Molecular Ecology (2007) **16**, 685–695 doi: 10.1111/j.1365-294X.2006.03161.x

Incorporating multiple mixed stocks in mixed stock analysis: 'many-to-many' analyses

BENJAMIN M. BOLKER, TOSHINORI OKUYAMA, KAREN A. BJORNDAL and ALAN B. BOLTEN
Department of Zoology & Archie Carr Center for Sea Turtle Research, University of Florida, PO Box 118525, Gainesville, FL 32611-8525, USA

Abstract

Traditional mixed stock analyses use morphological, chemical, or genetic markers measured in several source populations and in a single mixed population to estimate the proportional contribution of each source to the mixed population. In many systems, however, different individuals from a particular source population may go to a variety of mixed populations. Now that data are becoming available from (meta)populations with multiple mixed stocks, the need arises to estimate contributions in this 'many-to-many' scenario. We suggest a Bayesian hierarchical approach, an extension of previous Bayesian mixed stock analysis algorithms, that can estimate contributions in this case. Applying the method to mitochondrial DNA data from green turtles (*Chelonia mydas*) in the Atlantic gives results that are largely consistent with previous results but makes some novel points, e.g. that the Florida, Bahamas and Corisco Bay foraging grounds have greater contributions than previously thought from distant foraging grounds. More generally, the 'many-to-many' approach gives a more complete understanding of the spatial ecology of organisms, which is especially important in species such as the green turtle that exhibit weak migratory connectivity (several distinct subpopulations at one end of the migration that mix in unknown ways at the other end).

Keywords: Bayesian hierarchical model, *Chelonia mydas*, connectivity, mixed-stock analysis, mtDNA haplotype, spatial population structure

Received 2 May 2006; revision received 1 September 2006; accepted 15 September 2006

Introduction

Many animals spend discrete phases of their lives in widely separated geographical areas, either migrating seasonally (caribou, temperate/tropical birds, monarch butterflies) or moving through ontogenetic shifts among foraging grounds or between foraging grounds and spawning or breeding grounds with adults returning to the natal site to reproduce (salmon, sea turtles, whales). The entire population may exhibit strong migratory connectivity (*sensu* Webster *et al*. 2002) by moving between only two sites (one at either end of the migration route). Often, however, migratory connectivity is weak, with several distinct subpopulations existing at one end of the migration route and mixing in unknown ways at the other end. Both as a matter of basic biological knowledge and to inform coherent conservation strategies, biologists are interested in understanding the flows of these shifting populations: where do the individuals from a given source population go, and where do individuals from a given mixed population originate?

If individuals can be reliably tagged or tracked between sites, or if each source subpopulation contains unique morphological or genetic markers, then assigning origins for individuals in mixed populations and destinations for individuals in source populations is easy, at least once the data are collected. Often, however, tagging and tracking are not feasible, and markers overlap among subpopulations. The tool of mixed stock analysis (Manel *et al*. 2005; Pella & Masuda 2005) has been developed to estimate, for a single mixed stock and set of source populations, what fraction of the individuals in the mixed stock come from each source population. Mixed stock analyses have grown in sophistication, incorporating the effects of sampling error (Pella &

Correspondence: Ben Bolker, Fax: +1352-392-3704; E-mail: bolker@zoo.ufl.edu

Milner 1987; Smouse *et al*. 1990; Bolker *et al*. 2003) and source population size (Okuyama & Bolker 2005). To date, however, all mixed stock analyses have focused on a single mixed stock without taking the entire structure of the population into account.

In this study, we introduce a method for 'many-to-many' mixed stock analysis that simultaneously estimates the origins and destinations of individuals in a metapopulation made up of multiple source populations and multiple mixed stocks. We apply the method to mitochondrial DNA (mtDNA) data to infer green turtle (*Chelonia mydas*) movements in the Atlantic Ocean. Green turtles, like most sea turtle species, exhibit a complex life-history pattern (Bolten 2003) with an early dispersal of hatchlings from the nesting beaches into oceanic waters. After several years, unknown cues prompt immature turtles to shift to neritic foraging grounds where they may undertake extensive developmental migrations among neritic foraging grounds until sexual maturity is attained, after decades. Once mature, green turtles make periodic reproductive migrations to nesting beaches that may be thousands of kilometres from their foraging areas. Rookery sources have been identified for individual mixed stocks of green turtles on foraging grounds (Bass *et al*. 1998; Lahanas *et al*. 1998; Bass & Witzell 2000; Luke *et al*. 2004).

