

## SEA TURTLE STOCK ESTIMATION USING GENETIC MARKERS: ACCOUNTING FOR SAMPLING ERROR OF RARE GENOTYPES

BENJAMIN BOLKER,<sup>1,3</sup> TOSHINORI OKUYAMA,<sup>1</sup> KAREN BJORN DAL,<sup>1,2</sup> AND ALAN BOLTEN<sup>1,2</sup>

<sup>1</sup>Zoology Department, University of Florida, P.O. Box 118525, Gainesville, Florida 32611-8525 USA

<sup>2</sup>Archie Carr Center for Sea Turtle Research, P.O. Box 118525, University of Florida,  
Gainesville, Florida 32611-8525 USA

**Abstract.** The contributions of different sea turtle rookeries to mixed-stock populations on foraging grounds can only be estimated by indirect methods such as analysis of mitochondrial DNA samples from the mixed stocks and rookery populations. We explain and evaluate methods for genetic stock estimation using simulations and data from previous studies. We focus on Markov Chain Monte Carlo (MCMC) estimation, a relatively new method. MCMC differs from older combinations of maximum likelihood (ML) with non-parametric bootstrapping in (1) using a Bayesian prior to quantify previous knowledge; (2) taking account of multiple modes in the probability distribution of contributions; and (3) incorporating sampling error more flexibly, allowing for the possibility that rare haplotypes actually present in a particular rookery were not detected in a small sample. In the context of sea turtle stock analysis, the differences in point estimates between ML and MCMC methods are relatively small, but MCMC gives wider and more accurate confidence limits than ML with bootstrapping, which tends to underestimate small contributions as zero.

**Key words:** *bootstrap*; *Caretta caretta*; *Chelonia mydas*; *Markov Chain Monte Carlo*; *maximum likelihood*; *migration*; *mtDNA*; *sea turtle*; *stock analysis*; *uncertainty*.

### INTRODUCTION

Sea turtles are a group of threatened and endangered species that are being intensely studied in the hopes of discovering enough about their ecology to guide conservation efforts. They have many of the typical traits of endangered species: long generation time, restricted reproductive habitats, and few nonhuman predators as adults. They also are subject to typical threats: legal and illegal harvesting, bycatch in commercial fisheries, and habitat destruction. Reproductive females may migrate thousands of kilometers to their natal beaches from foraging grounds with wide geographic distributions (Bowen and Karl 1997). These strong homing instincts maintain the separation of maternal lineages in different rookeries, leading to discrete stocks identifiable by their maternal (mitochondrial) DNA haplotypes. After emerging from their nests, hatchlings of most marine turtles (with the exception of the Australian flatback, *Natator depressus*) enter an oceanic stage, followed by recruitment to neritic habitats. This paper focuses on Atlantic loggerhead (*Caretta caretta*) and green (*Chelonia mydas*) turtles, for which we have good data, but our general conclusions should also apply to other regions and species. Loggerheads in the North Atlantic spend about eight years in oceanic habitats in the eastern Atlantic (Bjorndal et al. 2000), where individuals from a number of rookeries combine

in mixed stocks, before recruiting to neritic habitats (Bolten et al. 1998). The location and duration of the early life stage of green turtles are still unknown. In the Atlantic, green turtles recruit to neritic habitats at ~25 cm carapace length and then move among a number of foraging habitats where further mixing of stocks occurs (Lahanas et al. 1998, Bass and Witzell 2000).

Ecologists must gauge the proportions of a mixed population that originate in geographically disparate rookeries, both to improve management efforts and to understand the population dynamics of the full population, including rookeries and mixed populations on foraging grounds. Elucidating the source rookeries of mixed-stock foraging aggregations identifies regions that should be included in regional management plans and, because different rookeries have different survival outlooks, focuses protective measures on those foraging grounds with higher proportions of individuals from more threatened rookeries. Because of the difficulty of tracking individuals during the oceanic stage, assessments of the contributions of different rookeries to a particular mixed population must use indirect methods, such as the statistical analysis of mitochondrial DNA (mtDNA) haplotype composition in rookeries and mixed populations.

Maximum likelihood (ML) methods that use the distribution of genetic markers to estimate the contributions of different source pools to a mixed stock have a long history in fisheries (Miller 1987, Pella and Milner 1987, Utter and Ryman 1993). ML analyses of green and loggerhead turtle mtDNA data have shown that these species recruit to mixed populations in the

Manuscript received 5 November 2001; accepted 23 May 2002; final version received 19 September 2002. Corresponding Editor: H. B. Shaffer.

<sup>3</sup> E-mail: bolker@zoo.ufl.edu

Caribbean and eastern Atlantic from a variety of rookeries throughout the eastern Atlantic and Mediterranean (Bowen et al. 1996, Bolten et al. 1998, Lahanas et al. 1998). ML methods, however, suffer from some technical problems, particularly in their handling of rare and apparently missing haplotypes. A new (to conservation ecology) statistical method called Markov Chain Monte Carlo estimation (MCMC) has recently been used to address these problems (Pella and Masuda 2001). In this paper, we apply ML and MCMC methods to simulated data sets to assess the effectiveness of different methods; we reanalyze existing turtle mtDNA data with MCMC; and finally we discuss the implications of the statistical methods for turtle conservation and for stock analysis in general.

#### Conditional and unconditional maximum likelihood

Conditional maximum likelihood (CML) and unconditional maximum likelihood (UML) estimation are well described elsewhere (Pella and Milner 1987), but we start with a brief description of these methods to put the problem in perspective. Our data are the numbers of individuals with each mitochondrial haplotype ( $h$ ) sampled in each rookery ( $r$ ),  $F_{rh}$ , and the numbers sampled in the mixed population,  $M_h$ . The number of rookeries is  $R$  and the total number of distinct haplotypes represented in all stocks is  $H$  (in general, we will use capital letters to denote numbers of samples and lower case to denote frequencies). We want to estimate  $c_r$ , the proportion of the individuals in the mixed population contributed by each rookery. CML and UML are both based on finding the set of parameters with the highest likelihood: the probability of observing the sampled data given a particular set of parameters. CML assumes that the true haplotype frequencies in each rookery are equal to the actual frequencies observed in the sample:  $f_{rh} = F_{rh} / \sum_{s=1}^H F_{sh}$  where  $s$  is a summation variable and the summation is over all rookeries sampled (Pella and Milner 1987). Given these assumed frequencies, the expected frequencies in the mixed population are equal to  $m_h = \sum c_r f_{rh}$ , and the likelihood is the probability of drawing a multinomial sample  $M_h$  from a population with true frequencies equal to  $m_h$ . (The negative log likelihood, which is useful for computation, is equal to  $-\sum M_h \log m_h$  plus a constant, which can be ignored when maximizing the likelihood.)

Once the likelihood is defined, finding the maximum likelihood estimates is a straightforward computational problem, searching among parameter combinations for the combination that gives the highest likelihood (or lowest negative log likelihood). There are certain technical difficulties (contributions must be between 0 and 1, and must sum to 1), but these difficulties can be handled by standard transformations and numerical methods (see Appendix A: *CML/UML methods*).

The main problem with CML is the assumption that the true rookery haplotype frequencies are exactly equal to the frequencies observed in the rookery sam-

TABLE 1. A small simulated data set.

Haplotype	Rookery		Mixed population
	A	B	
I	10	50	10
II	50	10	50
III	0	1	1

ples. If sample sizes are small, or if some haplotypes are rare, this assumption is questionable; sampling error leads to discrepancies between the underlying frequencies and the sample frequencies.

