A Versatile Rapture (RAD-Capture) Platform for Genotyping Marine Turtles

Lisa Komoroske[1,2], Michael Miller[3], Sean O'Rourke[3], Kelly R. Stewart[2,4], Michael P. Jensen[2] and Peter H. Dutton[2]

[1] Department of Environmental Conservation, University of Massachusetts Amherst, Amherst, MA 01003, USA

[2] Marine Mammal and Turtle Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, La Jolla, CA, 92037, USA

[3] Department of Animal Science, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA

[4] The Ocean Foundation, Washington, DC, United States

Keywords: conservation genetics, reptiles, sea turtle, Rapture, RAD-Seq, high-throughput sequencing

Running Title: *Marine Turtle Rapture*

Abstract

Advances in high-throughput sequencing (HTS) technologies coupled with increased interdisciplinary collaboration is rapidly expanding capacity in the scope and scale of wildlife genetic studies. While existing HTS methods can be directly applied to address some evolutionary and ecological questions, certain research goals necessitate tailoring methods to specific study organisms, such as high-throughput genotyping of the same loci that are comparable over large spatial and temporal scales. These needs are particularly common for studies of highly mobile species of conservation concern like marine turtles, where life history traits, limited financial resources and other constraints require affordable, adaptable methods for HTS genotyping to meet a variety of study goals. Here, we present a versatile marine turtle HTS targeted enrichment platform adapted from the recently developed Rapture (RAD-Capture) method specifically designed to meet these research needs. Our results demonstrate consistent enrichment of targeted regions throughout the genome and discovery of candidate variants in all species examined for use in various conservation genetics applications. Accurate species identification confirmed the ability of our platform to genotype over 1,000 multiplexed samples, and identified areas for future methodological improvement such as optimization for low initial concentration samples. Finally, analyses within green turtles supported the ability of this platform to identify informative SNPs for stock structure, population assignment and other applications over a broad geographic range of interest to management. This platform provides an additional tool for marine turtle genetic studies and broadens capacity for future large-scale initiatives such as collaborative global marine turtle genetic databases.

Introduction

Marine turtles are migratory, long-lived megafauna of conservation concern, with populations of all species classified in high risk categories on the IUCN Red List of Threatened Species (IUCN 2017). The complex behaviors and life history traits marine turtles exhibit can make them highly susceptible to human impacts, while also posing challenges to understanding critical aspects of their biology required for their conservation (Wyneken *et al.* 2013). Over the past several decades, genetic approaches have provided key insight to important research questions in marine turtle biology and conservation, including natal homing to breeding grounds, connectivity between distant foraging grounds and nesting beaches, delineation of broad stocks and distinct population segments (DPS) for management (ESA 1973), and quantifying proportional impacts of fisheries across populations (reviewed in Jensen *et al.* 2013; Komoroske *et al.* 2017). Yet despite this progress, a diversity of unresolved research questions persist (Rees *et al.* 2016), many of which are well-suited to being addressed with emerging genetic and genomic approaches.

Genomic technological capabilities, especially high-throughput technologies (HTS), have rapidly expanded over the past decade to tackle a broader variety of questions in ecology and evolution (Ekblom & Galindo 2011; Ellegren 2014; Romiguier *et al.* 2014). Whole genome sequencing (WGS) and reduced representation approaches (such as targeted enrichment, transcriptome and restriction-site associated nuclear DNA sequencing; RNA-Seq and RAD-Seq, respectively) are becoming increasingly common with the continued decline in HTS costs and improvement of reference genome availability (Andrews *et al.* 2016; De Wit *et al.* 2015; Jones & Good 2016; Genome 10K 2009; Todd *et al.* 2016). However, resource development and applications in some taxa, especially many of conservation concern, have lagged behind others (Shafer *et al.* 2015; Garner *et al.* 2016). This is true for marine turtles and other non-mammalian vertebrates, highlighted by the fact that mammals comprise only 8% of the total number of vertebrate species, but represent over 70% of existing vertebrate genomes currently on *Ensembl* (Flicek *et al.* 2014). This has been in part due to limited resources and logistical constraints sampling animals with protected status and complex life histories, but also because these approaches are not compatible or cost effective with some of the highest priority research needs for these species. For example, WGS or reduced representation approaches that can be directly applied with little to no *a priori* genomic resources (RNA- and RAD-Seq) are well suited to address some research topics like phylogenomics and adaptive variation (Jarvis *et al.* 2014; Prince *et al.* 2017). However, other methods are needed for

studies that necessitate background knowledge and tailoring approaches to yield informative variants (particularly single nucleotide polymorphism; SNPs) for specific study organisms and goals, such as research requiring cost-effective high-throughput genotyping data that are comparable over large spatial or temporal scales. This latter scenario is common in conservation research (Hunter *et al.* 2018) and monitoring of wide-ranging, long-lived species such as marine turtles, where samples often need to be compared across regions, continents and generations, such as fisheries bycatch DPS assignment and genetic capture-recapture studies (Komoroske *et al.* 2017; Shamblin *et al.* 2017; Stewart *et al.* 2016).

Several methods have recently emerged to meet these needs, including Genotyping-in-Thousands by sequencing (GT-Seq; Campbell *et al.* 2015), Rapture (RAD-Capture; Ali *et al.* 2016), and microhaplotypes (an adaptation of GT-Seq; Baetscher *et al.* 2017). Each of these approaches has demonstrated utility and strong potential for future broader application in conservation research under different study objectives and contexts. Marine turtle conservation researchers frequently encounter needs to genotype samples for different species, sample quantities, numbers of loci (e.g., for stock structure vs. relatedness studies), yet have limited time and financial resources to develop informative markers tailored to each study goal. Additionally, despite being one of the largest and most threatened vertebrate groups (Shaffer *et al.* 2015), there are currently limited reference genomes or transcriptomes for non-avian reptiles in general (but see Tzika *et al.* 2015; Shaffer *et al.* 2013; Wang *et al.* 2013), making it challenging to identify informative SNP loci *a priori* from existing genomic resources. Finally, researchers often deal with samples of varying tissue types, storage conditions, quality and quantity due to field, resource, and permitting and other limitations (e.g., samples from decomposing stranded animals, limited refrigeration in tropical study sites, and international CITES and shipping regulations). Thus, while no one approach provides an *a priori* solution to all of these research needs, we sought out to develop a robust, flexible platform that could be employed across a variety of research projects by adapting the Rapture method developed by Ali et al. (2016). In particular, we leveraged an existing molecular collection to test the utility of our approach with samples spanning the conditions frequently encountered in marine turtle research and combined initial RAD-Seq with Rapture target design to achieve this without *a priori* knowledge of good candidate regions. Here, we present our results and highlight the strengths, limitations, and future applications of this platform and general approach in marine turtle biology and conservation research.

Materials and Methods

Sample Selection, Processing and RAD-Sequencing

We selected 96 samples from the national Marine Mammal and Sea Turtle Research Collection (MMASTR) housed at NOAA Southwest Fisheries Science Center (La Jolla, CA) that collectively were representative of the genetic diversity among and within global leatherback populations. Samples were collected from 1988-2016, including nesting females, adult males, hatchlings (sex undetermined), as well as in-water foraging, stranded and bycaught animals of both sexes. Sample selection was weighted toward Pacific leatherbacks to contribute to a complementary project investigating fine-scale population structure in the Pacific. Tissue samples (skin, blood or muscle) were preserved in saturated salt when available, shipped, and stored in the NOAA-National Marine Fisheries Service MMASTR Collection at -20°C. Genomic DNA (gDNA) was isolated from sub-samples of tissue using one of the following standard extraction techniques: phenol/chloroform (Sambrook *et al.* 1989), sodium chloride (Miller *et al.* 1988), a modified DNeasy Qiagen extraction kit (Qiagen, Valencia, California), or Qiagen reagents on a Corbett CAS-1200 extraction robot (Corbett Robotics, San Francisco, California) or PerkinElmer JANUS robot (Waltham, MA). After extraction, gDNA was stored at -80°C until use in downstream analyses. All candidate samples were checked for DNA quantity and quality via Qubit Fluorometry (Thermo Fisher Scientific, Waltham, MA) and a 4200 TapeStation System (Agilent, Santa Clara, CA), respectively. Samples with adequate concentrations and the best quality (i.e., high molecular weight) were normalized and included in the final sample set for each location. Libraries were prepared following the updated RAD protocol as described in Ali et al. (2016) using *SbfI*-HF and NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA) and sequenced at UC Davis Genomics Core Facility for paired-end 100 bp reads in 25% of a lane on an Illumina HiSeq 3000 instrument.

