Summary Report

Project title: Developing molecular markers to assess sex ratios in hatchling alligators (*Alligator mississippiensis)* in conservation applications

Samantha Bock
PhD candidate
Odum School of Ecology
University of Georgia

Supervisor: Dr. Benjamin Parrott

Associated citation: Bock, S.L., Smaga, C.R., McCoy, J.A., & Parrott, B.B. (2022). Genome-wide DNA methylation patterns harbor signatures of hatchling sex and past incubation temperature in a species with environmental sex determination. *bioRxiv*, https://doi.org/10.1101/2022.05.03.490459.

Introduction

Climate change threatens species persistence in wild populations via its influence on key demographic parameters. In ectotherms, shifting temperature regimes impact juvenile recruitment (Laloë et al., 2017; Noble et al., 2018; Pruett & Warner, 2021; Schwanz et al., 2010), adult fecundity (Crews et al., 1998; Wang et al., 2016; Warner & Shine, 2008), and, in some cases, population sex ratios (Bock et al., 2020; Janzen, 1994; Jensen et al., 2018; Reneker & Kamel, 2016). Fortunately, efforts to predict organismal and population level responses to changing environmental conditions are rapidly advancing with improvements in the precision and geographic resolution of global climate models and their integration with empirical data from physiological, ecological, and evolutionary studies (Carlo et al., 2018; Laloë et al., 2014; Mitchell et al., 2008; Riddell et al., 2018; Seebacher et al., 2015). Yet, one major challenge moving forward will be ground truthing these predictions to adapt conservation strategies accordingly. Genomic and epigenomic datasets represent powerful resources for conservationists in these efforts (Formenti et al., 2022; Layton & Bradbury, 2021; Waldvogel et al., 2020). The use of genomic tools in conservation is already widespread, allowing for assessments of population structure and connectivity (Wright et al., 2020), inbreeding depression (Dussex et al., 2021; Kardos et al., 2016), contemporary evolutionary responses to environmental change (Bi et al., 2019; Catullo et al., 2019), and adaptive capacities to cope with future change (Bay et al., 2018; Flanagan et al., 2018; Harrisson et al., 2014; Layton et al., 2021). The potential of epigenomic tools, on the other hand, is only just beginning to be realized. DNA methylation (DNAm), the covalent addition of a methyl group to a cytosine base, is one of the best-studied epigenetic marks, and variation in DNAm within and across individuals harbors key biological and ecological insights (Angers et al., 2010; Schübeler, 2015; Ziller et al., 2013). In the face of a rapidly shifting global climate, variation in genomic DNAm patterns has the potential to reveal cryptic organismal traits relevant to conservation and may provide insight into the environmental conditions from which they arose.

This potential holds unique relevance for thermally sensitive species facing climate-induced population collapse. Among these vulnerable taxa are species with temperature-dependent sex determination (TSD). Due to their reliance on temperature cues during incubation to determine offspring sex, species with TSD, including many turtles, all crocodilians, and some fish, are predicted to experience highly skewed population sex ratios as a result of climate change (Bock et al., 2020; Janzen, 1994; Laloë et al., 2014; Mitchell et al., 2008, 2010; Ospina-Alvarez & Piferrer, 2008; Reneker & Kamel, 2016). Such skews in sex ratios are already being detected in wild populations of TSD species (Jensen et al., 2018) and may exert lasting adverse impacts on population dynamics given the long generation times of many TSD species (Böhm et al., 2016). Despite these observations, efforts to actively monitor cohort sex ratios in the field remain limited. This is in part due to the lack of reliable, non-lethal methods of sexing individuals at early life stages. Since hatchlings do not tend to exhibit sexually dimorphic morphology, specialized training is required to determine hatchling sex (Allsteadt & Lang, 1995), and genetic markers of sex (i.e., sex-linked genes or alleles) are non-existent in TSD species (or may not be indicative of phenotypic sex in the case of temperature-induced sex-reversal (Holleley et al., 2015)). DNAm patterns in blood

represent a novel avenue by which those tasked with managing populations of TSD species under climate change might gain insight into ongoing sex ratio shifts as well as changes in field incubation conditions.