For example, consider the data set in Table 1. The pattern of the common haplotypes (I and II) suggests that rookery A contributes almost all of the mixed population; in this relatively large sample, the haplotype frequencies in the mixed population match those in rookery A, and are quite different from those in rookery B. In contrast, the CML estimate of the contributions (0.92 from rookery A, 0.08 from rookery B) is strongly influenced by the evidence of two individuals with haplotype III, even though sampling error could easily explain the absence of haplotype III in rookery A.

One solution to this problem is unconditional maximum likelihood (Smouse et al. 1990), which allows for sampling error by simultaneously estimating the true frequencies in the rookeries ( $f_{rh}$ ) and the contributions from the rookeries ( $c_r$ ). The negative log likelihood is the sum of the negative log likelihood of the match between expected and observed frequencies in the mixed population ( $-\sum M_h \log m_h$  plus a constant) and the match between the expected and observed frequencies in the rookeries ( $-\sum F_{rh} \log f_{rh}$  plus a constant). UML can ascribe lack of fit either to an unlikely sample from the mixed population, given the true haplotype proportions (which are a function of the contributions of different rookeries and of the true haplotype proportions in the rookeries), or to an unlikely sample from the rookeries. The balance between these sources of error depends on the sample sizes from the rookeries and the mixed population (the method prefers to attribute sampling errors to a smaller multinomial sample, where they are more likely) and the detailed pattern of the data. In the previous example, UML suggests that it is more likely that haplotype III is actually present in rookery A, but was not sampled, and that 0.995 of the mixed population really comes from rookery A.

UML demands more computing power than CML, and specialized algorithms have been developed for it (Pella and Milner 1987), but (especially in light of ever-increasing computational power) it can also be done with standard algorithms as detailed in Appendix A: *CML/UML methods*.

#### BOOTSTRAPPING

In addition to the maximum likelihood (point) estimates given by CML and UML, we need confidence

limits describing the range of uncertainty in the estimate. Both large sample sizes and large differences between haplotype frequencies in different rookeries contribute to accurate estimates. The nonparametric bootstrap resamples the data from each rookery and from the pooled population with replacement, and then estimates the rookery contributions anew from the new bootstrapped data set. This procedure is repeated many times, and the confidence limits are estimated from the distribution of bootstrapped estimates: for example, the 25th- and 975th-largest estimates from 1000 bootstrap samples give the estimate of the 95% confidence limits. The nonparametric bootstrap is equivalent to drawing a new multinomial sample with the sample sizes and haplotype frequencies observed in the original data. However, nonparametric bootstrapping still suffers from the problem of missing haplotypes. Resampling a rookery where no individuals of a particular haplotype were initially sampled can never produce that haplotype in the bootstrapped data set (cf. Walsh 2000).

#### MARKOV CHAIN MONTE CARLO

An alternative to UML with nonparametric bootstrapping is Markov Chain Monte Carlo (MCMC), a very general approach to estimating parameters in situations with multiple levels of random variation, such as the current problem where there is uncertainty in true haplotype frequencies and in sampling. Although these kinds of situations are common in ecology and conservation ecology, the use of MCMC is rare in ecology (Gibson 1997, Gibson and Renshaw 1998, Gottwald et al. 1999). MCMC can flexibly account for some of the uncertainties that nonparametric bootstrap misses; in particular, observed samples of zero can be resampled as nonzero values in a way that is consistent with other information in the data set.

MCMC refers to a particular way of sequentially picking random guesses (based on the data, on prior information, and on the prior guess) at plausible values of the parameters. When done according to the appropriate recipes, this procedure produces a set of values that are an estimate of the probability distribution of the parameters for which we are looking. The advantages of MCMC are (1) it accounts better for sampling error (or multiple sources of random variation) than does bootstrapping; (2) with a sufficiently clever implementation, it can be as fast or faster than bootstrapping; and (3) the foundations of MCMC lie in Bayesian estimation, which means that we can incorporate prior information and interpret our results as probability distributions of parameters in a natural way. The advantage of Bayesian frameworks is still controversial; in this paper, we will simply compare what the methods have to tell us about turtles.

In the most general sense, all we have to do to take one step in a MCMC analysis is to pick a new, random set of parameters that are consistent with the data, the previously guessed set of parameters, and possibly

some amount of prior information about the parameters. The core property is that the relative probability of moving from one set of parameters to another is proportional to the relative probability of those parameters, given the evidence of the data (the likelihood, as previously defined) and any prior information. There are (of course) many technical details, but any set of rules for picking new parameters that has this property will lead, in the long run (after discarding parameters from an appropriate "burn-in" period), to an appropriate set of estimates.

The *Methods* section will present more detail on the particular implementation of MCMC for estimating contributions from different rookeries, introduce a simulation framework for comparing the effectiveness of different methods, and describe the criteria we use for evaluating point estimates and confidence limits. In *Results*, we show results for both a broad range of simulations and for the available loggerhead and green turtle data. Finally, we discuss the implications of MCMC estimation for stock analysis, in general, and for turtle stock analysis, in particular.

#### METHODS

##### *MCMC implementation*

Our implementation of MCMC uses a two-stage algorithm (an example of a so-called Gibbs sampler, which resamples parameters sequentially) for picking new sets of parameters (Pella and Masuda 2001). The first stage (Fig. 1, Step A) starts with an initial guess at the parameters (both the true haplotype frequencies,  $f_{rh}$ , and the contributions of different rookeries to the mixed population,  $c_r$  (Fig. 1, Step 0), and estimates the probability of a turtle in the mixed population with a particular haplotype coming from a given rookery. With given estimates of  $f_{rh}$  and  $c_r$ , the expected contribution of a particular haplotype ( $h$ ) from a particular rookery is  $c_r f_{rh}$ ; the probability of a mixed-stock individual with haplotype  $h$  coming from rookery  $r$  is this contribution divided by the total contribution from all rookeries,  $c_r f_{rh} / \sum_{r=1}^R c_r f_{rh}$ . We then pick multinomial samples  $I_{rh} \sim \text{Mult}(M_h, f_{rh})$  with these probabilities, which give a consistent guess for the unknown origins of the individuals sampled from the mixed pool. Another way of describing this procedure is that we impute rookery origins to each of the turtles in the mixed population in a way that is random, but consistent with our current set of parameters.

The second stage (Fig. 1, Step B) takes these guesses of rookery origins for granted, estimates the probability distributions of the rookery haplotype frequencies and contributions, and picks random samples out of these distributions. (These distributions are so-called Dirichlet distributions; for details, see Appendix B.) For this step, we must also specify prior information on the haplotype frequencies ( $F_{rh}^p$ ) and rookery contributions ( $C_r^p$ ) (Fig. 1, PRIOR); we use upper case for these priors

DATA							
a) Numbers ( $F_{rh}, M_h$ )			b) Frequencies ( $F_{rh} / \sum_{h=1}^H F_{rh}$ )				
	Rook A	Rook B	Mixed		Rook A	Rook B	Average
I	10	50	10	I	0.167	0.83	0.17
II	50	10	50	II	0.833	0.15	0.82
III	0	1	1	III	0	0.02	0.01
<b>sum</b>	$F_1 = 60$		$F_2 = 61$				

PRIOR					
c) Contribution, $C_r^p$ (all equal)			d) Haplotype freq., $f_{rh}^p$ (overall average)		
	Rook A	Rook B		Rook A	Rook B
	1	1	I	0.17	0.17
			II	0.82	0.82
			III	0.01	0.01

STEP 0					
e) Initial contributions, $c_r$ (all equal)			f) Initial hap. freq., (= observed freq.)		
	Rook A	Rook B		Rook A	Rook B
	0.5	0.5	I	0.167	0.83
			II	0.833	0.15
			III	0	0.02