RAD Data Analysis & Capture Target Design

We demultiplexed samples by assigning reads with complete matching barcodes (Ali *et al.* 2016) and assessed raw sequence data quality with FASTQC (Andrews 2010). The leatherback turtle genome has not yet been assembled, and the green turtle is the closest related species with reference genome. Although divergence of the *Dermochelidae* - *Cheloniidae* families is estimated at approximately 100 million years before present (Duchene *et al.* 2012), given the evidence for slower rates of DNA

evolution among turtles relative to many other vertebrates (Avise *et al.* 1992) and the potential benefits of using a common reference genome relative to *de novo* assembly for our project goals, we aligned the leatherback RAD data to the green turtle genome (Wang *et al.* 2013) with the Burrows-Wheeler Aligner (BWA v0.7.5; Li & Durbin 2009) and evaluated mapping performance. We used *SAMtools* (v1.3; Li *et al.* 2009) to sort, filter for proper pairs and index alignments, remove PCR duplicates, and calculate summary statistics. After observing high mapping success (see results), we proceeded using these alignments to identify candidate SNPs and cross-species Rapture target loci. In brief, we employed a *SAMtools* genotype likelihood model in the program *ANGSD* (Korneliussen *et al.* 2014; Nielsen *et al.* 2012) to infer major and minor alleles and minor allele frequencies (MAF) for sites with data for at least one individual, mapping quality score ≥10 and base quality score ≥20. Specifically, we inferred major and minor alleles and estimated MAF using genotype likelihoods with a fixed major allele and unknown minor allele (Kim *et al.* 2011), adapted with an expectation-maximization algorithm as implemented in *ANGSD*. We then identified good candidate regions for targeted enrichment as regions with consistent coverage (~84 bp length), paired both up and downstream of an identified restriction site in a high proportion of total individuals (≥68% for all samples; ≥80% for Pacific leatherbacks only), and without any suspected polymorphisms within the restriction site or unknown nucleotide identity (*N*) in the reference sequence. Within regions that passed these criteria, we then randomly selected one of the paired regions (i.e., either up- or downstream of the restriction site) and created candidate lists for two target types: (1) potential candidate SNP loci (MAF ≥0.1≤0.4, allowing only one variable site within 150bp from the restriction site; preferentially including those with a SNP within the first 84bp), and (2) no additional filters, to serve as a random locus set for unbiased genome representation within and across marine turtle species. We used corresponding sequences from the green turtle genome to design a custom MYBaits in-solution DNA target enrichment kit set (120bp baits, Arbor Biosciences, formerly MYcroarray Inc., Ann Arbor, MI) with ~1000 targets for each of the two categories (2007 targets total) according to manufacturer protocols and quality control filters (e.g., probe compatibility, repeat masking, and melting temperature filters) with minor modifications to address initial failure of higher GC content baits (see below and Appendix S1 for details).

Rapture Sample Selection, Library Preparation & Sequencing

We selected DNA samples from the MMASTR collection encompassing a cross section of covariates to examine the versatility of this method for the varied conditions frequently encountered in our studies (e.g., sample location, sex, life stage, collection method, tissue type, DNA concentration, DNA quality and collection year; 1342 samples total). In particular, we included samples with detectable concentrations at or below 5 ng/ul, which are frequently encountered in minimally invasive sampling of sensitive wildlife species, but below typical recommended concentrations for many reduced representation genome protocols. Although sample selection was again weighted toward leatherbacks for a complementary study, samples from six of the seven extant sea turtle species were included to evaluate target enrichment success across species and geographic regions, as well as green turtle samples representative of all currently defined global distinct population segments (DPS; Seminoff *et al.* 2015) to confirm the consistency of these genome-wide markers with established management delineations. We prepared RAD libraries as described above (Ali *et al.* 2016; 16 libraries total), with the modification of including samples with initial gDNA concentrations across the range frequently obtained from wild marine turtle samples (i.e., not selecting higher concentration samples only). A total gDNA of 50 ng was targeted as starting material for each library across all samples with a maximum input volume of 10 ul (i.e., samples with initial concentrations < 5 ng/ul had lower starting input). We quantified and normalized libraries, followed by targeted enrichment following manufacturer's protocols, with the exception of doubling the capture reaction to include all RAD libraries (i.e., ~1/8 capture reaction per RAD library). During amplification steps in RAD library and capture enrichment protocols, we estimated the minimum number of PCR cycles required for each library to minimize PCR clones.

The library enrichment process described above was conducted in two replicate trials after results from the first trial indicated a strong effect of GC bait content on enrichment success (Figure S1). After confirming with the manufacturer that our probe design met all quality control standards, a new, exact replicate MyBaits kit was synthesized. Library enrichment was repeated on the same RAD libraries with the new kit for Trial 2, along with minor amendments recommended by MYcroarray, Inc. to the original manufacturer protocol. For both trials, enriched libraries were combined and sequenced at the UC Davis Genomics Core Facility on an Illumina HiSeq 3000 instrument in a full lane (Trial 1: paired-end 100-bp reads, Trial 2: paired-end 150-bp reads). Here, except where specified, we focus on results from analyses of Trial 2 data only. However, we include a semi-quantitative comparison between the two trials with regards to on-target coverage to

emphasize the importance of these technical details to inform effective MYBait design and application in future projects.

Rapture Data Quality Assessment & Analyses

We demultiplexed samples as described above and assessed assignment error by quantifying the absolute and proportional number of raw reads (1) assigned to unused Illumina indexes or blanks (i.e., staggered wells without DNA within each plate/library) or (2) had barcodes on both forward and reverse reads. We assessed sequence data quality with *FASTQC* and *MultiQC* (Andrews 2010; Ewels *et al.* 2016), and calculated summary statistics in *R* (R Core Team 2016) to examine depth and evenness of coverage across predictor factors (e.g., library, species, tissue type, input concentration, sample location, and collection year). We used *BWA* and *SAMtools* as described above to map sequences and filter alignments. We qualitatively examined mapping quality using the *Integrative Genomics Viewer* (IGV; Robinson *et al.* 2011) and quantitatively assessed by locus and sample coverage at a representative position within target regions (relative position 20) with *Bedtools* (Quinlan & Hall 2010) and *R*. We combined information from raw read distributions and target loci coverage to establish quality (success/failure) thresholds, and only samples that passed these thresholds were included in subsequent data analyses. To quantify rates of on-target capture, we mapped forward reads to a reference of target loci only using the same pipeline described above with the exception of omitting PCR duplicate removal.

To examine and compare the success of our approach to generate SNPs within and across species and populations informative for various genotyping applications, we conducted SNP discovery, inferred major and minor alleles, and estimated allele frequencies for variable sites using *ANGSD* (Korneliussen *et al.* 2014; Nielsen *et al.* 2012) on a series of sample sets: (1) all turtle samples, (2) hardshell (Cheloniid spp.) turtles only, (3) green turtles only, (4) all leatherback samples, and (5) a representative leatherback population. For each sample set, we employed a genotype likelihood model and applied quality filters similar to RAD data as described above, additionally only including samples that passed initial QC thresholds and alignments that were proper pairs and uniquely mapped. Polymorphic sites were identified and retained in downstream analyses only if there were data for at least 50% of individuals within the group being tested, MAF $\geq$0.05, and p-value of being variable $\leq$1e-6. To examine relationships of coverage and predictor variables with genotyping success at multiple stringency levels, we estimated genotype posterior probabilities for a

223 set of *a priori* candidate SNP positions (identified in RAD analysis described above) using an allele-
224 frequency based prior and called genotypes with threshold cutoffs of 80, 90, and 95%.
225
226 Species Confirmation & Population Structure Analyses
227 To validate our highly multiplexed approach, we first confirmed species identification with
228 principal components analyses (PCA) by generating a covariance matrix without calling genotypes
229 using the ngsCovar function in *ngsTools* (Fumagalli *et al.* 2014; Fumagalli *et al.* 2013) on all hardshell
230 turtles, including a small sample set of suspected hybrids (based on morphological characteristics).
231 To reduce influence of variance in depth of coverage between samples, we used *SAMtools* to
232 randomly subsample alignments at multiple thresholds to balance information and sample retention
233 in subsequent analyses (Ali *et al.* 2016). These analyses were also repeated including only less
234 represented groups in the total hardshell dataset (i.e., loggerhead, olive ridley and Kemp's ridley),
235 where the higher proportion of green turtle samples could obstruct distinguishing variation. We also
236 estimated admixture proportions of individuals using a maximum-likelihood-based clustering
237 algorithm with the program *NGSAdmix* (Skotte *et al.* 2013) and genetic distances for a representative
238 subset of samples across species and geographic regions using *ngsDist* (branch support based on
239 bootstrapping 1000 replicates with 500 SNP blocks; Vieira *et al.* 2016) and plotted as a tree with
240 *FastME* (BME iterative taxon addition method with NNI tree refinement; Lefort *et al.* 2015) and the
241 R packages *phanhorn* (Schliep 2011) and *ape* (Popescu *et al.* 2012).
242 Secondly, we included green turtle samples from nesting grounds over a geographic range of
243 interest to management in order to explore how our platform would perform delineating population
244 structure within species. Thus, our goal was to evaluate the utility of the identified SNPs with this
245 preliminary dataset to discern if they were likely to be informative markers in future, larger-scale
246 analyses of stock structure and population assignment. We employed methods described above for
247 PCA, admixture and genetic distances, and also estimated allele frequency spectra using *ANGSD*
248 and *realSFS* to calculate pairwise F_{ST} values. Although it is common to accompany F_{ST} estimation
249 with permutation tests to assess significant differences among the *a priori* defined groups, such
250 analyses would have limited confidence given the restricted group sample sizes in our exploratory
251 dataset, and are more suitable for future stock structure studies employing these markers with robust
252 sample sizes and comprehensive geographic coverage.

