Gulf of Mexico Science

Volume 33
Number 2 *Number 2*

Article 8

2016

Thousands of Single Nucleotide Polymorphisms in the Critically Endangered Kemp's Ridley Sea Turtle (Lepidochelys kempii) Revealed by Double-Digest Restriction-Associated DNA Sequencing: Opportunities for Previously Elusive Conservation Genetics Research

Luis A. Hurtado
Texas A&M University

Xochitl De La Rosa-Reyna
Texas A&M University

Mariana Mateos
Texas A&M University

Donna J. Shaver
Padre Island National Seashore

Richard Metz
Texas A&M University

DOI: 10.18785/goms.3302.08

See next page for additional authors

Thousands of Single Nucleotide Polymorphisms in the Critically Endangered Kemp's Ridley Sea Turtle (Lepidochelys kempii) Revealed by Double-Digest Restriction-Associated DNA Sequencing: Opportunities for Previously Elusive Conservation Genetics Research

Authors

Luis A. Hurtado, Xochitl De La Rosa-Reyna, Mariana Mateos, Donna J. Shaver, Richard Metz, John Thiltges, Joshua Hill, Miguel A. Reyes-Lopez, and Charlie Johnson

SHORT PAPERS AND NOTES

Gulf of Mexico Science, 2016(2), pp. 214–218

THOUSANDS OF SINGLE NUCLEOTIDE POLYMORPHISMS IN THE CRITICALLY ENDANGERED KEMP'S RIDLEY SEA TURTLE (*LEPIDOCHELYS KEMPII*) REVEALED BY DOUBLE-DIGEST RESTRICTION-ASSOCIATED DNA SEQUENCING: OPPORTUNITIES FOR PREVIOUSLY ELUSIVE CONSERVATION GENETICS RESEARCH.—

Among sea turtles, the Kemp's ridley is the most endangered and geographically restricted, with its distribution mostly confined to the Gulf of Mexico (NMFS and USFWS, 2015). After experiencing a severe and sustained bottleneck that put this species on the verge of extinction, it appeared to be rebounding successfully, as evidenced by an exponential growth in the number of nests observed per nesting season, following decades of Mexico–United States bi-national efforts aimed at its recovery (Heppell et al., 2007). Unfortunately, nesting was severely reduced by ~35% during 2010 (the year of the BP Deepwater Horizon oil spill in the Gulf of Mexico), as compared to nesting rates in 2009 (NMFS and USFWS, 2015). Although nesting rebounded during 2011 and 2012 to levels similar to that of 2009, nesting declined drastically again during 2013 and experienced a further drop during 2014 (NMFS and USFWS, 2015; Shaver et al., 2016). The number of nests in 2014 represents a 46% decrease from 2012, which was the year with the highest recorded number of nests since 1965 (Sarti, 2014). Should nesting continue to decline, long-term species recovery efforts will be compromised. Therefore, there is deep concern about the future of the Kemp's ridley, and data to inform and assess bi-national management and conservation measures are urgently needed (Plotkin and Bernardo, 2014). Population genetics information crucial to the long-term conservation of the Kemp's ridley, including baseline data required for monitoring its future status, is lacking. This includes estimations of genomic diversity, effective population size, and number of breeders; assessment of levels of population differentiation; and detection of genomic signatures of bottlenecks.

Past research on basic population genetics of the Kemp's ridley was limited by the paucity of informative genetic markers. Three previous population genetics studies of the Kemp's ridley were based on a low number of microsatellite markers [three in Kichler et al. (1999), four in Stephens (2003), and two in Lara Rivera (2012)], which prevented robust inferences of genetic diversity, effective population size, and population structure. Five of the combined six microsatellite markers used in the above studies were originally characterized from other sea turtle species, in which the variability was generally higher (FitzSimmons et al., 1995). The additional marker was developed from the Kemp's ridley (Kichler, 1996), but subsequent sequencing revealed it is an imperfect (compound) microsatellite (Stephens, 2003). A more recent study (Frey et al., 2014) used 10 microsatellite loci characterized from other sea turtle species to genetically infer the annual numbers of Kemp's ridley nesting females in the Texas coast, based on dead embryo and hatchling specimens. However, no other genetic information was inferred. Development of additional markers that represent genome-wide variation should benefit conservation genetics research related to the Kemp's ridley.