STEP A						
g) Estimated prob. of origin: $\{w_{rh}\} = \frac{c_r f_{rh}}{\sum c_r f_{rh}}$			h) Imputed origin: $\{I_{rh}\}$			
	Rook A	Rook B		Rook A	Rook B	
I	0.168	0.832	→	I	0	10
II	0.847	0.153	multinomial	II	44	6
III	0	1	sample	III	0	1
<b>sum</b>				<b>sum</b>	$I_1 = 44$	$I_2 = 17$

STEP B						
New parameters: $\{f'_{rh}\} = \text{Dirichlet}(\{F_{rh} + I_{rh} + f_{rh}^p\})$ , $\{c'_r\} = \text{Dirichlet}(\{c_r^p + I_r\})$						
i) Dirichlet parameters			j) New values ( $f'_{rh}, c'_r$ )			
	Rook A	Rook B		Rook A	Rook B	
I	$10 + 0.17 + 0 = 10.17$	$50 + 0.17 + 10 = 60.17$	→	I	0.116	0.852
II	$50 + 0.82 + 14 = 94.82$	$10 + 0.82 + 6 = 16.82$	Dirichlet	II	0.882	0.132
III	$0 + 0.01 + 0 = 0.01$	$1 + 0.01 + 1 = 2.01$	sample	III	0.002	0.016
<b>Contrib.</b>	$1 + 14 = 45$	$1 + 17 = 18$		<b>Contrib.</b>	0.617	0.383

FIG. 1. One round of the Gibbs sampler algorithm (Pella and Masuda 2001) for estimating turtle stock mixtures, with the data presented in Table 1 (a, b); initial estimates of the haplotype frequencies (f) equal to the observed haplotype frequencies (b) and initial estimate of equal contributions from both rookeries (e); prior estimates of haplotype frequencies (d) equal to the average observed frequencies (b); and prior estimates of rookery contributions set equal (c). See *Methods: MCMC implementation* for description of the procedure, which cycles repeatedly between Step A and Step B. In DATA (b), average refers to the harmonic average of the rookery haplotype frequencies.

because they can be expressed in terms equivalent to numbers of individuals in a prior sample, although the numbers need not be integers. The probability distribution of rookery haplotype frequencies is based on the number of a particular haplotype observed in a rookery (Fig. 1a), plus the number of turtles with that haplotype in the mixed population that are imputed to be from that rookery (Fig. 1h), plus any prior information on rookery haplotypes (Fig. 1d). The proba-

bility distribution of contributions is simply based on the number in the mixed pool imputed to be from a given rookery (Fig. 1h), plus prior information (Fig. 1c). To finish the second stage (and return to the beginning of the cycle), we pick frequencies from the probability distributions of rookery haplotypes and of contributions.

To use MCMC (or any Bayesian method), we have to decide what prior information to incorporate, in this

case, information on haplotype rookery frequencies and on contributions. We follow the methods of Pella and Masuda (2001) for picking priors that are weak, although not completely uninformative. The contribution priors are set equal with a strength of  $1/R$ ; a prior strength of 1 corresponds to complete ignorance, and a prior  $<1$  actually shifts the estimates toward larger contributions from large-contributing rookeries and smaller contributions from small-contributing rookeries, although the effect is not strong. Pella and Masuda's priors for the haplotype frequencies are proportional to the among-rookery harmonic means of rookery frequencies, set to a strength that minimizes the sum of squared deviations between the observed frequencies and the weighted average of the priors and the observations. These priors allow for some undetected overlap of haplotypes among rookeries.

MCMC only converges to the appropriate distribution of parameter values in the long run, after the starting values of the parameters (e.g., Fig. 1, Step 0) have been "forgotten." There are standard methods for evaluating how long this so-called "burn-in" period should be, and how long the chain must be run to get reasonable estimates; we use the implementation in the (R version of) the publicly available CODA package (version 0.5–12, 6/02).<sup>4</sup> In particular, we first use the Raftery-Lewis criterion, which estimates how long a single chain should be to achieve a particular level of accuracy in estimating a particular quantile of the parameter distributions (Raftery and Lewis 1996). We estimate the time required to estimate the 97.5% quantiles within a 2% margin of error with a 95% probability. Once we have a chain long enough to pass this criterion, we double-check with the Gelman-Rubin criterion (Gelman et al. 1995), which makes sure that the variance between a set of MCMC chains started from different points is not much larger than the variance within chains. This makes sure that the different chains have all moved away from their starting points and are covering the same region of parameter space. We use one chain for each rookery, each starting with an estimate of 95% contributions from that rookery and the remainder evenly split ( $5/(R - 1)\%$  each) among the others. For the simulations and data that we will present, our burn-in times are surprisingly short (on the order of 100 steps) and our convergence times are on the order of 20 000 steps.

Masuda and Pella have built a stand-alone Windows program that implements these algorithms, using the same input formats as their previous stock analysis programs.<sup>5</sup> We re-implemented the algorithms using R, a public domain statistics and programming language (Ihaka and Gentleman 1996; available online),<sup>6</sup> as a

		Rookery				
		A	B	C	D	E
Haplotype	I	$c$	$i$	$r$	$r$	$r$
	II	$i$	$c$	$r$	$r$	$r$
	III	$r$	$r$	$c$	$i$	$r$
	IV	$r$	$r$	$i$	$c$	$i$
	V	$r$	$r$	$r$	$i$	$c$

FIG. 2. Haplotype frequency structure:  $c$ , common haplotype;  $i$ , haplotype that occurs at intermediate frequencies; and  $r$ , a "rare" haplotype. The turtle rookeries are organized into two blocks, each of which shares common and intermediate haplotypes. Common haplotypes for one block are found only rarely in the other block (in simulations,  $r$  is set to either 0% or 4%).

graphical and data management framework. The actual MCMC code was written in C for speed (and was compiled for Windows and Linux; native R code is also available, although slower). An R library for stock analysis by CML, UML, and MCMC is available.<sup>7</sup> Our library can only handle haploid data, not the full range of possible genetic data (e.g., microsatellites or diploid markers), and requires the R package. On the other hand, for those who do choose to use R, it may provide a more convenient interface.

#### Simulations

We constructed a framework for simulation that allowed us to evaluate different stock analysis methods for a variety of possible configurations of rookery genotypes and contributions. Each simulation run incorporates the true distribution of haplotypes among rookeries, the true contributions of rookeries to the mixed population, and the sizes of samples taken from different rookeries and from the mixed population. The main properties of the haplotype distribution are the dominance of the most common haplotypes, the overlap of haplotypes among rookeries, and the characteristics of the "tail" of the haplotype distribution (whether rare haplotypes really appear at low frequency throughout all rookeries, or whether they are really confined to one or a few rookeries).

We set up a fairly general structure that mimics the observed structure of haplotypes among sea turtle rookeries (Fig. 2).