We call the standard approach to mixed stock analysis a 'many-to-one, foraging-ground-centric' approach (Fig. 1a). That is, for a single foraging ground, one estimates the contribution of each known rookery (since we are working with turtle data, we will refer to source populations as 'rookeries' and mixed populations as 'foraging grounds' hereafter). However, we are most interested in a 'many-to-many' approach (Fig. 1b) that takes the entire web of relationships among different rookeries and foraging grounds into account: we can express the results of such an analysis either in a 'foraging-ground centric' way, as the proportions of individuals in each foraging ground contributed by different rookeries, or in a 'rookery-centric' way, as the proportions of individuals in each rookery going to each foraging ground. While analyses have often been limited in the past by the availability of marker data for only a single mixed stock, multimixed stock data are now becoming available for a variety of threatened and/or economically important species (Baker *et al*. 1998; Shaklee *et al*. 1999; Dalebout *et al*. 2005). The methods presented here will extract broad-scale information on gene and population flows among subpopulations from multistock data and provide insights into the spatial ecology and extent of connectivity of migratory species.

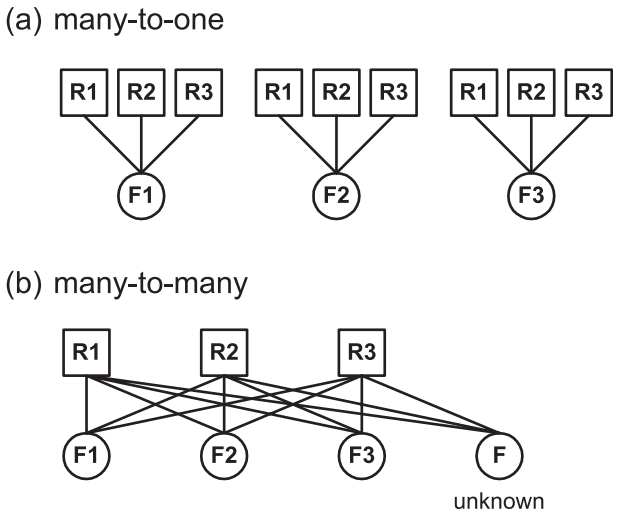

Fig. 1 Schematic diagrams of traditional and new mixed stock analyses. (a) Many-rookeries-to-one foraging-ground (foraging-ground-centric approach; traditional mixed stock analysis); (b) many-to-many approach. R, rookery (square); F, foraging ground (circle). The many-to-one method divides the mixed stock analysis into two or more independent problems and ignores the dependence among contributions of one rookery to several different foraging grounds.

Materials and methods

In principle, if we know the proportions of origins of individuals in each of a number of foraging grounds (multiple many-rookery-to-one-foraging-ground analyses), and if we know the relative size of the different rookeries, we can solve for the destinations of individuals in each rookery (multiple one-rookery-to-many-foraging ground analyses) using a generalized inverse (Venables & Ripley 1999; p. 100). However, this approach performs poorly in simulated examples, even for large sample sizes.

Instead, we turn to a modification of an approach initially introduced by Pella & Masuda (2001) and successfully used by us and others to do many-rookery-to-one-foraging ground mixed stock analyses (Bolker *et al*. 2003; Bass *et al*. 2004; Luke *et al*. 2004; Ruzzante *et al*. 2004; Koljonen *et al*. 2005). We construct a Bayesian hierarchical model that incorporates observations of mtDNA haplotype data from both rookeries and foraging grounds as well as (relative) rookery sizes, and fit it to the data using a standard Markov chain Monte Carlo algorithm. Our model inherits many of the basic structures from the original Bayesian mixed stock analysis model (Pella & Masuda 2001) and from our hierarchical extension to it that incorporates rookery size as an ecological covariate (Okuyama & Bolker 2005). In this case, however, we build rookery size in as a strict constraint rather than as a regression parameter. Because we are now assessing movements among all sources and all mixed populations rather than from all sources to a single mixed population, the assumption that the overall contributions of a rookery are strictly proportional to its size is more reasonable. Since we assume only that contributions are proportional to size, and do not try to estimate

the actual proportionality constant between numbers of individuals in rookeries and in foraging grounds, we only need relative rather than absolute rookery sizes. 'Orphan' haplotypes (haplotypes found in foraging grounds but not in rookeries) are excluded from the analyses, and all sets of haplotypes that are found only in a single rookery are lumped together (the identities of the different haplotypes within the rookery provide no information on the mixture).