253 Finally, we also estimated allele frequency spectra to calculate genetic diversity statistics
254 (Watterson's estimator, θ_w, based on number of segregating sites, and Tajima's estimator, θ_π or π,
255 based on pairwise differences between sequences) in *ANGSD* and *realSFS* among species
256 (Korneliussen *et al.* 2014; Korneliussen *et al.* 2013; Tajima 1989; Watterson 1975). Unequal sample
257 sizes, population structure and upstream filtering for SNPs can cause biases in nucleotide diversity
258 estimations (Lozier 2014; Subramanian 2016; confirmed with subsampling simulations on this
259 dataset), potentially creating issues in our dataset with variable sample sizes across populations with
260 likely differing demographic histories and current status (e.g., recovering, declining, etc.). To address
261 this, we included only the random set of targeted loci as described above with selected subsets of 4-6
262 QC passed individuals from representative populations from each species, and report results on
263 semi-quantitative evaluation of descriptive statistics only. Thus, although inference from these
264 metrics is constrained, we include them demonstrate the utility of this platform for research
265 employing these metrics in robust sample sets within or across species.
266
267 **Results**
268 RAD-Sequencing & Rapture design
269 We recovered 95.7 million total raw sequences, and 89.0% of which were retained based on
270 sample assignment criteria. *FASTQC* confirmed consistent high sequence quality across the library
271 with no evidence of contamination. After removal of four failed samples (defined as <2% of average
272 number of sequences assigned to sample), an average of 93.9% (±7.3% S.D.) of sequences mapped
273 to the green turtle genome, an average of 51.2% (±4.1% S.D.) of which remained after filtering out
274 PCR clones. These results of strong concordance supported the use the green turtle genome as a
275 reference, so we proceeded using these alignments for further Rapture bait development. We
276 identified a total of 7,282 RAD tags with paired regions that met initial filtering criteria. A total of
277 1,379 of these candidate regions further met our SNP criteria (see methods) and were included in
278 bait design, as well as 1,400 additional randomly selected regions from this list. From these 2,779
279 final candidates, we were able to design a custom MYBaits kit that met MYcroarray's QC criteria
280 with 2,007 targets for Rapture genotyping in marine turtles.
281
282 Rapture data quality analysis

283 In Trial 2, we recovered 396 million total raw sequences, with only 0.38% of these sequences
284 removed due to assignment to unused Illumina indexes or the presence of barcodes on both
285 forward and reverse reads. *FASTQC* and *MultiQC* results confirmed high quality scores across and
286 within libraries and no issues of contamination. Assignment of raw sequences to blanks dispersed
287 across libraries was extremely low (average= 245, min/max=27/818). Based on sequence count
288 distributions, we determined an initial sample failure/success threshold of 10,000 raw sequences,
289 which 1127 samples passed (84%; hereafter referred to as 'QC passed samples'). Read counts varied
290 across library and samples, but we did not observe any clear patterns of success or failure between
291 input factors, particularly among species or DNA input. Samples more recently collected and with
292 higher DNA initial concentrations more consistently passed initial quality thresholds, but many low
293 concentration and older samples did as well.
294
295 Rapture target coverage and genotyping success
296 Samples exhibited very high percentages of mapping and on-target sequence capture, with
297 Trial 2 having even higher on-target success than Trial 1 (Fig. 1A & S1; see methods and Appendix
298 S1 for details). For Trial 2 data, mapped filtered (PCR clones removed) fragments for QC-passed
299 samples were an average of 20.8% (±6.9% S.D.) of the total sequenced fragments per individual,
300 and this was correlated with sample initial gDNA concentration (Fig. 1B). Average coverage per
301 locus in filtered QC-passed samples was 26.6 (±10.1 S.D.; min/max=0.9/99.1; see Fig. S2 for
302 coverage distributions). Samples generally reached ≥ 4x coverage across loci with approximately
303 50,000-75,000 filtered alignments (Fig. S3a). However, we identified samples that passed initial QC
304 thresholds, but had lowered numbers of filtered alignments and few Rapture loci covered at ≥ 4x
305 (Fig. S3b), prompting us to implement an additional filter of a minimum of 5,000 filtered alignments
306 in further downstream analyses. Of these new QC-passed samples (1097 total), we were able to
307 genotype over 50% of *a priori* identified SNPs in Rapture loci at all posterior probability thresholds
308 tested (Fig. 2a). Genotyping capacity increased with depth of coverage but began reaching saturation
309 at approximately 150,000 sequenced fragments per individual (depending on posterior probability
310 threshold and sample). However, genotyping capacity was also clearly affected by the relative
311 position of the SNP within the Rapture locus region (Fig. 2b), displaying a distinct break at
312 approximately relative position 100, despite the use of longer 150bp paired-end sequencing.
313

314 Cross Species Capture Success & SNP discovery

315 We observed consistent success in coverage of Rapture loci across all species tested,
316 confirming the broad utility of this approach for genotyping studies across marine turtle species. A
317 reduction in the maximum loci covered regardless of total depth of coverage was observed in non-
318 green hardshell turtle species (Fig. 3), indicating that a small percentage of selected targets in this
319 particular enrichment set are not useful for other hardshell species, likely due to polymorphisms in
320 *SbfI* restriction sites or other compatibility issues. Nevertheless, we identified ample candidate
321 polymorphic SNPs suitable for within-species genotyping studies (Table 1). However, we emphasize
322 that because SNP identification is inherently determined by analysis parameters and input sample
323 composition, determining informative SNPs within Rapture target regions should be conducted
324 using samples and filtering thresholds aligned with research goals to avoid ascertainment bias (e.g.,
325 demonstrated here by comparing SNP discovery results in all leatherback samples versus within one
326 specific population; Table 1).
327
328 Species Confirmation and Green Turtle Population Structure

329 Individuals strongly separated by species as expected in the first two PC components for all
330 hardshell species, with the exception of the two ridley species (Fig. 4a) that resolved in further PC
331 axes in the combined analysis, as well as separate analyses omitting green and hawksbill turtle
332 samples (Fig. 4b). Clear species separation was similarly observed in admixture proportion results,
333 but with even more pronounced effects of the unbalanced sample groups when all hardshell samples
334 were included (i.e., strong breaks in population structure within green turtles began to emerge before
335 the separation of the ridley species; Fig. 4c,d). Estimated genetic distances among species were
336 largest as expected between leatherbacks and hardshell turtles, followed by green turtles relative to
337 other hardshell species (loggerhead, hawksbill, Kemp's ridley, and olive ridley; Fig. S4). Several
338 hybrids were identified, including three green-loggerhead hybrids and one green-hawksbill hybrid,
339 however for several other suspected hybrids both PCA and admixture proportion results support
340 only genetic contributions from olive ridley.

341 In green turtles, pairwise F_{st} values, genetic distances and PCA discerned strong breaks in
342 population structure between major ocean regions aligned with previous studies based on mtDNA
343 and microsatellites and green turtle DPS designations (Jensen *et al.* in press; Seminoff *et al.* 2015;
344 Figs. 5 & S5; Table S1). Tree topology branch support of genetic distances as well as F_{st} values were

345 higher in the Atlantic compared to the Pacific Ocean. In the western Pacific, PCA clustering of
346 samples by location for several groups are congruent with potential finer-scale population structure
347 (Fig. S5b), further supporting the utility of these SNP markers for future stock structure and
348 population assignment studies.
349
350 Genetic Diversity Estimates
351 Patterns within groups were consistent between θ_w and π, and within species, with the
352 exception of Costa Rica hawksbills that had substantially higher values for both metrics (Fig. 6).
353 Generally, green turtles exhibited the highest nucleotide diversity, while leatherbacks displayed the
354 lowest. In particular, all four groups of Pacific leatherbacks had lower levels of variation relative to
355 the Atlantic population included (Brazil).
356
357 **Discussion**
358 Technological advances combined with increased interdisciplinary collaboration has rapidly
359 expanded both the scope and scale of genetic studies over the past decade, yet for many species of
360 conservation concern such as marine turtles, the realized potential of these advances is only just
361 beginning (Garner *et al.* 2016; Komoroske *et al.* 2017; Shafer *et al.* 2015). This is in part because life
362 history traits and protected status of these taxa can create unique research challenges, but also
363 because the resources required for method development (which often needed to be repeated to
364 generate informative markers tailored to each species and study goal) often has made it infeasible for
365 conservation researchers. Our results demonstrate that the adaptation of the Rapture method
366 developed by Ali et al. (2016) provides a flexible platform for marine turtle research. While
367 limitations and room for further improvement remain, the addition of our platform and general
368 approach to the marine turtle genetic toolbox opens the door to a diversity of rapid, cost-efficient
369 genotyping applications. These data can be comparable across laboratories, geographical regions,
370 and timescales, which can be particularly important in such highly mobile species that can migrate
371 across entire ocean basins and necessitate international collaboration for effective conservation
372 (Shamblin *et al.* 2014). Though our specific selected regions for targeted enrichment will not be
373 suitable for all populations or research questions, our study also demonstrates how initial RAD-
374 Sequencing can be used to develop a Rapture platform suited to specific research needs.
375 Additionally, these target regions can be adapted to other genotyping platforms that may be better

suited to meet some research needs but require prior knowledge of genomic variants, e.g., GT-Seq that may have improved performance on lower quality and concentrations samples (Campbell *et al.* 2015) or microhaplotypes that may provide increased power for relationship inference (Baetscher *et al.* 2017).

Our results highlight several key strengths of this platform in meeting the diverse needs of marine turtle genotyping applications. First, researchers often need to analyze few or many samples at few or many loci, depending on study goals. Our data demonstrate that samples can be combined and effectively genotyped at the same loci with moderate sequencing coverage using partial capture reactions. This not only facilitates cost-effective, time-efficient analysis of large sample sets, but also combining samples for different projects. For example, researchers working on large nesting beaches often have many samples to analyze at the end of the season (Shamblin *et al.* 2017), while those genotyping samples from fisheries bycaught animals or some foraging population assessment projects may have smaller sample sets collected intermittently over the year. In the latter case, it has been particularly problematic to determine how to move from manual analysis with traditional markers to next-generation sequencing approaches where much of the reduced cost and time efficiency is related to multiplexing and high-throughput processing. While genotyping high priority single samples that need to be analyzed in near real-time may still pose a challenge, the flexibility of the Rapture platform offers options to combine library preparation and sequencing across projects and species, or to create a libraries with fewer samples and reduce total sequencing depth (e.g., through the use of a lower output instrument such as an Illumina MiSeq, or coordinating with other researchers to use different library barcodes and share sequencing lanes). Additionally, we designed a custom MYBaits enrichment kit with ~2000 targets to satisfy the needs of a variety of study types, but this can be adapted to include fewer or more loci. For example, researchers interested in basic population structure and individual assignment may wish to design kits with a subset of only several hundred informative targets, increasing the per locus depth of coverage in each sample. Finally, the ability to repeatedly capture the same genomic regions facilitates studies conducted over broader time periods (e.g., examining trends across many nesting seasons or even generations) or spatial scales (e.g., collaborating labs can generate and share data between foraging and nesting grounds).