We varied the characteristics of a given simulation by changing the ratio of common to intermediate haplotypes ( $c/i$ ), and whether the "rare" haplotypes were really absent ( $r = 0$ ) or present at low frequency ( $r = 0.04$ ). With this scheme, the degree of overlap between neighboring rookeries is hard to manipulate independently of the dominance of common haplotypes within rookeries (both are controlled by  $c/i$ ), but the block structure indicated in Fig. 2 does change the overlap:

<sup>4</sup> URL: <http://www.cran.r-project.org/src/contrib/PACKAGES.html#codas>

<sup>5</sup> URL: <ftp://ftp.afsc.noaa.gov/sida/mixture-analysis/bayes/>

<sup>6</sup> URL: <http://www.r-project.org>

<sup>7</sup> URL: <http://www.zoo.ufl.edu/bolker/turtle>

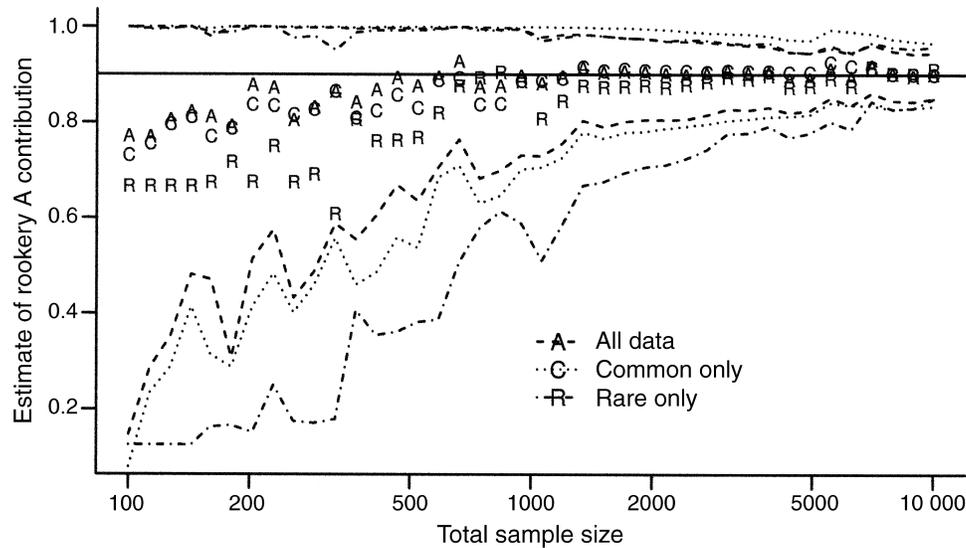


FIG. 3. MCMC estimates of known turtle rookery contributions (1) using all information, (2) discarding rare haplotypes and using only common haplotypes, and (3) discarding common haplotypes and using only rare haplotypes. Total sample sizes are allocated 0.50 in the mixed stock and 0.25 in each rookery. Each rookery has a dominant haplotype, which appears in the other rookery at a lower frequency, and two rare haplotypes, which are absent from the other rookery. Haplotype frequencies in rookery A are {0.65, 0.31, 0.01, 0.01, 0.0, 0.0}; in rookery B, they are {0.31, 0.65, 0.0, 0.0, 0.01, 0.01}. True contributions are 0.90 from rookery A (shown as a horizontal line) and 0.10 from rookery B. Mean and 95% confidence limits shown are averages of estimates from five MCMC estimates from each of five simulations for each sample size and choice of data.

there are two groups of more closely related rookeries with less haplotype overlap between them. For rookery contributions, the characteristics of dominance (ratio of common contributions to intermediate contributions) and of rarity (presence or absence of rookeries that contribute at low levels) are both important, and we can specify both in a way similar to the definitions for the haplotypes. Specifying sample sizes is also straightforward; we used various sample sizes per rookery (25, 50, and 100 were our default values), and doubled this sample size in the mixed population.

In addition to these general simulations, which we used to evaluate a broad range of conditions, we also developed two more specific simulation protocols to answer particular questions. First, in order to understand how the power to detect the absence of contributions from a particular rookery varies with total sample size, we started with the estimates of contributions of green turtles to the mixed population and set the contributions from low-contribution rookeries (Suriname, Brazil, Mexico, Ascension, and Cyprus) to zero. We tried a range of total sample sizes, of which half were taken from the mixed population, with the other half evenly divided among the nine rookeries. Simulations took multinomial samples with these sample sizes from the estimated haplotype frequencies in each rookery and from the expected haplotype frequencies in the mixed population.

Second, in order to understand how rare and common haplotypes contribute to statistical power (Fig. 3), we took

a simple six-haplotype, two-rookery example with rookery A haplotype frequencies {0.65, 0.31, 0.01, 0.01, 0.0, 0.0}; rookery B haplotype frequencies {0.31, 0.65, 0.0, 0.0, 0.01, 0.01}; and true contributions, 0.90 from rookery A and 0.10 from rookery B. We then used MCMC to estimate the contributions using (1) all of the haplotype data; (2) common haplotypes ( $>0.01$ ) only, discarding rare haplotypes; or (3) rare haplotypes ( $\leq 0.01$ ) only, discarding common haplotypes.

#### Criteria

In order to decide whether ML or MCMC methods are better suited to turtle stock analysis, we have to establish criteria that define a good estimate, for both point estimates and confidence intervals. Given a set of simulations in which one knows the true values of the rookery contributions, and given an estimate such as the ML estimate, one can calculate the bias (expected deviation of the estimate from the true value) and the variance (expected variance of the estimate around its expected value). There is a fundamental trade-off between minimizing bias and minimizing variance. To determine the best balance between bias and variance, one can also quantify the total squared error of the estimate around the mean, which is equal to the bias squared plus the variance.

Traditional frequentist estimation methods like CML and UML attempt to find the single best-fit value of the parameters, which, in these cases, means the maximum likelihood value or, in Bayesian terms, the mode

of the probability distribution of parameters. MCMC and other Bayesian methods estimate the entire probability distribution, so one can choose among different summaries of the probability distribution: the mode is one possibility, but the mean is more commonly used. Using the mean rather than the mode of the distribution means that low-probability but important possibilities (such as large contributions by a particular rookery) are included in the point estimate. Thus, even though MCMC and UML are essentially estimating the same distribution, they give different answers. It is also possible (as recommended by Pella and Masuda [2001]) to use the Bayesian mode as one's point estimate when using MCMC, bringing the ML and MCMC answers closer together; this is essentially a matter of taste or philosophy. In our simulations (and, we suspect, in those of Pella and Masuda [2001]), the mean and the mode are nearly equivalent in statistical efficiency, representing a simple trade-off between bias and variance. One may reduce the bias slightly by using the mode, but only at the cost of increasing variance and, hence, total error.

The standard criterion for the confidence interval is the coverage: the percentage of the time in repeated simulations that the estimated confidence interval includes the true value. If the coverage is greater than the nominal size of the confidence limits (e.g., 98% instead of 95%), then the confidence limits are larger than they should be, or too pessimistic; if the coverage is smaller, then the confidence intervals are too optimistic.

## RESULTS

We have computed CML, UML, and MCMC estimates (point estimates and confidence limits), according to the algorithms described in the *Methods* section and in the technical appendices, for a wide range of simulations and for the available loggerhead and green turtle data. Although stock estimation has traditionally focused more on the performance of point estimators than on confidence intervals, we will focus (after a brief discussion of point estimates) on the performance of the confidence intervals. We do care about the accuracy of point estimates (despite statisticians' cautions, managers still want a "best" estimate), but accurate confidence intervals may be more critical for avoiding truly worst case scenarios and coming up with management strategies that are robust to uncertainty. We can only know the performance of the estimators relative to true values (bias, variance, coverage) for the simulation runs, in which we know the true values; we discuss these results first and then turn to what the different estimators tell us about the turtle data.

### *Simulations*

*Point estimates.*—The bottom line is that there is little difference between the point estimates for ML and MCMC methods (Table 2); after all, both are es-

entially using the same set of data to solve the same likelihood problem. What differences there are appear when estimation is especially difficult because of overlapping haplotypes in rookeries, small contributions by a number of rookeries, and small sample sizes. The differences in the estimates represent different trade-offs between bias and variance (neither method is really more accurate) and can be ascribed to the effects of the prior and to the differences between the mean and the mode of the distribution of estimates.