In this study, we investigate sex- and temperature-associated variation in the DNA methylomes of hatchling blood cells in the American alligator, a species with TSD. DNAm dynamics in blood cells are examined in the context of other sex-linked biological correlates such as circulating levels of sex steroids and gonadal DNAm patterns established during TSD. We further use these DNAm patterns to develop models that predict hatchling sex and past incubation temperature based on the methylation status of a subset of loci. Using these predictive models, we assess the potential utility of DNAm-based markers in the conservation of species with TSD under rapid environmental change. Due to the unique temperature-by-sex ratio reaction norm of the alligator, this species serves as a particularly insightful model to identify sex- and temperature-associated DNAm patterns. As in all other crocodilians, females are produced at both low and high incubation temperatures, while males are produced at intermediate incubation temperatures (Lang & Andrews, 1994). From the perspective of experimental design, this pattern of TSD allows us to disentangle DNAm variation linked to sex from that linked to temperature through the inclusion of females from both the low and high ranges of the reaction norm. The shape of the crocodilian temperature-by-sex ratio reaction norm also yields greater uncertainty in predictions of sex ratio shifts in response to climate change (Bock et al., 2020). Thus, monitoring sex ratios in the field is especially pertinent for these species and our results hold unique relevance for future conservation.

Methods

In June 2013, seven clutches of *A. mississipiensis* eggs were collected from Lake Apopka, FL and transported to Hollings Marine Laboratory in Charleston, SC (McCoy et al., 2016). Detailed methods regarding the incubation experiment can be found in the associated report, which sought to decipher temperature-dependent inter- and intra-sexual variation in gonadal gene expression (McCoy et al., 2016). Briefly, viable eggs were maintained at 30˚C until Ferguson stage 19, the opening of the canonical thermosensitive period (Lang & Andrews, 1994), after which eggs from each clutch were distributed across four incubation temperatures (30˚C, 32˚C, 33.5˚C, 34.5˚C), where they were maintained until hatch. At seven days post-hatch, a blood sample was taken from the postcranial sinus. Animals were then euthanized with a lethal dose of pentabarbital and necropsied to collect the whole brain and gonad-adrenal-mesonephros (GAM) complex. Fresh tissues were preserved in RNAlater and kept at -80˚C. Whole blood samples were stored in lithium heparin Vacutainer tubes on ice prior to centrifugation at 1500 rcf for 20 min. Plasma was isolated and RNAlater was added to the residual pelleted blood cells. Plasma was stored at -80˚C and blood cells in RNAlater were stored at -20˚C. Plasma concentrations of 17ß-estradiol (E2) and testosterone (T) were quantified via radioimmunoassays as described in (McCoy et al., 2016). Hatchling sex was assigned based on gonadal expression of *AMH* and *CYP19A1* in (McCoy et al., 2016). DNA was isolated from blood cells using a modified column approach based on the SV total RNA isolation system (Bae et al., 2021).

Reduced representation sequencing (RRBS) libraries were constructed with DNA from 24 blood samples – including those from eight 30˚C females, eight 33.5˚C males, four 34.5˚C females and four 34.5˚C males – using the Premium RRBS Kit (Diagenode, Cat. No. C02030032) according to manufacturer instructions. Resulting libraries were pooled and sequenced on a single Illumina NextSeq 500 High Output flowcell (75bp, single-end) with 20% Illumina PhIX control at the Georgia Genomics and Bioinformatics Core (University of Georgia, Athens, GA). Detailed methods regarding the bioinformatic pipeline used to process and analyze sequencing data can be found in the associated citation (Bock et al., 2022). Briefly, temperature- and sex-associated differentially methylated cytosines (DMCs) were identified via pairwise comparisons of treatments groups using logistic regression with correction for overdispersion in the R package methylkit (Akalin et al., 2012).

To develop a DNAm-based model to predict hatchling sex, we divided our samples into a training set of 20 individuals and a test set of four individuals. The test set was composed of one randomly selected individual from each temperature-sex combination. A percent methylation matrix for loci identified as having sex-associated differential methylation patterns was then modeled using penalized binomial logistic regression with elastic net regularization (alpha = 0.5) in the R package glmnet to identify sites for which methylation status was most predictive of sex in the training set (Zou & Hastie, 2005). In this modeling approach, the coefficients of loci that

are weak predictors of sex are reduced to zero, thereby eliminating them from the model. The performance of the model was then assessed by the correct assignment of sex for individuals in the test set. To develop a DNAm-based predictive model for incubation temperature, a similar approach was taken to the one described for developing a predictive model of sex. Loci exhibiting differential methylation with respect to temperature were used as input for the predictor selection. In this case, we fit a penalized linear regression with elastic net regularization. Model fit on the training set and performance on the test set was evaluated through the R-squared of the relationship between actual and predicted incubation temperature and the absolute error of the predicted incubation temperatures.

Results

A total of 462,236 cytosines was retained in our analyses following filtering. Of the CpGs analyzed, 120 (0.026%) loci displayed sex-associated differential methylation patterns across all females and males irrespective of temperature. When we relaxed the requirement that loci be differentially methylated across all females and males, we identified 1,307 (0.28%) sex-associated DMCs. This group included loci that were only differentially methylated between sexes in hatchlings incubated at 34.5˚C. Of the CpGs analyzed, 707 (0.15%) loci exhibited differential methylation with respect to incubation temperature and not sex.