Despite these exciting opportunities, our data also clearly show that our current Rapture platform has some limitations that are relevant to situations frequently encountered in wildlife genetics studies. First, although we were able to effectively perform high on-target sequencing and

genotyping for samples across tissue types, DNA extraction methods, species, and other co-factors, a portion of our test samples failed to sequence well. Though no clear patterns emerged with sample age or molecular weight thresholds, it is likely that highly degraded or contaminated samples (e.g., due to natural conditions, collection and storage methods) were more likely to fail. While this problem is often easily circumvented in controlled experimental settings, in many conservation applications these issues can be unavoidable, such as working with museum collections or opportunistic sampling of animals that have had substantial exposure to natural elements post-mortem. However, we emphasize that many samples in our study that exhibited evidence of some degradation were successful, including those that fall into these sub-optimal categories (e.g., stranded and bycaught animals). Our results support the initial findings of Ali et al. (2016) that this new RAD protocol is more robust than previous RAD methods for partially degraded samples, but there may be a point beyond which it is not a suitable approach. However, it may be possible to generate comparable genotype data for these samples at a subset of informative Rapture loci with highly-multiplexed PCR based methods such as GT-Seq (Campbell *et al.* 2015) that amplify short DNA fragments and thus be more robust to sample degradation. Secondly, we observed a substantial proportion of sequenced fragments that were PCR clones, and this was correlated with initial sample DNA concentration. The latter observed effect may be a product of the increased influence of measurement and pipetting error at low concentrations, which could be targeted for improvement in a future protocol adaptation. However, since PCR clones are in effect wasted sequences, in practice this currently means that it is less cost effective to sequence samples with low initial DNA concentrations, and that calculations of required sequencing to attain a targeted depth of coverage must take these factors into account. Although sequencing costs are likely to continue to decrease such that genotyping can still be achieved despite this loss, future efforts to reduce clonality would improve the efficiency and cost of this approach. Finally, although costs and technological accessibility have vastly improved in recent years, access to the equipment and financial resources to conduct genetic studies is far from universally available. This makes continued collaboration essential to advancing our understanding of marine turtles, with researchers with access to such resources working to increase capacity elsewhere, such as through visiting scientist training partnerships and creation of shared genetic databases. Particularly given the influence that bioinformatics parameters (e.g., filtering criteria, assembly methodology, genotyping thresholds) can

437 have on results (O'Leary *et al.* 2018), it is imperative for researchers to include metadata and analysis
438 details to ensure robust and comparable data across laboratories and over time.
439 We present results of conducting SNP discovery independently for each species and within a
440 representative leatherback population to demonstrate that substantial variation exists within our
441 targeted regions to meet a variety of study goals, but also to highlight the importance of appropriate
442 test data and analyses parameter thresholds to avoid ascertainment bias (i.e., discerning informative
443 SNPs appropriate for a given study goal; Lachance & Tishkoff 2013). For example, intra-population
444 questions can require variable SNPs within a target population, which may not be identified in
445 broader analysis including many populations depending on filtering thresholds and sample sizes
446 (Andrews *et al.* 2018). One advantage to the flexible Rapture platform is that researchers can
447 generate data for many genomic regions and then hone in on informative SNPs to genotype without
448 *a priori* knowledge and the need to develop different markers tailored to each study goal, which can
449 be cost and time prohibitive. However, as discussed previously, if desired, researchers can also use
450 preliminary RAD or Rapture data with a representative test dataset to identify the most informative
451 markers for their study and design new MYBaits kit or GT-Seq primers to focus exclusively on
452 those targets.
453 Principal components and admixture proportion analyses identified clear separation of all
454 species examined and our tree depicting relationships among species was in general agreement with
455 previous research (Duchene *et al.* 2012; Naro-Maciel *et al.* 2008). It is important to note that these
456 studies were focused on resolving phylogenetic relationships among all marine turtle species, and
457 thus the methods employed were much more in-depth than our analyses; additionally, we were not
458 able to include any flatback turtle samples in our study. Thus, clarifying any discrepancies or further
459 confirmation using our genome-wide markers would require additional studies. However, for the
460 purpose of our primary study goals, since species were randomized across and within RAD libraries
461 and we observed low number of sequences assigned to blank wells, our results show that sequences
462 can be assigned correctly to individuals using this highly-multiplexed approach and our analyses
463 criteria. Cross-species targeted enrichment may not be as effective in other taxa with high genomic
464 diversity or for studies that require tens to hundreds of thousands of SNPs, and researchers working
465 with other species may wish to omit targets from our panels that only yielded coverage in green or
466 leatherback turtles.

We identified several hybrids, in agreement with preliminary evaluation of these samples with three nuclear loci and the mitochondrial control region (Dodge *et al.* 2006), though additional analyses with larger sample sizes from contributing species at the same locations would further validate these findings and provide insight into the prevalence of hybridization in these populations. Hybridization and complex introgression patterns have been previously documented, primarily in southeast Atlantic populations (Reis *et al.* 2010; Vilaça *et al.* 2012), but the frequency of such events elsewhere and the relative hybrid fitness is largely unknown. Given recent concern that increasingly skewed female-biased sex ratios due to climate change (Jensen *et al.* 2018) and other anthropogenic pressures (Gaos *et al.* 2018) could cause interspecies mating events to become more prevalent and further destabilize populations, additional research is needed to better understand these processes and monitor changes over time; our Rapture platform offers an additional tool for such studies

Our exploratory green turtle analyses determined that our platform can also successfully amplify targeted regions within species across broad geographic locations and identify informative SNPs for stock structure, population assignment and other management applications. A recent study of green turtle global phylogeography using mtDNA control region sequences identified eleven divergent lineages that each encompass a few to many genetically differentiated distinct management units (MUs) with more recent shared ancestry but deemed to be demographically independent (Jensen *et al.* in press). This comprehensive study builds on previous work within regions documenting restricted gene flow attributed to female natal philopatry and generally little genetic differentiation among nesting beaches within 500km (reviewed in Jensen *et al.* 2013; Jensen *et al.* in press; Komoroske *et al.* 2017). While instrumental for our understanding of green turtle evolutionary history and contemporary stock structure patterns, there is a clear need to complement this work with studies employing nuclear markers to identify the roles of male-mediated gene flow and higher marker resolution. With additional refinement of the SNPs identified here specifically to meet these goals (e.g., narrower filtering criteria to remove any biases due to physical linkage or inconsistent coverage), these markers will serve as a valuable resource for such studies over large spatial and temporal scales, further advancing our understanding of green turtle population connectivity, MU designation, and human impacts.

Finally, comparisons of genetic variation among populations and species can be informative for a variety of conservation relevant research, such as understanding how genetic diversity may differ among healthy, recovering, and declining populations (Lozier 2014). While our current sample set

498 was not designed to address these questions specifically, the ability to consistently amplify over a
499 thousand regions across the genome for all marine turtles, enables our platform can be effectively
500 employed for such research goals within or across species. For example, we found that Pacific
501 leatherbacks exhibited the lowest levels of nucleotide diversity relative to all other groups evaluated,
502 including the (Atlantic) Brazilian nesting stock. While further robust analysis is needed to confirm
503 this preliminary finding, this could be related to the continued decline of Pacific leatherback
504 populations in contrast to Atlantic populations.

505 In conclusion, our Rapture platform provides a tool that is complementary to existing traditional
506 genetic markers as well as other emerging genomic techniques suited to address a broad diversity of
507 research questions in marine turtle ecology, evolution and conservation (e.g., transcriptome, other
508 reduced representation, and whole genome sequencing to study adaptive variation and genome-
509 phenome linkages). Though some limitations still hinder widespread adoption of these techniques,
510 such as cost and well-assembled and annotated genomic resources, as technologies continue to
511 advance we anticipate continued application and creative adaptations to meet the challenging needs
512 of conservation researchers. If realized, this could generate capacity for large-scale initiatives such as
513 the creation of global genetic databases akin to those that have begun emerging recently for other
514 taxa (e.g., Deck *et al.* 2017). This would not only expand the scope of research questions that can be
515 investigated, but also provide traditionally resource-limited marine turtle programs with the ability to
516 incorporate genetic information in their research and monitoring efforts. Such endeavors will
517 inevitably present many new challenges, but the successes of analogous initiatives such as the State
518 of the World's Sea Turtles (SWOT) and the Atlantic-Mediterranean Loggerhead Genetics (LGWG;
519 Shamblin *et al.* 2014) working groups among others have demonstrated the power of such global
520 collaborative efforts to answer the major outstanding research questions in these wide-ranging,
521 complex megafauna.
522
523 **Acknowledgements**
524 We would like to thank members of the Marine Mammal and Turtle Division for assistance in the
525 laboratory, logistics and project design suggestions, especially A. Frey, P. Morin, V. Pease, E.
526 LaCasella, A. Lanci, G. Serra-Valente and S. Roden. Funding was provided by the National Oceanic
527 and Atmospheric Administration, Southwest Fisheries Science Center, the National Research
528 Council of the National Academies of Science, Engineering and Medicine to LMK, through a

529 Lenfest Ocean Program grant to KRS (the views expressed are those of the authors and do not
530 necessarily reflect the views of the Lenfest Ocean Program or The Pew Charitable Trusts), and The
531 Sea Turtle Census Initiative at The Ocean Foundation. We would like to thank T. Summers, D.
532 Graff, A. Tagarino, N. Fitzsimmons, A. Gaos, M. Liles, K. Dodge, H. Guzman, N. Marcovaldi, J.
533 Thome, B. Bowen, J. Mortimer, E. Harrison and G. Balazs, Centro Tamar/ICMbio and Projeto
534 Tamar/Fundação Pro Tamar, The Sea Turtle Conservancy, American Samoa Department of Marine
535 and Wildlife Resources, CNMI Division of Fish and Wildlife & Department of Lands & Natural
536 Resources and NOAA-Fisheries Pacific Islands Science Center, for sample contributions.