When haplotypes, contributions, and sample sizes are such that estimation is fairly easy, there are few differences between the results of CML, UML, and MCMC estimation. In order to make estimation easy, one needs a few common haplotypes that are fairly specific to particular rookeries, roughly equal contributions from different rookeries, and large sample sizes. Even in the opposite case (with overlapping haplotypes, a single dominant rookery with other rookeries contributing little or nothing, and small sample sizes), the differences in point estimates are relatively minor and stem from two sources: (1) the difference between using the mean vs. the mode as a point estimate and (2) the MCMC prior.

The other difference between ML and MCMC methods is that MCMC methods (like all Bayesian approaches) necessarily incorporate some estimate of prior information. The prior is typically made weak (one assumes that there is little prior information), but it still has an effect in any part of the estimation problem where there is little information in the available data. The priors suggested by Pella and Masuda (equal contributions from all rookeries, and equal haplotype frequencies in all rookeries) tend, in those places where the data shed little light, to shift the estimate toward equal contributions and to predict greater uncertainty in the estimates. Depending on whether the assumptions made by the priors match the "true" situation (the assumptions or parameters in the simulation), the priors may make the estimates either better or worse. For example, if some haplotypes are widely distributed among rookeries but are at low frequency in each rookery, such that they will often be missed by sampling error, the Bayesian prior will improve the estimates by making the right assumption; conversely, if haplotypes are really restricted to a few rookeries and are completely absent elsewhere, the prior will degrade the estimates by making the wrong assumption. We conclude that ML methods are more or less at the limit of efficiency, extracting as much usable information as is present in the data set; without making further assumptions in the form of a prior, one can play with the bias-variance trade-off, but cannot reduce the total error of estimation.

*Confidence limits.*—In contrast to the minor differences in point estimates, the confidence limits estimated by ML and MCMC are different in an important way (Table 2). The use of a Bayesian prior in the

TABLE 2. Simulation results: bias, variance, error, and minimum/mean/maximum of coverage by turtle rookery, for UML (unconditional maximum likelihood) and MCMC (Markov Chain Monte Carlo) methods.

Haplo- type ratio, <i>c/i</i>	Factors			UML						MCMC					
	Rare con- trib.† (%)	Rare hap.‡ (%)	Sam- ple size§	Estimate   (%)			Coverage¶			Estimate   (%)			Coverage¶		
				Bias	Var.	Error	Min.	Mean	Max.	Bias	Var.	Error	Min.	Mean	Max.
5	0.0	0	25	0.754	43	43.7	82	88.6	96	1.08	27.9	29.2	87	96.2	100
5	0.0	0	50	0.431	25.7	25.9	83	90.4	97	1.5	18.5	21.4	82	95.4	100
5	0.0	0	100	0.312	10.8	11	80	89.6	97	1.72	8.08	12.6	93	97.2	100
5	0.0	4	25	2.39	68.8	76.1	71	84.8	94	2.26	36.3	43.2	94	96.8	98
5	0.0	4	50	1.67	35.7	39.4	76	87.8	98	1.56	21.7	24.3	96	97.2	99
5	0.0	4	100	1.06	19.4	21	77	88.8	97	1.48	13.6	16	94	96.6	99
5	2.5	0	25	1.03	45.5	46.7	81	88.2	98	0.855	31.3	32.1	92	97.4	100
5	2.5	0	50	0.61	24.6	25.1	91	94.0	96	1.1	17.9	19.8	95	96.8	98
5	2.5	0	100	0.126	11.5	11.5	91	94.6	99	1.32	9.72	13.5	92	94.4	96
5	2.5	4	25	0.904	65.9	67.2	90	92.0	95	1.06	38.9	40.3	95	98.0	100
5	2.5	4	50	0.872	41.3	42.4	93	94.6	96	0.74	27.7	28.6	96	98.2	100
5	2.5	4	100	0.417	19.8	20	92	96.2	99	1.38	13.8	17.3	95	98.0	100
8	0.0	0	25	0.427	30.7	30.9	83	89.2	96	1.02	22.5	23.6	87	95.6	100
8	0.0	0	50	0.25	17.7	17.8	82	90.0	96	1.28	15	17.4	93	97.2	100
8	0.0	0	100	0.231	7.44	7.51	73	87.8	97	1.21	7.94	10.3	92	96.2	100
8	0.0	4	25	1.77	47	51.5	66	81.8	95	2.11	30.1	35.3	94	94.6	95
8	0.0	4	50	1.35	26.2	28.9	74	86.4	96	1.28	20.4	22.1	94	95.8	97
8	0.0	4	100	0.84	13.6	14.8	71	85.4	96	1.26	12.3	14	93	95.4	98
8	2.5	0	25	0.606	33.3	33.7	77	87.4	96	0.867	25.3	26.2	89	96.6	100
8	2.5	0	50	0.323	19.4	19.5	89	94.4	97	1.17	17.4	19.6	92	95.2	98
8	2.5	0	100	0.179	8.72	8.76	90	94.2	98	1.08	9.69	12.1	88	93.0	97
8	2.5	4	25	1.09	52.9	54.6	88	89.6	93	1.09	34.5	35.8	91	96.0	100
8	2.5	4	50	0.702	28.8	29.5	92	95.4	97	0.897	22.7	24.1	94	97.0	99
8	2.5	4	100	0.412	14.3	14.6	95	96.4	97	1.1	12.9	15.1	94	96.6	100

Notes: Simulations are based on five rookeries (structured as in Fig. 2). Different simulation runs vary the ratio of common to intermediate haplotype frequencies (*c/i*).

† Frequency of rare haplotype contributions from rookeries to the mixed stock.

‡ Frequency of rare haplotypes in turtle rookeries.

§ Sample size is per turtle rookery; the value is doubled in a mixed population.

|| The mean absolute value of bias in percentage contribution of haplotypes; the mean variance and the mean error in percentage contribution.

¶ Coverage is of nominal 95% confidence intervals.

MCMC results does affect the answers given: it tends to increase the uncertainty in areas where the data shed little light, because it smears haplotype frequencies across rookeries. More important, MCMC treats uncertainty in a fundamentally different way than does the nonparametric bootstrapping used with ML methods. Rather than simply resampling the observed counts, it allows the actual underlying frequencies to vary. This resampling allows MCMC to take a larger range of variation into account and produces wider confidence intervals. In the worst case scenario for MCMC, when contributions from low-contributing rookeries are really zero, the wider confidence limits are inappropriate and lead to coverages somewhat greater than 95%. However, the confidence intervals produced by MCMC are more robust across the board than those from bootstrapping with UML, which fail badly when many rookeries contribute sparsely.

To follow up on the effects of sampling rare and common haplotypes, we ran the simulation described in *Methods* in which we estimated point estimates and confidence limits using all data, only common haplotypes, or only rare haplotypes.

Fig. 3 shows that throwing away any information, from either common or rare haplotypes, degrades our ability to estimate contributions. However, rare haplotypes alone fall far short, even when they are really distributed across rookeries in an informative way.

#### LOGGERHEAD AND GREEN TURTLE DATA

Having checked the properties of MCMC estimates using simulations, we now turn to reanalyzing mtDNA data on loggerhead and green turtles sampled from individuals caught in rookeries and in the mixed populations. These data were originally presented by Lahanas et al. (1998) and Bolten et al. (1998), where they were analyzed using UML; we redo the analysis with CML, UML, and MCMC for comparison. The only difference from the originals in our analysis is that we do not lump Suriname and Aves Island together in the analysis of the green turtle data (Table 3).