Haplotype frequency of each rookery

The sample of haplotypes from the r th rookery ($r = 1, 2, \ldots, R$, where R is the number of rookeries) is expressed as $y_r = (y_{r1}, y_{r2}, \ldots y_{rH})$, where H is the number of unique haplotypes, and y_{rh} is the number of individuals with haplotype h sampled from rookery r. We can model the distribution of these samples as multinomial distribution

$$y_r \sim \text{Multi}(Y_r, \mathbf{q}_r),$$

where Y_r is the total sample size in the rth rookery and $\mathbf{q}_r$ are the parameters determining the haplotype proportions in the rth rookery. We use a Dirichlet prior for $\mathbf{q}_r$ with hyperparameters determined by a pseudo-Bayes method as detailed in Pella & Masuda (2001).

Migration of individuals from each rookery to foraging grounds

The contribution of individuals from the rth rookery to each foraging ground, $p_r = (p_{r1}, p_{r2}, \ldots, p_{r(F+1)})$, is a vector of values where p_{rf} is the proportion of individuals from the rth rookery that migrates to the fth foraging ground. F is the number of known foraging grounds, and we allocate the $(F + 1)^{st}$ foraging ground as an unknown foraging ground (i.e. the model does not assume all the foraging grounds are known). Because these values are proportions and must sum to one, we model the prior distribution of p_r as a Dirichlet distribution; we use an uninformative or flat prior with all Dirichlet parameters equal to 1.

If N_r is the size of the rth rookery, then the contribution to the fth foraging ground from the rth rookery is $\theta_{fr} = N_r p_{rf} / T_f$, where $T_f = \sum_k N_k p_{kf}$ is the total (relative) size of the fth foraging ground (i.e. the sum of contributions from all rookeries). Thus, we can calculate the contribution vector $\theta_f = (\theta_{f1}, \theta_{f2, \ldots,} \theta_{fR})$ for each foraging ground.

Let $x_{nf} = (x_{nf1}, x_{nf2, \ldots,} x_{nfH})$ represent the nth sample from the fth foraging ground where $x_{nfh} = 1$ if the sample is of haplotype h and zero otherwise. (This extended definition in terms of individuals, rather than a simpler definition that just gives the number of samples of haplotype h in foraging ground f, is helpful in defining a Gibbs sampler via the Bayesian analysis program BUGS (Spiegelhalter *et al*. 2003), which we use throughout.) If we know the rookery origin ($z_{nf} \in \{1 \ldots R\}$) of $\mathbf{x}_{nf}$, the distribution of haplotype $\mathbf{x}_{nf}$ should follow a multinomial sample with the underlying haplotype frequency of rookery z_{nf},

$$\mathbf{x}_{nf} \sim \text{Multi}(1, \mathbf{q}_{z_{nf}})$$

where z_{nf} is distributed as a categorical variable with parameters (i.e. the vector of probabilities of each category) $\boldsymbol{\theta}_f$.

Uncertainty in rookery sizes

Most of our analyses assumed that rookery sizes were known exactly. To explore the effects of uncertainty in rookery size, we ran additional analyses assuming that the rookery sizes were drawn from normal distributions with means equal to the estimated mean for each rookery and variances set so that the 90% confidence intervals were equal to ±25% of the mean ($\sigma = 0.25\,\mu/1.64$; see discussion of rookery size estimates below).

