brenna-levine/Levine-ParthenoGenius
7. Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinform.* **27,** 2987-2993.
8. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience*. **10,** giab008. doi: 10.1093/gigascience/giab008.
9. Lischer HEL, Excoffier L. 2012. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinform.* **28,** 298-299.
10. Saunders IW, Brohede J, Hannan GN. 2007. Estimating genotyping error rates from Mendelian errors in SNP array genotypes and their impact on inferences. *Genomics.* **90,** 291-296.

**Supplemental Figures**

Figure S1. American Crocodile (*Crocodylus acutus*) exhibit at Parque Reptilandia, Costa Rica.

Figure S2. American Crocodile (*Crocodylus acutus*) egg showing banding pattern indicative of potential egg viability.

Figure 3. Stillborn parthenogenetic American Crocodile (*Crocodylus acutus*) fetus



Figure 4. Stillborn parthenogenetic American Crocodile (*Crocodylus acutus*) fetus.