We developed DNAm-based predictors of sex and past incubation temperature based on the methylation status of a subset of loci. For the predictive model of hatchling sex, 23 cytosines passed the threshold to be included in variable selection via elastic net regularization, and 20 of those cytosines were selected to be included in the final model. This model correctly assigned sex to all 20 of the training set individuals with minimal variability in the model predictions (Mean ± SE predicted probability male: Females = 0.07 ± 0.01; Males = 0.93 ± 0.01; Figure 1A). This model also correctly assigned sex to all 4 of the test set individuals, though there was more variability in the model predictions (Mean ± SE predicted probability male: Females = 0.28 ± 0.07; Males = 0.71 ± 0.01; Figure 1B). For the predictive model of past incubation temperature, 152 cytosines passed the threshold to be included in variable selection, and 24 of those cytosines were selected to be included in the final model. This model fit the training set well, with an R-squared greater than 0.99 between the actual incubation temperatures and fitted incubation temperatures. The mean absolute error for the training set individuals was 0.05˚C (Figure 1C). However, the model did not perform as well on the test set individuals. The R-squared for the test set was 0.81 and the mean absolute error of the model predictions was 1.20˚C (Figure 1D).

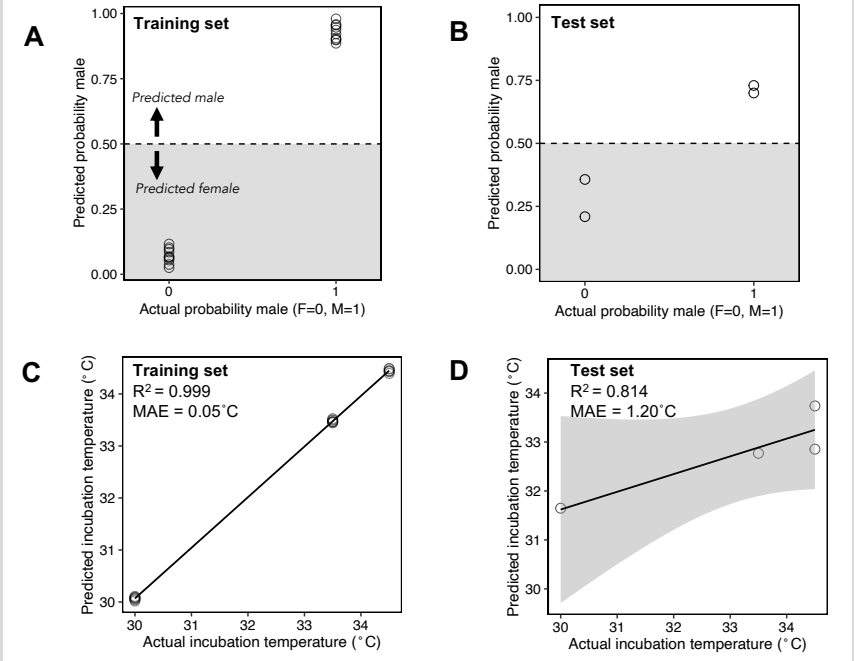


Figure 1. Predictive models of hatchling sex and past incubation temperature based on the methylation status of a subset of cytosines. (A) Hatchling sex predictions resulting from penalized binomial logistic regression with elastic net regularization for training set (n = 20). (B) Hatchling sex predictions for test set (n = 4). (C) Incubation temperature predictions resulting from penalized linear regression with elastic net regularization for training set (n = 20). (D) Incubation temperature predictions for test set (n = 4). The model predicting hatchling sex was based on the methylation status of 20 loci, while the model predicting incubation temperature was based on the methylation status of 24 loci. MAE = mean absolute error.

Conclusions

In the present study, we demonstrate that the DNA methylomes of blood cells from hatchling alligators are sexually dimorphic and retain signatures of past incubation temperature. We further show that DNAm patterns at a subset of loci can be used to predict sex and past incubation temperature. Collectively, our results support the potential utility of DNAm-based markers as tools for ecological monitoring and conservation in the face of environmental change. Further, these findings point to several critical considerations relevant to the implementation of epigenomic tools in conservation more broadly.

In the face of rapidly changing environmental conditions, monitoring the responses of key demographic parameters, like offspring sex ratios, to climate change in real time will play a critical role in the conservation of thermally sensitive species. This is the first study to use somatic DNAm patterns to predict phenotypic sex in a species with environmental sex determination. Intriguingly, we also demonstrate that DNAm patterns hold the potential to provide information about past incubation temperatures, though this finding requires validation in the context of fluctuating incubation temperatures as well as in older animals. Understanding how genetic and environmental variation contribute to these methylation patterns represents a critical target for future research that will advance efforts to realize the full potential of epigenomic tools in conservation contexts.