Model analysis

Because the model cannot be solved analytically, we use the Markov chain Monte Carlo (MCMC) method to obtain the posterior distributions of the parameters of interest. The basic idea of the MCMC method, and in particular of the Gibbs sampling algorithm we use, is that the program steps through the unknown parameters (rookery haplotype frequencies and mixture proportions) one at a time, estimating the posterior distribution of each parameter conditional on the current values of all the other parameters and then sampling a random value from the posterior distribution. This procedure can be shown to converge eventually on the posterior distributions of all the parameters. More detailed discussions of MCMC models applied to mixed stock analysis can be found elsewhere (Pella & Masuda 2001), as can the properties of the distributions used in the model and the more general theory of MCMC and Bayesian statistics (Gelman *et al*. 1996; Gilks *et al*. 1996). Implementation of MCMC can be automated by a variety of open source tools, including WINBUGS (Spiegelhalter *et al*. 2003) or JAGS (http://www-fis.iarc.fr/~martyn/software/jags/); here we use WINBUGS (see Appendix for BUGS code). Various researchers have developed diagnostic statistics to tell whether Markov chains have converged to a stationary distribution, which is required for inference (Robert & Casella 1999); we used the Gelman-Rubin diagnostic test (Gelman & Rubin 1992) on two chains started from overdispersed starting points — one where almost all individuals were assumed to come from the first rookery (Mexico) and one where almost all individuals were assumed to come from the third rookery (Costa Rica).

Table 1 Green turtle rookeries and foraging grounds included in the analysis. Rookery size is an estimate of the annual number of nesting females (see text for reference sources); n is number of genetic samples; the Haplotypes column lists the number of haplotypes; h is haplotype diversity; π is nucleotide diversity. References are sources of haplotype frequencies

	Rookery size	n	Haplotypes	h	π	References
Rookeries						
Quintana Roo, Mexico (MX)	1547	20	7	0.816	0.0052	Encalada *et al*. (1996)
Tortuguero, Costa Rica (CR)	24 000	433	5	0.163	0.0033	Bjorndal *et al*. 2005
Florida, USA (FL)	759	60	6	0.624	0.0038	Encalada *et al*. (1996); Bjorndal & Bolten, unpublished
Aves Island, Venezuela (AV)	267	55	2	0.137	0.0029	Lahanas *et al*. (1998), Bjorndal & Bolten, unpublished
Surinam (SU)	1800	46	4	0.167	0.0011	Encalada *et al*. (1996); Bjorndal & Bolten, unpublished
Northeast Brazil* (BR)	125	69	7	0.463	0.0009	Encalada *et al*. (1996); Bjorndal *et al*. in press
Trindade, Brazil (TR)	3000	99	7	0.505	0.0012	Bjorndal *et al*. in press
Ascension Island (AS)	4000	207	10	0.306	0.0007	Encalada *et al*. (1996); Formia (2002)
Guinea Bissau (GB)	1500	70	1	0	0	Encalada *et al*. (1996); Formia (2002)
Gulf of Guinea** (GG)	680	76	7	0.330	0.0012	Formia (2002)
Cyprus (CY)	100	9	2	0.222	0.0005	Lahanas *et al*. (1998)
Foraging grounds						
Nicaragua (NIC)	—	70	3	0.208	0.0041	Bass *et al*. (1998), Bjorndal & Bolten, unpublished
Florida Atlantic, USA (FLF)	—	362	16	0.626	0.0036	Bass & Witzell (2000), Bagley 2003
Bahamas (BAH)	—	560	23	0.612	0.0061	Bjorndal & Bolten, unpublished
Mochima, Venezuela (MOC)	—	14	5	0.769	0.0118	Bjorndal & Bolten, unpublished
Barbados (BAR)	—	60	8	0.773	0.0105	Luke *et al*. (2004)
Northeast Brazil* (BRF)	—	32	6	0.589	0.0019	Bjorndal *et al*. in press
Corisco Bay, Gulf of Guinea (CBG)	—	239	15	0.455	0.0024	Formia (2002)

*Atol das Rocas and Fernando de Noronha.
**Bioko, São Tome and Principe.

Data sources

The 11 rookeries and 7 foraging grounds included in our analysis are characterized in Table 1. Haplotypes are based on the sequences of the 481 base-pair fragment at the 5′ end of the control region of the mitochondrial genome. Genetic diversity of individual rookeries and foraging grounds was characterized by haplotype diversity (h) and nucleotide diversity (π) using the software ARLEQUIN (version 2.000; Schneider *et al*. 2000). Sequence divergence between haplotypes was estimated using the Tamura–Nei model of nucleotide substitutions with no gamma correction (Tamura & Nei 1993). Haplotype sequences and frequencies used in our analyses are available upon request.