Literature Cited

Ali O, O'Rourke S, Amish S, Meek M, Luikart G, Jeffres C, Miller M (2016) RAD Capture (Rapture): Flexible and Efficient Sequence-Based Genotyping. *Genetics* **202**, 389–400.

Andrews KR, Adams JR, Cassirer EF, Plowright RK, Gardner C, Dwire M, Hohenlohe PA, Waits LP (2018) A bioinformatic pipeline for identifying informative SNP panels for parentage assignment from RADseq data. *Mol Ecol Resour*. https://doi.org/10.1111/1755-0998.12910

Andrews KR, Good J, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*. **17**, 81–92.

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

Avise JC, Bowen BW, Lamb T, Meylan AB, Bermingham E (1992) Mitochondrial DNA evolution at a Turtle's Pace: evidence for low genetic variability and reduced microevolutionary rate in the testudines. *Mol Biol Evol*, **9**, 457–473.

Baetscher DS, Clemento AJ, Ng TC, Anderson EC, Garza JC (2017) Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. *Mol Ecol Resour*, **18**, 296-305.

Campbell N, Harmon S, Narum S (2015) Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Mol Ecol Resour*, **15**, 855-867.

De Wit P, Pespeni M, Palumbi S (2015) SNP genotyping and population genomics from expressed sequences –current advances and future possibilities. *Mol Ecol*, **24**, 2310–2323.

Deck J, Gaither M, Ewing R, Bird C, Davies N, Meyer C, …, Crandall ED (2017) The Genomic Observatories Metadatabase (GeOMe): A new repository for field and sampling event metadata associated with genetic samples. *PLoS Biol*, **15**, e2002925.

Dodge K, LeRoux R, Frey A, Dutton P (2006) Confirmation of marine turtle hybrids in northwest Atlantic waters. In: *Twenty Sixth Annual Symposium on Sea Turtle Biology and Conservation* (eds. Frick M, Panagopoulou A, Rees AF, Williams K), p. 376 pp. International Sea Turtle Society, Athens, Greece.

Duchene S, Frey A, Alfaro-Nunez A, Dutton PH, Gilbert MTP, Morin PA (2012) Marine turtle mitogenome phylogenetics and evolution. *Mol Phylogenetics Evol*, **65**, 241-250.

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, **107**, 1-15.

Ellegren H (2014) Genome sequencing and population genomics in non-model organisms. *TREE*, **29**, 51-63.

571 Endangered Species Act of 1973. U.S. Fish and Wildlife Service Home,
572 www.fws.gov/laws/lawsdigest/ESACT.html. Accessed August 24, 2018.

573 Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: Summarize analysis results for
574 multiple tools and samples in a single report. In: *Bioinformatics*. http://multiqc.info/

575 Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, …, Searle SM (2014) Ensembl 2014.
576 *Nucleic Acids Res*, **42**, D749-755.

577 Fumagalli M, Vieira F, Linderoth T, Nielsen R (2014) ngsTools: methods for population genetics
578 analyses from next-generation sequencing data. *Bioinformatics*, **30**, 1486-1487.

579 Fumagalli M, Vieira FG, Korneliussen TS, Linderoth T, Huerta-Sánchez E, Albrechtsen A, Nielsen
580 R (2013) Quantifying Population Genetic Differentiation from Next-Generation Sequencing Data.
581 *Genetics*, **195**, 979-992.

582 Gaos AR, Lewison RL, Liles MJ, Henriquez A, Chavarría S, Yañez IL, Stewart K, …, Dutton PH
583 (2018) Prevalence of polygyny in a critically endangered marine turtle population. *JEMBE,* **506**, 91-
584 99.

585 Garner B, Hand B, Amish S, Bernatchez L, Foster J, Miller K, …, Luikart G (2016) Genomics in
586 conservation: case studies and bridging the gap between data and application. *TREE*, **31**, 81-83.

587 Genome 10K Community of Scientists (2009) Genome 10K: a proposal to obtain whole-genome
588 sequence for 10,000 vertebrate species. *J Hered*, **100**, 659–674.

589 Hunter ME, Hoban SM, Bruford MW, Segelbacher G, Bernatchez L (2018) Next-generation
590 conservation genetics and biodiversity monitoring. *Evol Appl*, **11**, 1029-1034.

591 Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, …, Zhang G (2014) Whole-genome analyses
592 resolve early branches in the tree of life of modern birds. *Science*, **346**, 1320-1331.

593 Jensen M, FitzSimmons N, Dutton P (2013) Molecular Genetics of Sea Turtles. In:Wyneken J,
594 Lohmann K, Musick J (eds) Biology of Sea Turtles CRC Press. pp. 135-162.

595 Jensen MP, Allen CD, Eguchi T, Bell IP, LaCasella EL, Hilton WA, Hof CAM, Dutton PH (2018)
596 Environmental Warming and Feminization of One of the Largest Sea Turtle Populations in the
597 World. *Curr Biol*, **28**, 154-159.

598 Jensen MP, Fitzsimmons NN, Bourjea J, Hamabata T, Reece J, Dutton PH (in press*)* The
599 evolutionary history and global phylogeography of the green turtle *(Chelonia mydas). J Biogeogr.*

600 Jones M, Good J (2016) Targeted capture in evolutionary and ecological genomics. *Mol Ecol*, **25**, 185-
601 202.

602 Kim S, Lohmueller K, Albrechtsen A, Li Y, Korneliussen T, Tian G, …, Nielsen R (2011)
603 Estimation of allele frequency and association mapping using next-generation sequencing data. *BMC*
604 *Bioinformatics*, **12**, 231.

605 Komoroske LM, Jensen MP, Stewart KR, Shamblin BM, Dutton PH (2017) Advances in the
606 Application of Genetics in Marine Turtle Biology and Conservation. *Frontiers in Marine Science*, **4**, 156.

607 Korneliussen TS, Albrechtsen A, Nielsen R (2014) ANGSD: Analysis of Next Generation
608 Sequencing Data. *BMC Bioinformatics*, *15*, 356.

609 Korneliussen TS, Moltke I, Albrechtsen A, Nielsen R (2013) Calculation of Tajima's D and other
610 neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinformatics*, **14**, 289.

611 Lachance J, Tishkoff SA (2013) SNP ascertainment bias in population genetic analyses: why it is
612 important, and how to correct it. *Bioessays*, **35**, 780-786.

613 Lefort V, Desper R, Gascuel O (2015) FastME 2.0: A Comprehensive, Accurate, and Fast Distance-
614 Based Phylogeny Inference Program. *Mol Biol Evol*, **32**, 2798-2800.

615 Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform.
616 *Bioinformatics*, **25**, 1754-1760.

617 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, …, Durbin R, Subgroup GPDP
618 (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, **25**, 2078-2079.

619 Lozier JD (2014) Revisiting comparisons of genetic diversity in stable and declining species:
620 assessing genome-wide polymorphism in North American bumble bees using RAD sequencing. *Mol*
621 *Ecol*, **23**, 788-801.

622 Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from
623 human nucleated cells. *Nucl Acids Res*, **16**, 1215

624 Naro-Maciel E, Le M, FitzSimmons NN, Amato G (2008) Evolutionary relationships of marine
625 turtles: A molecular phylogeny based on nuclear and mitochondrial genes. *Mol Phylogenet Evol*, **49**,
626 659-662.

627 IUCN (2017) The IUCN Red List of Threatened Species. Version 2017-3.
628 http://www.iucnredlist.org. Accessed 05 December 2017.

629 Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J (2012) SNP Calling, Genotype Calling, and
630 Sample Allele Frequency Estimation from New-Generation Sequencing Data. *PLoS ONE*, **7**,
631 e37558.

632 O'Leary SJ, Puritz JB, Willis SC, Hollenbeck CM, Portnoy DS (2018) These aren't the loci you're
633 looking for: Principles of effective SNP filtering for molecular ecologists. *Mol Ecol*,
634 https://doi.org/10.1111/mec.14792.

635 Popescu A-A, Huber KT, Paradis E (2012) ape 3.0: New tools for distance-based phylogenetics and
636 evolutionary analysis in R. *Bioinformatics*, **28**, 1536-1537.

637 Prince DJ, O'Rourke SM, Thompson TQ, Ali OA, Lyman HS, Saglam IK, …, Miller MR (2017)
638 The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics
639 for informing conservation. *Science Advances*, **3**, e1603198.

640 Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features.
641 *Bioinformatics*, **26**, 841-842.