*Point estimates.*—Fig. 4 shows the results of running the MCMC algorithm on data from green (*Chelonia mydas*) and loggerhead (*Caretta caretta*) turtles (taken from Lahanas et al. [1998] and Bolten et al. [1998], respectively). The point estimates (which are the mean

TABLE 3. Point estimates from MCMC for loggerhead and green turtles.

Rookery	Method		
	CML	UML	MCMC
A) Loggerheads			
NWFL	†	†	0.109
SOFL	0.684	0.706	0.542
NEFL.NC	0.197	0.187	0.139
Mexico	0.119	0.107	0.158
Greece	†	†	0.0499
Brazil	†	†	0.00153
B) Green turtles			
FL	0.0546	0.0537	0.0386
MEXI	†	†	0.0113
CR	0.783	0.785	0.789
AVES	0.149	0.148	0.107
SURI	†	†	0.0382
BRAZ	†	†	0.00341
ASCE	†	†	0.00444
AFRI	0.0133	0.0133	0.005552
CYPR	†	†	0.00174

† Estimated contributions  $\leq 0.0001$ .

values given in the MCMC chain) show few qualitative differences from previous estimates derived using UML with (nonparametric) bootstrap or jackknife confidence intervals. There is a general tendency for MCMC to predict nonzero contributions from more

different rookeries. For example, MCMC estimates that all green turtle rookeries contribute  $\geq 0.001$  of the mixed stock (Florida, Mexico, Costa Rica, Aves, Suriname) contribute  $>0.005$ ; UML predicts that only four rookeries contribute  $>0.005$ , and none of the others contributes  $>0.001$ . We believe this is not simply the result of a different bias toward equal contributions, because the estimated contribution from Costa Rica stays the same at  $\sim 0.78$ . (Note that in the original analysis of these data, Suriname and Aves were combined because their sampled haplotype frequencies are not significantly different; the correlation between estimates of Aves and Suriname contributions in the MCMC output is  $r = -0.5$ , suggesting uncertainty about the relative contributions of these two rookeries.) Similarly, for the loggerhead data, the UML results say that only SOFL, NEFL.NC, and Mexico contribute significantly ( $>0.01$ ), whereas MCMC says that all rookeries except Brazil contribute. These results appear to be driven largely by the difference between the mean and the mode of the posterior distributions; if we look at the Bayesian posterior modes (black squares in Fig. 4), they agree much more closely with the results from ML methods. Beyond this pattern of greater contribution from low-contributing rookeries (which we will discuss further), there is little qualitative difference be-

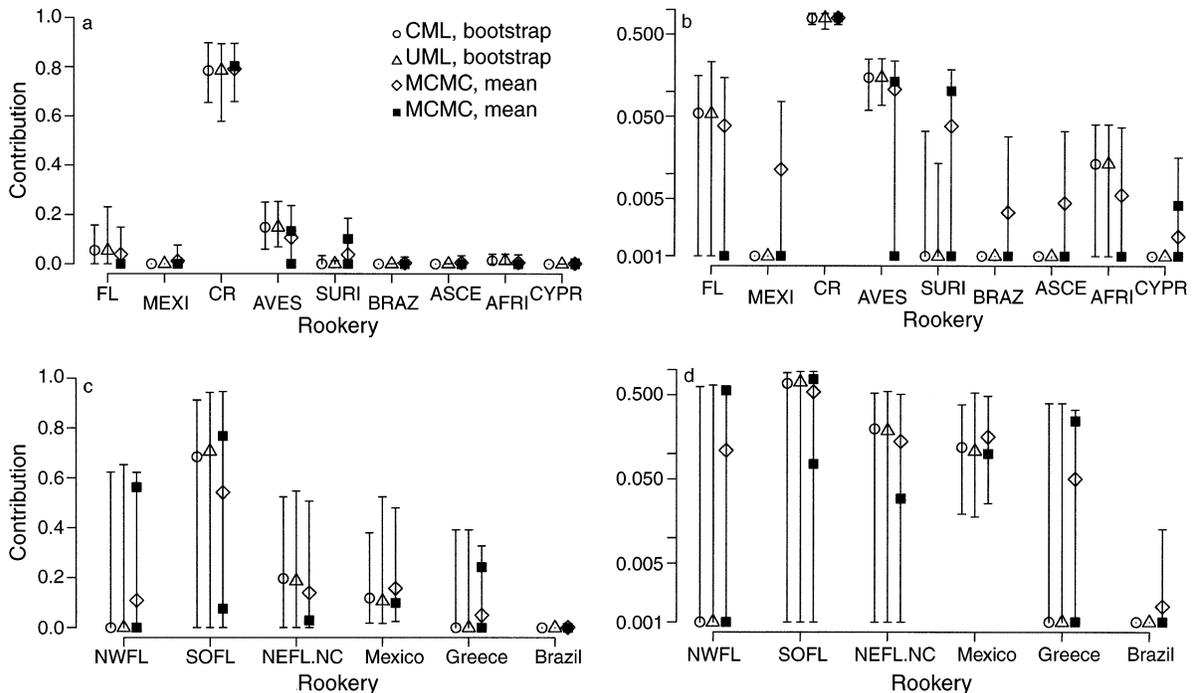


FIG. 4. Comparison of CML, UML, and MCMC results for data from Bolten et al. (1998) and Lahanas et al. (1998). Rookery abbreviations are: FL (Hutchinson Island, Florida, USA); MEXI (Yucatán, Mexico); CR (Tortuguero, Costa Rica); AVES (Aves Island, Venezuela); SURI (Matapica, Suriname); BRAZ (Atol das Rocas, Brazil); ASCE (Ascension Island, UK); NWFL (northwest Florida); SOFL (south Florida); and NEFL.NC (northeast Florida to North Carolina). Note the multiple modes (black squares) in the estimated distributions of contributions of some rookeries (for green turtles, Aves, Suriname, and Cyprus; for loggerheads, south Florida and Greece). (a) Green turtles, linear scale; (b) green turtles, log scale; (c) loggerhead turtles, linear scale; and (d) loggerhead turtles, log scale.

tween the point estimate results from MCMC and UML for these data sets.

**Confidence intervals.**—Finally, we examine the confidence intervals for the turtle data derived from UML with bootstrapping and from MCMC. The 95% confidence intervals from MCMC lead to major differences in conclusions about the presence or absence of contributions from particular rookeries. When we take sampling uncertainty into account appropriately, it becomes very difficult to say definitively that a particular rookery is not contributing some individuals to a particular mixed pool. Put another way, very large sample sizes are required to ensure that a few individuals, whose haplotypes would suggest a nontrivial contribution, have not been missed. For example, for the green turtle data, the upper 95% confidence for all rookeries is  $>0.01$ ; Surinam and Mexico both have plausible contributions  $>0.05$ . The confidence interval results for loggerheads show less striking differences from the bootstrap confidence intervals, possibly because there is a greater overall degree of uncertainty. The only noticeable difference is that MCMC suggests that Brazil could contribute up to 0.01 to the mixed population, rather than the definite 0.00 predicted by ML.

These results are somewhat sensitive to technical details of the estimation procedure, in particular, to the Bayesian prior one chooses. However, the general pattern is clear and independent of the details: the point estimates given by MCMC suggest that more different rookeries are contributing to the mixed stocks, and the confidence intervals suggest that even more rookeries may be contributing at significant levels.

To further illustrate this phenomenon, we present a rough simulation of the expected results from resampling the green turtle data at different levels (Fig. 5). The upshot is that contributions from Suriname could not be ruled out without a total sample size of 5000 individuals, whereas those from Mexico require 2000 individuals. Whether contributions of 0.01–0.02 significantly affect the population dynamics of the turtle mixed stock remain to be seen, but it is very hard to reject them on statistical grounds.