Recent advances in high-throughput DNA sequencing can facilitate the discovery of hundreds to thousands of molecular markers useful for conservation genetics research in nonmodel organisms (Narum et al., 2013). Commonly used methods for conservation genomics of nonmodel organisms consist of obtaining a reduced representation library that includes many regions across the genome, sequenced for multiple individuals of a target species, with the goal of identifying single nucleotide polymorphisms (SNPs). One approach that is increasingly employed is restriction-associated DNA sequencing (RAD-seq), which uses restriction enzymes and size selection to generate the reduced representation library (Narum et al., 2013). In this study, we tested the potential of a variant of this method, the double-digest restriction-associated DNA sequencing (ddRAD-seq) technique (Peterson et al., 2012), to identify SNPs in the Kemp's ridley.

Methods.—Hind flipper biopsies were obtained from 16 nesting Kemp's ridleys at Padre Island National Seashore, Texas, during 2014. DNA was isolated from ~40 mg of tissue. Samples were lysed in 1000 µl of extraction buffer (592 µl

distilled water, 68 µl 0.5 M ethylenediamine tetraacetic acid, and 340 µl 10% sodium dodecyl sulfate) and 5 µl proteinase K (New England BioLabs), incubated for 3.5 hr at 65°C. After incubation, RNA was eliminated with 5 µl RNAse A (Thermo Scientific) for 5 min. Subsequently, 50 µl of 3 M potassium acetate were added and samples were incubated on ice for 30 min. The supernatant was mixed with 500 µl of chloroform. The aqueous phase was separated and the DNA was precipitated with 500 µl of isopropanol. The DNA pellet was washed with 70% ethanol, after which it was eluted in 100 µl of nuclease-free water. DNA was quantified with a NanoDrop™ 1000 Spectrophotometer (Thermo Scientific). Aliquots of 50 µl per sample, containing at least 1,500 ng, were submitted to the TAMU AgriLife Genomics and Bioinformatics Services facility for sample processing, library preparation, and sequencing on Illumina HiSeq 2500.

Libraries were prepared following the ddRAD-seq protocol (Peterson et al., 2012), with some modifications. For each sample, 200 ng of DNA was digested with the restriction enzymes PstI and MluCI in a final volume of 40 µl. Following heat inactivation, restriction-site–compatible, bar-coded Illumina-specific adaptors were ligated to the cut DNA such that the PstI cuts received bar-coded P5 adapters and MluCI sites received half of the P7 adapter. Following heat inactivation, all samples were pooled and precipitated (we pooled 96 samples, which included 80 samples for other studies). The pooled DNA was then size-selected for 250–500–base pair (bp) inserts with the Pippin Prep method (Sage Science). The sample was then subjected to 20 cycles of polymerase chain reaction (PCR) with a biotinylated P5 primer and an indexed P7 reverse primer. PstI-MluCI–only fragments were then enriched using Dynabeads M-280 streptavidin magnetic beads (ThermoFisher). The resulting elutants were further subjected to eight cycles of PCR using the P5 and P7 primers. Final libraries were quantified by quantitative PCR (Kapa Biosystems) and sequenced on one lane of Illumina HiSeq 2500 with version 4 chemistry and the 125-bp paired-end recipe. Sequence cluster identification, quality prefiltering, base calling, and uncertainty assessment were done in real time using Illumina's HCS 2.2.38 and RTA 1.18.61 software with default parameter settings. Three metrics were examined for quality control of the sequences upon sample demultiplexing: (1) demultiplexing balance, (2) rate of prefiltering, and (3) the presence of technical adapter sequence at the ends of reads using cutadapt v.1.0 (Martin, 2011). Sequence quality for each specimen was assessed with FastQC (Andrews, 2010); the mean sequence quality score (Phred Score) was 37, and all reads passed the per base sequence quality.

Sequences were analyzed with Stacks v.1.24 (Catchen et al., 2013) to identify SNPs. Stacks does not directly support paired-end data, so each pair was combined into a single long read (sequences were not trimmed). Sequences were confirmed to have the enzyme site (AATT) present, or they were discarded. Unless otherwise noted, the Stacks pipeline programs were run with the default settings. The "ustacks" program was used to combine the reads from each sample into stacks. Considering the long read length, the maximum distance allowed between stacks was increased from 2 to 5 (-M 5) to ensure that similar regions were grouped together. The "cstacks" program was run to combine stacks from all samples into a catalog. Two mismatches were allowed (-n 2) when merging stacks. The "sstacks" program was run to compare each sample to the catalog. The populations program in Stacks was used to estimate the number of SNPs present after stacks were filtered to require that (1) each stack was represented by at least 50% of the individuals (-r 50); (2) each stack was represented by at least five reads per individual (-m 5); and (3) each SNP was present in at least 5% of the individuals (minimum allele frequency cutoff = 0.05). In a separate analysis, we also enforced the option (–write_single_snp) to record only one (i.e., the first one) SNP per stack.