642 Rees A, Alfaro-Shigueto J, Barata P, Bjorndal K, Bolten AB, Bourjea J, …, Godley BJ. (2016) Are
643 we working towards global research priorities for management and conservation of sea turtles?
644 *Endang Species Res*, **31**, 337-382.

645 Reis EC, Soares LS, Lôbo-Hajdu G (2010) Evidence of olive ridley mitochondrial genome
646 introgression into loggerhead turtle rookeries of Sergipe, Brazil. *Conserv Genet*, **11**, 1587-1591.

647 Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011)
648 Integrative Genomics Viewer. *Nature Biotechnol*, **29**, 24-26.

649 Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, …, Galtier N (2014)
650 Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*,
651 **515**, 261-U243.

652 Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring
653 Harbour Laboratory Press, New York.

654 Schliep KP (2011) phangorn: phylogenetic analysis in R. *Bioinformatics*, **27**, 592-593.

655 Seminoff J, Allen C, Balazs G, Dutton P, Eguchi T, Haas H, ..., Waples RS (2015) Status Review of
656 the Green Turtle *(Chelonia mydas)* under the U.S. Endangered Species Act. NOAA Technical
657 Memorandum, NOAA-NMFS-SWFSC-539. pp. 1–571, Silver Spring, MD.

658 Shafer ABA, Wolf JBW, Alves PC, Bergström L, Bruford MW, Brännström I, …, Zieliski P (2015)
659 Genomics and the challenging translation into conservation practice. *TREE*, **30**, 78-87.

660 Shaffer H, Gidis M, McCartney-Melstad E, Neal K, Oyamaguchi H, Tellez M, Toffelmier E (2015)
661 Conservation genetics and genomics of amphibians and reptiles. *Annu Rev Anim Biosci*, **3**, 113–138.

662 Shaffer HB, Minx P, Warren DE, Shedlock AM, Thomson RC, Valenzuela N, …, Wilson RK (2013)
663 The western painted turtle genome, a model for the evolution of extreme physiological adaptations
664 in a slowly evolving lineage. *Genome Biol*, **14**, R28.

665 Shamblin BM, Bolten AB, Abreu-Grobois FA, Bjorndal KA, Cardona L, Carreras C, …, Dutton PH
666 (2014) Geographic Patterns of Genetic Variation in a Broadly Distributed Marine Vertebrate: New
667 Insights into Loggerhead Turtle Stock Structure from Expanded Mitochondrial DNA Sequences.
668 *PLoS ONE*, **9**, e85956.

669 Shamblin BM, Dodd MG, Griffin DB, Pate SM, Godfrey MH, Coyne MS, …, Nairn CJ (2017)
670 Improved female abundance and reproductive parameter estimates through subpopulation-scale
671 genetic capture-recapture of loggerhead turtles. *Mar Biol*, **164**, 138.

672 Skotte L, Korneliussen TS, Albrechtsen A (2013) Estimating Individual Admixture Proportions
673 from Next Generation Sequencing Data. *Genetics*, **195**, 693-702.

674 Stewart K, LaCasella E, Roden S, Jensen M, Stokes L, Epperly S, Dutton PH (2016) Nesting
675 population origins of leatherback turtles caught as bycatch in the U.S. pelagic longline fishery.
676 *Ecosphere*, **7**, e01272.

677 Subramanian S (2016) The effects of sample size on population genomic analyses--implications for
678 the tests of neutrality. *BMC Genomics*, **17**, 123.

679 Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA
680 polymorphism. *Genetics*, **123**, 585–595.

681 Todd E, Black M, Gemmell N (2016) The power and promise of RNA-seq in ecology and
682 evolution. *Mol Ecol*, **25**, 1224-1241.

683 Tzika A, Ullate-Agote A, Grbic D, Milinkovitch M (2015) Reptilian Transcriptomes v2.0: An
684 Extensive Resource for Sauropsida Genomics and Transcriptomics. *Genome Biol Evol*, **7**, 1827–1841.

685 Vieira FG, Lassalle F, Korneliussen TS, Fumagalli M (2016) Improving the estimation of genetic
686 distances from Next-Generation Sequencing data. *Biol J Linnean Soc*, **117**, 139-149.

687 Vilaça ST, Vargas SM, Lara-Ruiz P, Molfetti É, Reis EC, LÔBo-Hajdu G, Soares LS, Santos FR
688 (2012) Nuclear markers reveal a complex introgression pattern among marine turtle species on the
689 Brazilian coast. *Mol Ecol*, **21**, 4300-4312.

690 Wang Z, Pascual-Anaya J, Zadissa A, Li WQ, Niimura Y, Huang ZY, …, Irie N (2013) The draft
691 genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of
692 the turtle-specific body plan. *Nature Genetics*, **45**, 701–706.

693 Watterson G (1975) On the number of segregating sites in genetical models without recombination.
694 *Theor Popul Biol*, **7**, 256–276.

695 Wyneken J, Lohmann K, Musick J eds. (2013) *Biology of Sea Turtles Volume III*. CRC Press, Boca
696 Raton, FL.
697
698 **Data Accessibility**

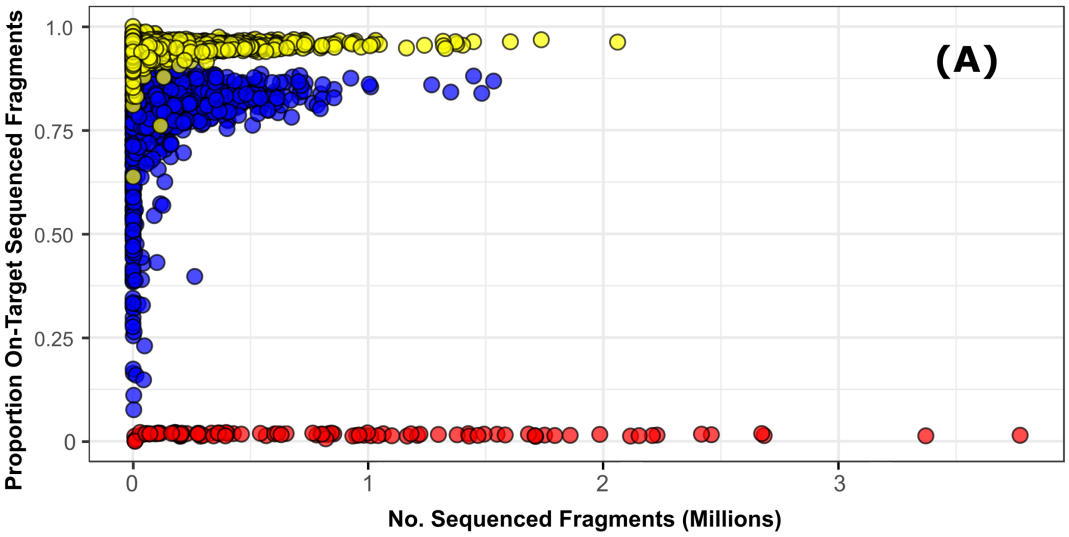
699 Data analyses scripts, documentation and Rapture platform probe sequences are available at
700 https://github.com/lkomoro/Marine_Turtle_Rapture_Methods. Illumina raw reads for Trial 2
701 hardshell turtles are deposited in NCBI Sequence Read Archive (Bioproject PRJNA487648).
702
703 **Author Contributions**

704
705 LMK, MM, SO, MPJ, KRS and PHD contributed to the conceptual design of the project. LMK,
706 MM and SO conducted laboratory, marker design, and data analyses. LMK, MPJ, KRS and PHD
707 assessed data interpretation for green turtles, and LMK and PHD wrote the manuscript.

Table 1. Initial SNP discovery per species with Rapture data for all QC passed samples (filters of MAF 0.05-0.4 and only sites with data for at least 50% individuals). Factors such as filtering thresholds, number of input samples, and source population of samples can affect identification of SNPs that are informative for different study goals.

Species	*C. mydas*	*C. caretta*	*E. imbricata*	*L. olivacea*	*L. kempii*	*D. coriacea*†	*D. coriacea*‡
No. Ind.	47	23	34	6	4	973	203
No. SNPs	11042	4502	6514	2048	1542	2835	2710

† All QC passed samples, global representation
‡ St. Croix nesting population QC passed samples



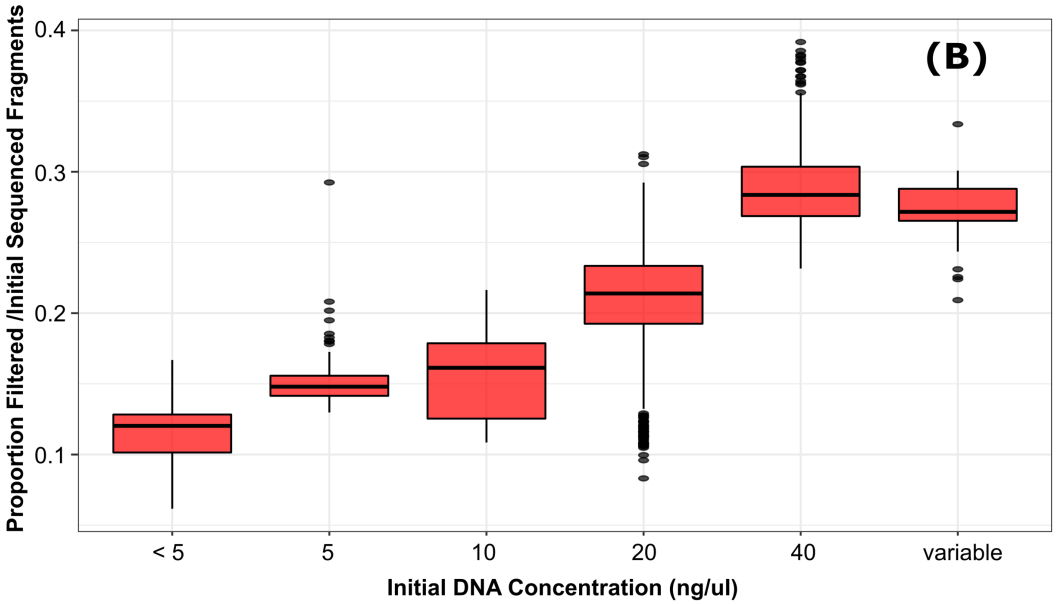


Figure 1. Panel (A) depicts the proportion of total sequenced fragments per individual that mapped to Rapture target loci from (1) initial RAD data (red circles), (2) Rapture data generated from original MYBaits protocol (Trial 1; blue circles), and (3) Rapture data generated from adapted MYBaits protocol (Trial 2; yellow circles). Note that one over-sequenced outlier with >7 million sequenced fragments was removed to improve visual interpretation. Panel (B) depicts the proportion of filtered mapped alignments/total sequenced fragments per individual for each category of initial DNA concentration (ng/ul).