## DISCUSSION

### *MCMC methods: pros and cons*

Markov Chain Monte Carlo methods present both opportunities and difficulties. The opportunities are that they take sampling error into account more broadly than does nonparametric bootstrapping, and that, as Bayesian estimates, they provide a complete account of our knowledge of the multidimensional distribution of contributions from different rookeries.

The difficulties stem from their novelty (to biology; they are well established within statistics), which means that biologists will have to learn what assumptions are involved, and will have to deal with differ-

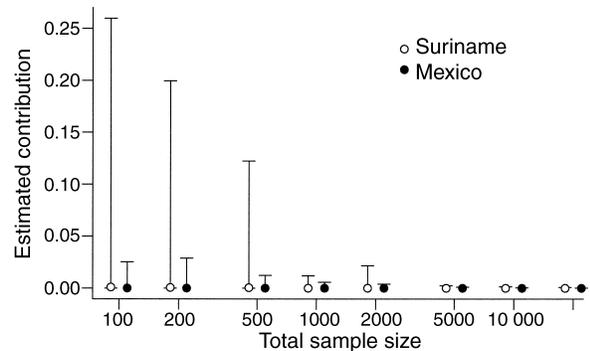


FIG. 5. Power to detect absence of contributions of green turtles from Suriname and Mexico. The UML estimates of contributions and haplotype frequencies were taken as the baseline parameters; contributions from low-contributing rookeries (Suriname, Brazil, Mexico, Ascension, Cyprus) were set to zero; and simulations were run by multinomial sampling. The figure shows the MCMC estimate and confidence limits for different total sample sizes. Half of the total simulated sample was taken in the mixed population, the other half was evenly divided between the nine rookeries.

ences in interpretation that come from switching to a Bayesian framework. For those who are unwilling to accept the baggage of Bayesian thinking along with the power of MCMC methods (Dennis 1996), it is possible that parametric bootstrapping (resampling the data, allowing not just the samples, but also the underlying frequencies, to vary randomly) can achieve some of the same goals. Indeed, MCMC can be interpreted in a non-Bayesian way (Geyer 1996): we are exploring some of these options.

**Advantages.**—In addition to their improved estimation of confidence intervals, MCMC methods can provide a complete picture of our uncertainty about rookery contributions. Because the MCMC chains represent an estimate of the posterior probability distribution of the contributions, we can use the chains to gather further information about the contributions. For example, the histogram of the contributions from Aves Island shows multiple modes: the estimation procedure tells us that Aves could contribute either 0.00 or  $\sim 0.12$  of the mixed population, whereas the UML estimate says only that the best estimate is  $\sim 0.10$  (Fig. 6).

The posterior mean contributions estimated by MCMC incorporate the effects of multiple modes. Rather than estimating just the most likely contribution, we can use MCMC to find the mean value (or the median) of the distribution of contributions. This difference from ML methods is not just a method for finding better point estimates, or a different trade-off between bias and variance, but an entirely different way of looking at point estimates.

In our implementation, the MCMC is considerably faster than bootstrapping with CML or UML, whether the likelihoods are found by direct search or by the

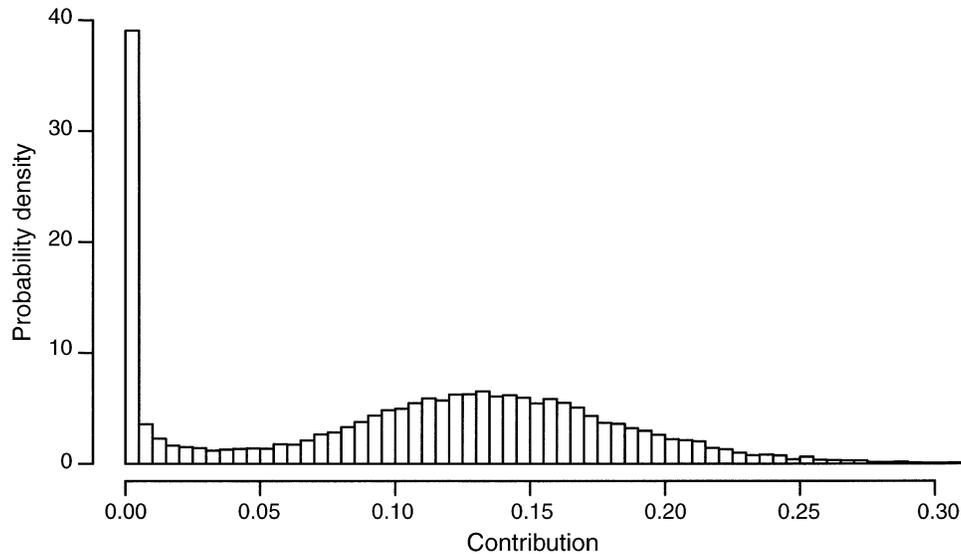


FIG. 6. Histogram of probability densities inferred from MCMC analysis of Aves Island contributions. Note the two modes at 0.00 and  $\sim 0.12$  (maximum UML estimate = 0.1).

expectation–maximization algorithm (Pella and Milner 1987), although the details vary between data sets.

Finally, MCMC is extensible: we can incorporate other information such as rookery size, distance from mixed stocks, and spatial correlation into the estimates (we discuss this further in the *Conclusions*). The model that we use here can be implemented in the BUGS package (Spiegelhalter et al. 1995), which is a general framework for MCMC sampling; this accessibility will make it easier to experiment with models incorporating alternative information. We can also use MCMC to incorporate simple assumptions about the genetic structure of sea turtle populations that will strengthen our estimation. For example, at present the procedure uses a weak prior that genotype frequencies are equal in all rookeries; as suggested by Pella and Masuda (2001), we could change this assumption to one in which genotype frequencies vary regionally (Mollié 1996). The real challenge, as is often the case with Bayesian methods, is deciding how strong to make the prior. We have begun to experiment with hierarchical Bayesian models (Gelman et al. 1995), which use the data to fit parameters of a submodel describing, e.g., geographic structure, but still allow variation in individual rookeries' contributions around the values expected from their geographic location.

*Disadvantages.*—The major disadvantages of MCMC methods are inherited from Bayesian statistics: novelty to ecologists, and the requirement of specifying prior probabilities. The novelty will wear off; despite its apparent complexity, the underlying statistical model is quite simple and incorporates the process of sampling error in a sensible way.

Bayesian priors are valuable when we actually have results of previous experiments or reliable data from

other sources that we want to include. Frequentists' main objections to priors are in cases, such as the present one, in which little other hard evidence is available (Edwards 1996). One can use a prior that corresponds to complete ignorance; another simple rule of thumb for setting prior strengths for the haplotype frequencies is to try to minimize the sum of squared deviations between the observed frequencies and the weighted average of the priors and the observations (as suggested by Pella and Masuda [2001]). Our simulations suggest that this is a good rule of thumb for the turtle data as well. One can (and should) always test the method with a range of prior strengths, especially to see whether weakening the priors changes the answers. We find that it does not (Edwards 1996).

#### *Confidence limits*

MCMC gives consistently wider confidence limits than ML methods for the contributions of different rookeries to mixed populations of green and loggerhead turtles. We do not know the “correct” answers for the green turtle and loggerhead turtle contributions; it is possible that the new, wider confidence intervals are simply an overly pessimistic estimate of our ignorance. The results of simulations do give us some guidance, however: for small sample sizes (or large overall samples thinly spread over many haplotypes and rookeries), MCMC's additional level of sampling uncertainty generally predicts larger confidence intervals. When rare haplotypes and rookery contributions are truly absent, UML confidence intervals are appropriate and MCMC's wider confidence intervals are slightly too pessimistic. When rare haplotypes and rookery contributions are rare but not absent and, hence, subject to sampling uncertainty, MCMC's confidence intervals

are appropriate and UML confidence intervals can be badly overoptimistic, if considered on a rookery-by-rookery basis.