Results and discussion.—Thousands of SNPs were identified in the genome of the Kemp's ridley using the ddRAD-seq technique. Based on the criteria used, 88,544 different stacks were retained, corresponding to a combined 22,224,544 bp, in which 35,252 SNPs were identified; thus, 0.16% of the sites were polymorphic. For comparison, the percentage of SNPs in the genomes of the giant panda and the Hawaii amakihi are 0.58% (Zhao et al., 2013) and 0.39% (Callicrate et al., 2014), respectively. Limiting to count only the first SNP per polymorphic stack, 23,696 SNPs were calculated. This number of markers constitutes a remarkable improvement over the limited number of markers that were available for this species. Addition of individuals in future studies will likely uncover more SNPs. Although further filtering in downstream analy-

ses might reduce the number of SNPs (e.g., Larson et al., 2014), a large number of SNPs is expected to be retained. The SNPs generated by the ddRAD-seq technique herein should enable robust population genomics inferences for the Kemp's ridley, important for its conservation and management, such as current genetic variability, signatures of past bottlenecks, effective population size, number of breeders per reproductive season, levels of population differentiation, and genetic identification of individual nesters.

Estimates of the current genetic variability are necessary for assessing the genetic health (e.g., levels of inbreeding) and evolutionary potential (i.e., ability to adapt in response to environmental change) of the Kemp's ridley. SNP genetic analyses may also reveal signatures of past bottlenecks (Hung et al., 2014) in the genome of this turtle. Given its recent history, it is likely that this species harbors very low levels of genetic diversity and shows signatures of recent bottlenecks. In addition, although the Kemp's ridley is a native nester at PAIS, an effort was undertaken to increase nesting there through translocation of thousands of eggs from Rancho Nuevo during the 1970s and 1980s, with the goal of enhancing the survival of the species should a catastrophe strike the primary nesting beach (Shaver and Wibbels, 2007; Shaver and Caillouet, 2015). Therefore, it is important to determine the degree to which turtles nesting and hatching at PAIS reflect the genetic diversity at Rancho Nuevo. Estimation of current levels of genetic diversity for the Kemp's ridley will provide important baseline information with which to monitor temporal changes, which will be useful to assessing the impact of drastic population reductions.

Because the Kemp's ridley is a long-lived iteroparous species, adult females collected during a nesting season likely represent multiple cohorts. Therefore, inferences of effective population size will yield an estimate across many generations, which will in turn provide a historical perspective relevant to understanding long-term evolutionary processes (Waples et al., 2014). On the other hand, genotyping of dead embryos and hatchlings from the same nesting season will provide estimates of the number of breeders that contributed to that specific season, which is important for understanding eco-evolutionary dynamics and mating systems (Waples et al., 2014) and will offer important baseline data for comparison among years. It is likely that previous bottlenecks have reduced the effective population size of this species. Similarly, recent reductions in the number of nests suggest a reduced number of breeders. Estimating the number of breeders per season is important for monitoring changes in the population.

SNP genomic data can also be used to assess levels of population differentiation, a necessary step to identifying conservation and/or management units. Although the Kemp's ridley is presumed to consist of a single panmictic population, this hypothesis has not been tested. Its current management as one "Regional Management Unit" (Wallace et al., 2010) would require reconsideration if genetic evidence for population subdivision is discovered. The existence of two different migratory routes from nesting sites revealed by satellite tracking of individuals (Shaver et al., 2015), differences in nesting behavior [e.g., arribadas vs solitary nesters (Bernardo and Plotkin, 2007)], and/or different nesting locations could be indicative of population subdivision.

Finally, the availability of a large number of genetic markers might enhance genetic census techniques and allow us to infer the number of Kemp's ridley nesting females at particular localities. A recent study (Frey et al., 2014) that used a 10-microsatellite panel provided more accurate estimations of the numbers of nesting females in the Texas coast than do current census techniques (i.e., counting nests or counting nesting females), because nesting female encounters are limited as a result of rapid nesting by this species and lack of coverage by workers. Genetic identification of nesters across years will enable a better estimation of population parameters (e.g., number of nests per female, internesting interval), which should enhance demographic inferences (Frey et al., 2014).