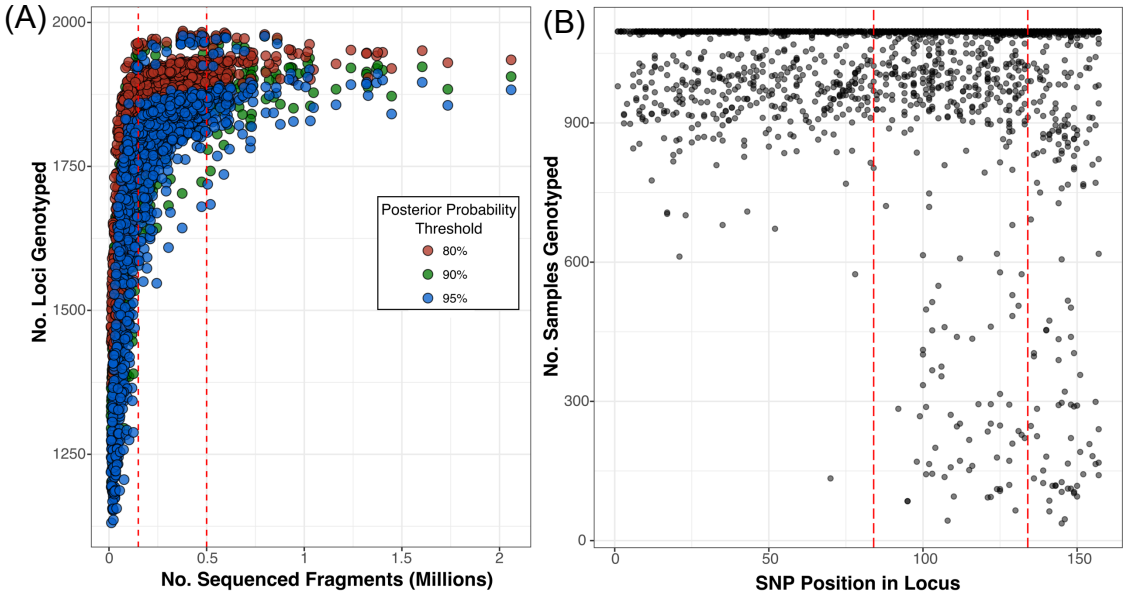


Figure 2. (A) Relationship between the number of sequenced fragments per individual and the number of *a priori* SNP loci genotyped, and (B) the relationship between the SNP relative position within a Rapture locus and the number of samples genotyped (visualized with 80% posterior probability threshold). Vertical lines added at relevant thresholds for visual interpretation (see text).

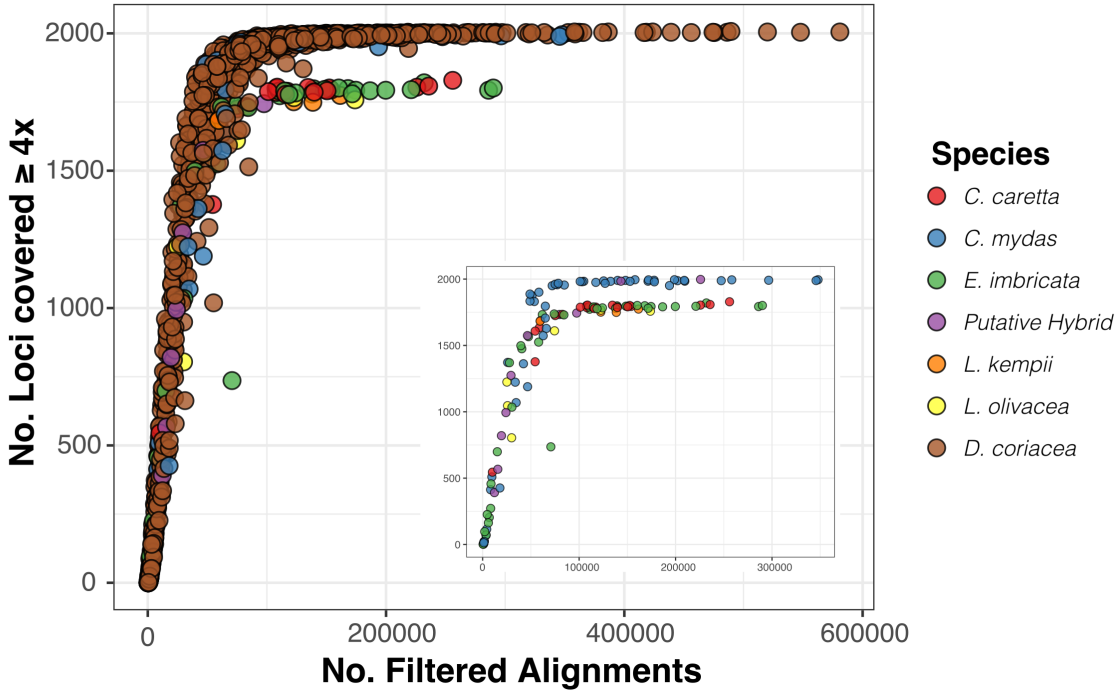


Figure 3. Number of Rapture loci covered ≥ 4x for all samples (one over-sequenced outlier with >1 million filtered alignments removed to improve visual interpretation). Inset depicts hardshell turtles to better visualize that only green turtles and green-hybrids attain coverage at all Rapture loci.

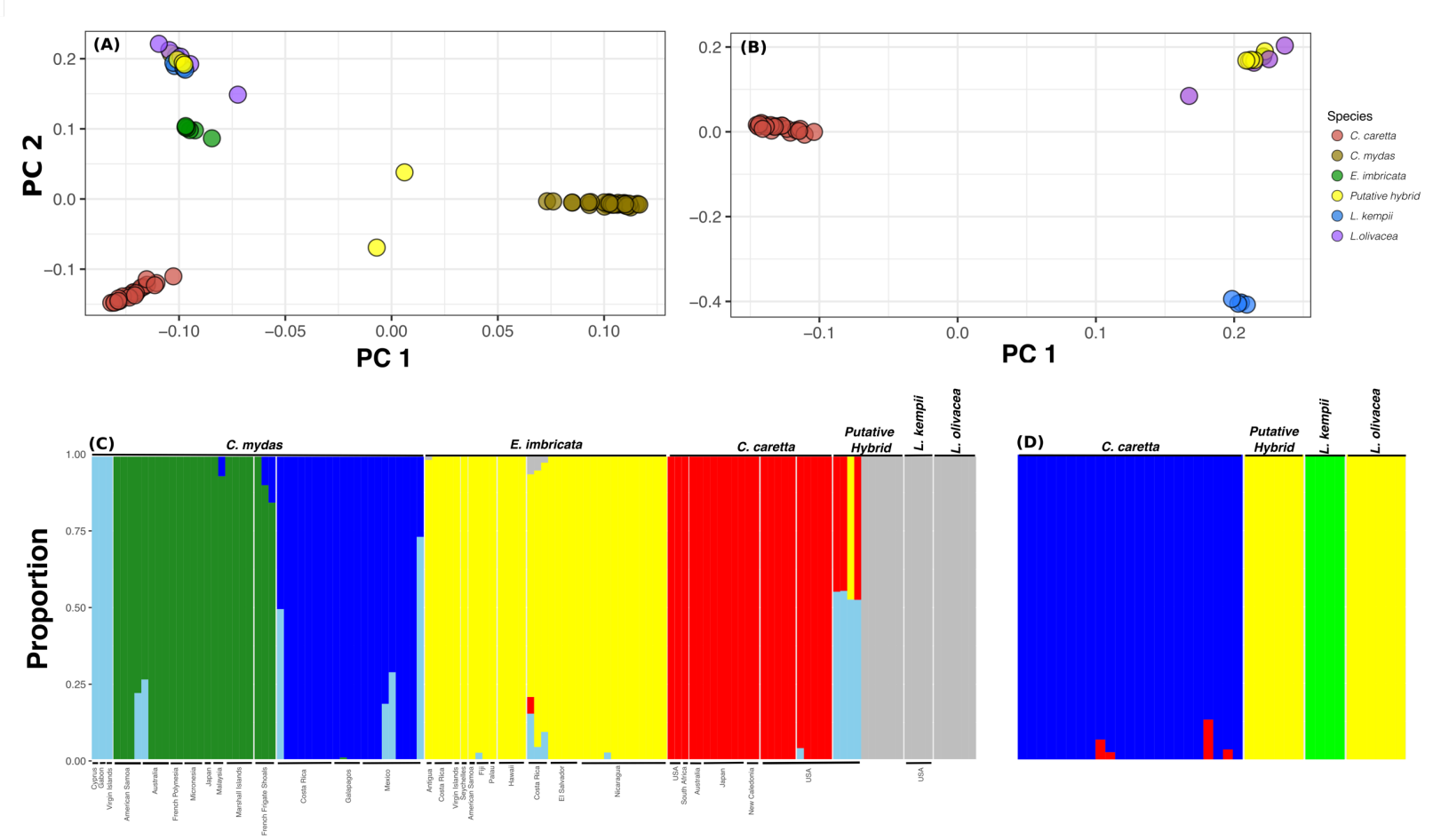


Figure 4. Species confirmation in hardshell turtles using principal components analyses (panels A and B) and admixture proportions (panels C and D). Panels (A) and (C) include all hardshell samples, while (B) and (D) include only of subsets of smaller groups, demonstrating how delineations among closer-related groups with smaller sample sizes can be masked in larger, disproportionate datasets. Only unresolved hybrids from the complete data set depicted in Panels A and C are included in Panels B and D.