As more detailed information on haplotypes becomes available, and as rookery areas are split into finer and finer geographic regions, the number of "rare" haplotypes will increase (our unpublished data suggest that this is already happening). The only ways to overcome sampling errors in this case are to throw away rare data entirely; to lump rookeries, or haplotypes, together in coarser groups; or, using Bayesian methods, to make some assumption that the underlying distribution of haplotypes is smoother than the apparent heterogeneity of rare haplotypes. Throwing away data entirely, or coarsening the level of description from the level painstakingly sampled in the field, is always unpleasant; we should consider moving to new methods that can handle sparse data of this kind.

#### *Inference about rare and missing haplotypes*

The main conclusion from employing novel (MCMC) methods to estimation of turtle origins is that we must not rely too heavily on rare and missing haplotypes to estimate rookery contributions. Incorporating an appropriate model of sampling error into the estimation procedure shows that apparent signals in the presence and absence of rare haplotypes can easily be caused by the process of sampling itself. Rare haplotypes will always be present: increasing numbers of sampled individuals will turn up new rare haplotypes that were previously below the detection threshold, and increasing sequencing resolution of existing haplotypes will distinguish more different haplotypes. Therefore, we need to use methods that weight this information appropriately. If common haplotypes are similar across rookeries, making inference difficult, we should still not rely solely on apparent differences in rare haplotypes to estimate the sources of mixed populations (Fig. 3).

#### CONCLUSIONS

Using ML and MCMC methods to analyze published data sets on the distribution of turtle genotypes reveals important differences in the results, particularly for the confidence limits. MCMC suggests considerably wider confidence limits for the contributions from rookeries that contribute small amounts to the mixed pool. This means that we can rule out neither the possibility that these rookeries contribute nothing at all, nor the possibility that their contributions are several times higher than currently estimated. Although these contributions would still be small in the context of the overall population dynamics of the mixed population, they may have important implications both for recolonization and for the politics of conservation.

#### ACKNOWLEDGMENTS

We would like to thank Jim Hobert for enjoyable technical discussions and guidance; Peter Smouse and an anonymous

reviewer for detailed, constructive comments; Andrew Shaw for early efforts; and the Disney Wildlife Conservation Fund and the University of Florida for funding.

#### LITERATURE CITED

- Bass, A. L., and W. N. Witzell. 2000. Demographic composition of immature green turtles *Chelonia mydas* from the east central Florida coast: evidence from mtDNA markers. *Herpetologica* **56**:357–367.
- Bjorndal, K. A., A. B. Bolten, and H. R. Martins. 2000. Somatic growth model of juvenile loggerhead sea turtles *Caretta caretta*: duration of pelagic stage. *Marine Ecology Progress Series* **202**:265–272.
- Bolten, A. B., K. A. Bjorndal, H. R. Martins, T. Dellinger, M. J. Biscotio, S. E. Encalada, and B. W. Bowen. 1998. Transatlantic developmental migrations of loggerhead sea turtles demonstrated by mtDNA sequence analysis. *Ecological Applications* **8**:1–7.
- Bowen, B. W., A. L. Bass, A. Graica-Rodriguez, C. E. Diez, R. van Dam, A. B. Bolten, K. A. Bjorndal, M. M. Miyamoto, and R. J. Ferl. 1996. Origin of hawksbill turtles in a Caribbean feeding area as indicated by genetic markers. *Ecological Applications* **6**:566–572.
- Bowen, B. W., and S. A. Karl. 1997. Population genetics, phylogeography, and molecular evolution. Pages 29–50 in P. L. Lutz and J. A. Musick, editors. *The biology of sea turtles*. CRC Press, Boca Raton, Florida, USA.
- Dennis, B. 1996. Discussion: should ecologists become Bayesians? *Ecological Applications* **6**:1095–1103.
- Edwards, D. 1996. Comment: the first data analysis should be journalistic. *Ecological Applications* **6**:1090–1094.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 1995. *Bayesian data analysis*. Chapman and Hall, New York, New York, USA.
- Geyer, C. J. 1996. Estimation and optimization of functions. Pages 241–258 in W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov Chain Monte Carlo in practice*. Chapman and Hall, London, UK.
- Gibson, G. J. 1997. Markov Chain Monte Carlo methods for fitting spatiotemporal stochastic models in plant epidemiology. *Journal of the Royal Statistical Society Series C* **46**: 215–233.
- Gibson, G. J., and E. Renshaw. 1998. Estimating parameters in stochastic compartmental models using Markov chain methods. *IMA [Institute of Mathematics and its Applications] Journal of Mathematics Applied in Medicine and Biology* **15**:19–40.
- Gottwald, T. R., G. J. Gibson, S. M. Garnsey, and M. Irey. 1999. Examination of the effect of aphid vector population composition on the spatial dynamics of citrus tristeza virus spread by stochastic modeling. *Phytopathology* **89**:603–608.
- Ihaka, R., and R. Gentleman. 1996. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**:299–314.
- Lahanas, P. N., K. A. Bjorndal, A. B. Bolten, S. E. Encalada, M. M. Miyamoto, R. A. Valverde, and B. W. Bowen. 1998. Genetic composition of a green turtle (*Chelonia mydas*) feeding ground population: evidence for multiple origins. *Marine Biology* **130**:345–352.
- Miller, R. B. 1987. Maximum likelihood estimation of mixed stock fishery composition. *Canadian Journal of Fisheries and Aquatic Sciences* **44**:583–590.
- Mollie, A. 1996. Bayesian mapping of disease. Pages 359–379, in W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov Chain Monte Carlo in practice*. Chapman and Hall, London, UK.
- Pella, J., and M. Masuda. 2001. Bayesian methods for analysis of stock mixtures from genetic characters. *Fisheries Bulletin* **99**:151–167.

- Pella, J. J., and G. B. Milner. 1987. Use of genetic markers in stock composition analysis. Pages 247–276 in N. Ryman and F. W. Utter, editors. Population genetics and fisheries management. University of Washington Press, Seattle, Washington, USA.
- Raftery, A. E., and S. M. Lewis. 1996. Implementing MCMC. Pages 115–130 in W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. Markov Chain Monte Carlo in practice. Chapman and Hall, London, UK.
- Smouse, P. E., R. S. Waples, and J. A. Tworek. 1990. A genetic mixture analysis for use with incomplete source population data. Canadian Journal of Fisheries and Aquatic Sciences **47**:620–634.
- Spiegelhalter, D. J., A. Thomas, N. Best, and W. R. Gilks. 1995. BUGS: Bayesian inference using Gibbs sampling, version 0.50. Technical report. Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge University, Cambridge, UK. [Online, URL: <www.mrc-bsu.cam.ac.uk/bugs/>.]
- Utter, F., and N. Ryman. 1993. Genetic markers and mixed stock fisheries. Fisheries **18**:11–21.
- Walsh, P. D. 2000. Sample size for the diagnosis of conservation units. Conservation Biology **14**:1533–1537.

#### APPENDIX A

A description of CML/UML methods is available in ESA's Electronic Data Archive: *Ecological Archives* A013-011-A1.

#### APPENDIX B

A description of Dirichlet distributions and shape parameters is available in ESA's Electronic Data Archive: *Ecological Archives* A013-011-A2.