The ddRAD-seq technique constitutes a powerful tool for obtaining the above important information. We plan to use this technique in a considerable number of Kemp's ridley samples that we have collected from 2010 (the year of the BP Deepwater Horizon oil spill), 2014, and 2015. We will continue sampling this turtle in Tamaulipas, Veracruz, and Texas during subsequent years. These samples, in conjunction with the ddRAD-seq technique, provide a remarkable opportunity for conservation genomic studies in the Kemp's ridley, which will in turn allow us to gather important baseline information for long-term monitoring of this critically endangered sea turtle. Given the density of SNPs

observed in the Kemp's ridley, it is likely that the ddRAD-seq technique will be also useful for conservation genetics research in other sea turtles. In addition, we have identified microsatellites in the ddRAD sequences of the Kemp's ridley and are in the process of optimizing their PCR amplifications. Development of such microsatellite loci will generate additional markers for conservation genetics research of the Kemp's ridley and possibly of other sea turtles.

Acknowledgments.—Support for genetic analyses was provided by Texas Sea Grant (to LAH and XRR) by an institutional grant (NA06OAR4170076) to the Texas Sea Grant College Program from the National Sea Grant Office, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, and by the Instituto Politécnico Nacional (grant SIP-20110067 to MRL). The National Park Service (NPS) provided funding for nesting detection and documentation and sample collection at North Padre Island, including PAIS. We thank Dimitra Guerrero, Rosalie Rossi, Cynthia Rubio, Jennifer Shelby Walker, and others who aided with collection and shipment of samples. Work by PAIS personnel was authorized under FWS Permit TE840727-3, Texas Parks and Wildlife Department Scientific Permit SPR-0190-122, and NPS Institutional Animal Care Protocols NPS IACUC 2011-15. This work was also conducted under the Animal User Protocol #2014-0037 from the TAMU Institutional Animal Care and Use Committee.

Literature Cited

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

Bernardo, J., and P. T. Plotkin. 2007. An evolutionary perspective of the arribada phenomenon and reproductive behavioral polymorphism of Olive Ridley sea turtles (*Lepidochelys olivacea*), p. 59–87. *In:* Biology and conservation of Ridley sea turtles. P. T. Plotkin (ed.). Johns Hopkins Univ. Press, Baltimore.

Callicrate, T., R. Dikow, J. W. Thomas, J. C. Mullikin, E. D. Jarvis, and R. C. Fleischer. 2014. Genomic resources for the endangered Hawaiian honeycreepers. BMC Genomics 15:1098.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22:3124–3140.

FitzSimmons, N. N., C. Moritz, and S. S. Moore. 1995. Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. Mol. Biol. Evol. 12:432–440.

Frey, A., P. H. Dutton, D. J. Shaver, J. S. Walker, and C. Rubio. 2014. Kemp's ridley *Lepidochelys kempii* nesting abundance in Texas, USA: a novel approach using genetics to improve population census. Endang. Species Res. 23:63–71.

Heppell, S. S., P. M. Burchfield, and L. J. Peña. 2007. Kemp's ridley recovery: how far have we come and where are we headed?, p. 325–335. *In:* Biology and conservation of Ridley sea turtles. P. T. Plotkin (ed.). Johns Hopkins Univ. Press, Baltimore.

Hung, C. M., P. J. Shaner, R. M. Zink, W. C. Liu, T. C. Chu, W. S. Huang, and S. H. Li. 2014. Drastic population fluctuations explain the rapid extinction of the passenger pigeon. Proc. Natl. Acad. Sci. USA 111:10636–10641.

Kichler, K. 1996. Microsatellites and conservation genetics: genetic variability and mating behavior of the Kemp's ridley sea turtle, *Lepidochelys kempii*. M.Sc. thesis, Texas A&M Univ., College Station, TX.

———, M. Holder, S. Davis, R. Marquez-M, and D. Owens. 1999. Detection of multiple paternity in the Kemp's ridley sea turtle with limited sampling. Mol. Ecol. 8:819–830.

Lara Rivera, A. L. 2012. Evaluación de la diversidad genética de la tortuga lora (*Lepidochelys kempii*) en Tamaulipas. M.Sc. thesis, Centro de Biotecnología Genómica, Instituto Politécnico Nacional Reynosa, Tamaulipas, Mexico. [In Spanish.]

Larson, W. A., L. W. Seeb, M. V. Everett, R. K. Waples, W. D. Templin, and J. E. Seeb. 2014. Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). Evol. Appl. 7:355–369.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17:10–12.