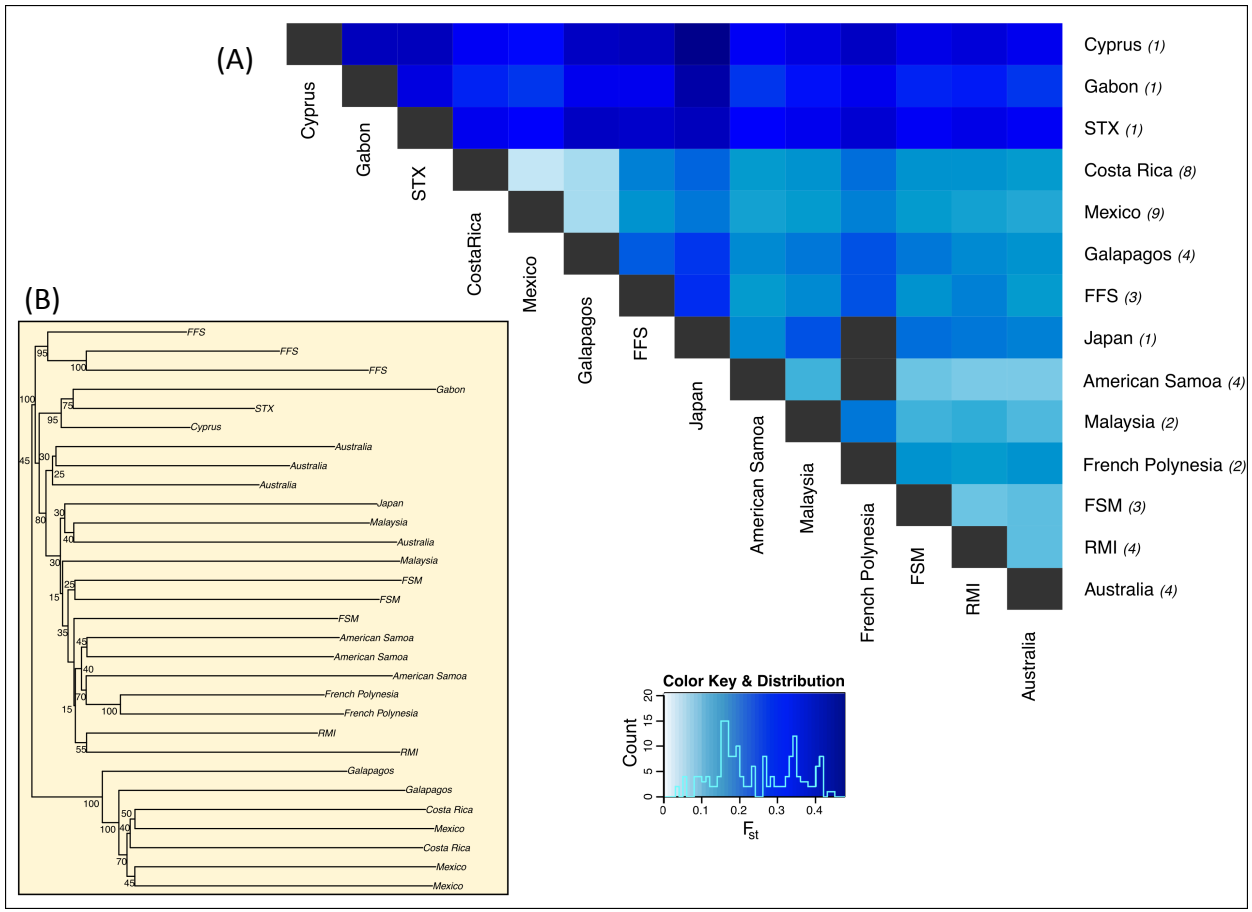
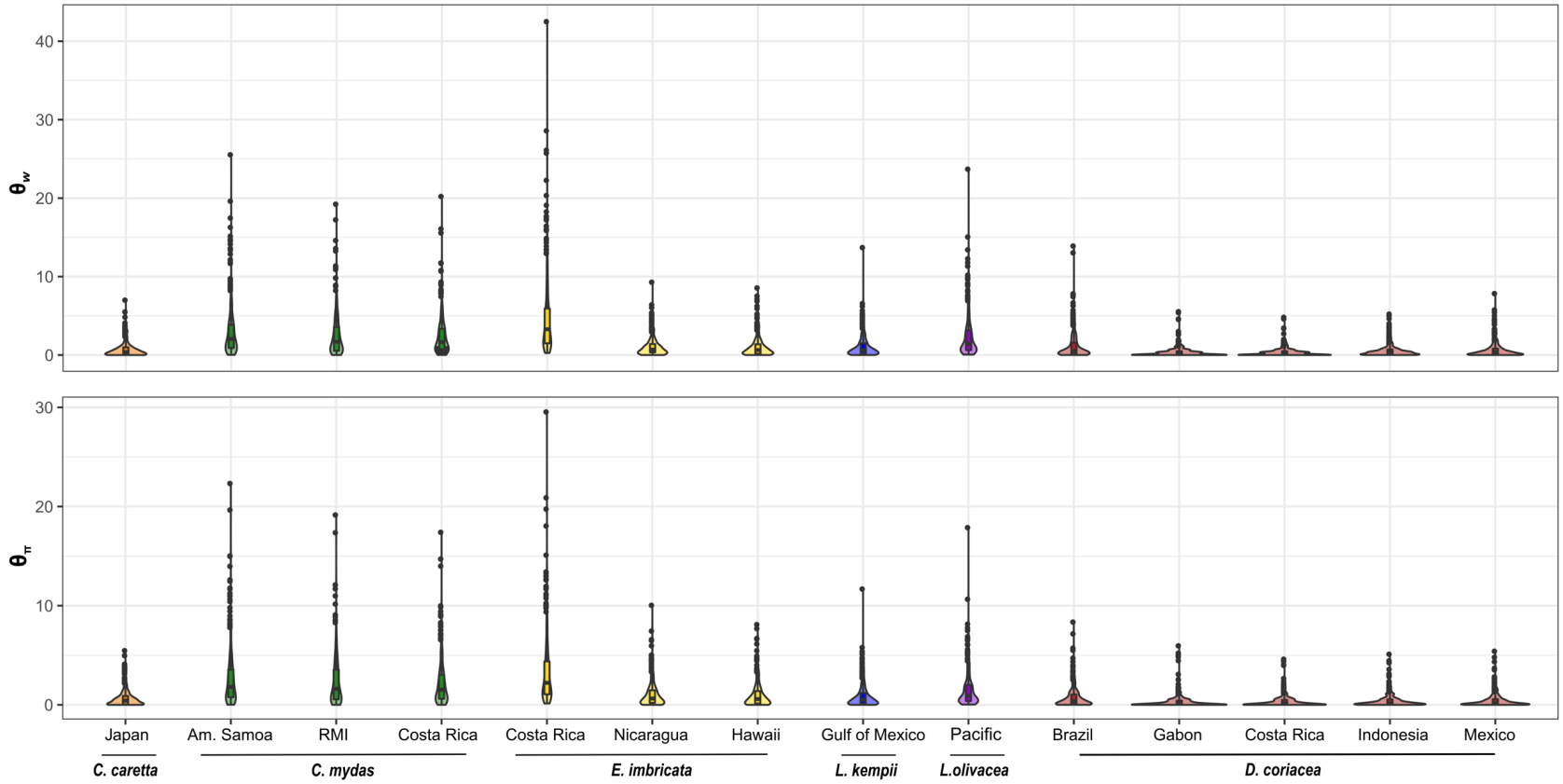


Figure 5. (A) Pairwise F_{st} values between green turtle nesting regions (sample sizes listed in italicized parentheses; black boxes indicates values could not be reliably calculated due to low sample size and sequencing coverage). (B) *FastME* tree of a representative subset of green turtle samples with topology and relative branch length based on genetic distances estimated in *ngsDist*. Branch support based on bootstrapping (1000 replicates, blocks of 500 SNPs). Abbreviations: STX=St. Croix, FFS=French Frigate Shoals, RMI= Republic of the Marshall Islands, FSM= Federated States of Micronesia.



748

Figure 6. Genetic diversity estimates (top: Watterson's estimator θ_w; bottom: Tajima's estimator θ_π) in representative groups for each species. Locations listed indicate nesting population with the exception of *L. olivacea* for which only bycatch samples with unknown nesting origin were available.

749