Narum, S. R., C. A. Buerkle, J. W. Davey, M. R. Miller, and P. A. Hohenlohe. 2013. Genotyping-by-sequencing in ecological and conservation genomics. Mol. Ecol. 22:2841–2847.

National Marine Fisheries Service and U.S. Fish and Wildlife Service (NMFS and USFWS). 2015. Kemp's ridley sea turtle (*Lepidochelys kempii*) 5-year review: summary and evaluation. Available from http://www.nmfs.noaa.gov/pr/listing/final_july_2015_kemp_s_5_year_review.pdf

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7:e37135.

Plotkin, P., and J. Bernardo. 2014. Sea turtle funding dries up. Science 343:484.

Sarti, L. 2014. Kemp's ridley nesting in Tamaulipas, Mexico—Rancho Nuevo, Barra del Tordo, Altamira, and Ciudad Madero. *In:* Second International Kemp's Ridley Symposium, November 18–19, 2014, Brownsville, TX.

Shaver, D. J., and C. W. J. Caillouet. 2015. Reintroduction of Kemp's ridley (*Lepidochelys kempii*) sea turtle to Padre Island National Seashore, Texas and its connection to head-starting. Herpetol. Conserv. Biol. 10:378–435.

———, K. M. Hart, I. Fukisaki, C. Rubio, A.R. Sartain-Ivreson, J. Peña, D. Gomez Gamez, R. D. J. Gonzalez-Diaz-Miron, P.M. Burchfield, H.J. Martinez, and J. Ortiz. 2016. Migratory corridors of adult female Kemp's ridley turtles in the Gulf of Mexico. Biological Conservation 194(2016):158–167.

———, C. Rubio, J. S. Walker, J. George, A. F. Amos, K. Reich, C. Jones, and T. Shearer. 2016. Kemp's ridley (*Lepidochelys kempii*) nesting on the Texas coast: geographic, temporal and demographic trends through 2014. Gulf Mex. Sci. 33:158–178.

———, and T. Wibbels. 2007. Head-starting the Kemp's ridley sea turtle, p. 297–323. *In:* Biology and conservation of Ridley sea turtles. P. T. Plotkin (ed.). Johns Hopkins Univ. Press, Baltimore.

Stephens, S. H. 2003. Genetic analysis of the Kemp's ridley sea turtle (*Lepidochelys kempii*) with estimates of effective population size. M.Sc. thesis, Texas A&M Univ., College Station, TX.

Wallace, B. P., A. D. DiMatteo, B. J. Hurley, E. M. Finkbeiner, A. B. Bolten, M. Y. Chaloupka, B. J. Hutchinson, F. A. Abreu-Grobois, D. Amorocho, K. A. Bjorndal, J. Bourjea, B. W. Bowen, R. Briseño Dueñas, P. Casale, B. C. Choudhury, A. Costa, P. H. Dutton, A. Fallabrino, A. Girard, M. Girondot, M. H. Godfrey, M. Hamann, M. López-Mendilaharsu, M. A. Marcovaldi, J. A. Mortimer, J. A. Musick, R. Nel, N. J. Pilcher, J. A. Seminoff, S. Troëng, B. Witherington, R. B. Mast 2010. Regional management units for marine turtles: a novel framework for prioritizing conservation and research across multiple scales. PLoS ONE 5:e15465.

Waples, R. S., T. Antao, and G. Luikart. 2014. Effects of overlapping generations on linkage disequilibrium estimates of effective population size. Genetics 197:769–780.

Zhao, S., P. Zheng, S. Dong, X. Zhan, Q. Wu, X. Guo, Y. Hu, W. He, S. Zhang, and W. Fan. 2013. Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. Nat. Genet. 45:67–71.

Luis A. Hurtado, Xochitl De La Rosa-Reyna, Mariana Mateos, Donna J. Shaver, Richard Metz, John Thiltges, Joshua Hill, Miguel A. Reyes-Lopez, and Charlie Johnson, *(LAH, XRR, MM) Department of Wildlife and Fisheries Sciences, Texas A&M University, College Station, Texas 77843-2258; (DJS) Division of Sea Turtle Science and Recovery, Padre Island National Seashore, Corpus Christi, Texas 78480-1300; (RM, JT, JH, CJ) TAMU Agrilife Genomics and Bioinformatics Services; (MRL) Centro de Biotecnología Genómica, Instituto Politécnico Nacional, Ciudad Reynosa, Tamaulipas, México, C.P. 88710.*