Estimates of rookery size — the annual number of nesting green turtles (Table 1) — are from the literature (Bellini & Sanches 1996; Bellini *et al*. 1996; Broderick *et al*. 2002; Formia 2002; Seminoff 2002). Where necessary, annual number of females was estimated from annual number of nests by assuming an average of three nests per turtle. Most green turtles sampled on the foraging grounds were immature, but some of the samples included mature turtles. The literature does not provide error estimates, and estimating the error is difficult: for the purposes of our analyses, we have treated the estimates as being accurate within 25% — specifically, that the 90% confidence interval for a rookery with mean μ is $(0.75\,\mu, 1.25\,\mu)$.

Significant spatial structure among rookeries and among foraging grounds is necessary to support mixed stock analyses. Waples & Gaggiotti's (2006) tests of methods for identifying population structure with multilocus data found that newly developed Bayesian methods are extremely conservative. While we expect that Bayesian methods will eventually become more powerful (and be adapted to the specific issue of migratory connectivity that we explore here), we opted to use more traditional frequentist methods in this analysis. Variation in mtDNA haplotype frequencies was partitioned using analysis of molecular variance (AMOVA) as implemented in ARLEQUIN 2.0 (Schneider *et al*. 2000). Significance was assessed by comparison to values generated from at least 20 000 random permutations of haplotypes among the aggregations. We also applied Monte Carlo methods for estimating heterogeneity in contingency tables, including both a permutation-based Pearson chi-squared test (R Development Core Team 2006) and a Monte Carlo-based version of Fisher's exact test (Raymond & Rousset 1995; Miller 1997).

Results and discussion

Population structure

Green turtle populations do differ enough in mtDNA frequency profiles to use these data for mixed stock

analysis; partitioning the variation in mtDNA haplotype frequencies revealed extensive structuring among rookeries (AMOVA, $F_{ST} = 0.669$, $P < 0.0001$) and foraging grounds ($F_{ST} = 0.277$, $P < 0.0001$); contingency-table-based analyses likewise suggested strong population structuring with $P < 0.0001$ both for rookeries and foraging grounds.

Monte Carlo analysis

As described in Materials and methods, two MCMC chains were run from overdispersed starting points. Running for 10 000 iterations was sufficient to reduce the Gelman-Rubin diagnostic to < 1.02 for all variables, with an upper 0.975 quantile < 1.15 (the usual rule of thumb for the Gelman-Rubin test is that values less than 1.2 indicate convergence). An additional 10 000 iterations were used to estimate the summary statistics of contributions.

Mixed stock analysis methodology

By incorporating all of the available information from the (meta)population into the estimate of movement, rather than estimating contributions to each foraging ground separately, the many-to-many approach improves the foraging-ground-centric estimates derived from a single many-to-one analysis. Figure 2 contrasts the estimates from our previous method of analysis, a hierarchical many-rookery-to-one-foraging-ground estimate that incorporates rookery size, with a foraging-ground-centric view of the results of the many-to-many analysis. As expected, the results are qualitatively similar: we are, after all, refining the estimates from a reasonably sophisticated analysis rather than replacing a biased or incorrect method. However, because it imposes a stronger constraint on the basis of rookery size, and also recognizes the dependence among rookeries' contributions to different foraging grounds, the many-to-many analysis does change some results. The new analysis suggests decreased contributions from the relatively small Florida rookery to the Florida and Bahamas foraging grounds, with compensating increases in the contributions from Mexico. Similarly, contributions to Corisco Bay from NE Brazil drop, compensated by those from Ascension Island and the Gulf of Guinea. The results were fairly insensitive to uncertainty in rookery sizes; running the analysis with 25% variation in rookery sizes gave qualitatively similar answers (not shown).

As well as changing some of the point estimates, the many-to-many analysis also increases the precision of foraging-ground-centric estimates in many cases. We can summarize this increase in precision in various ways. While the average standard deviation of the contributions of rookeries to foraging grounds drops only slightly (from 0.042

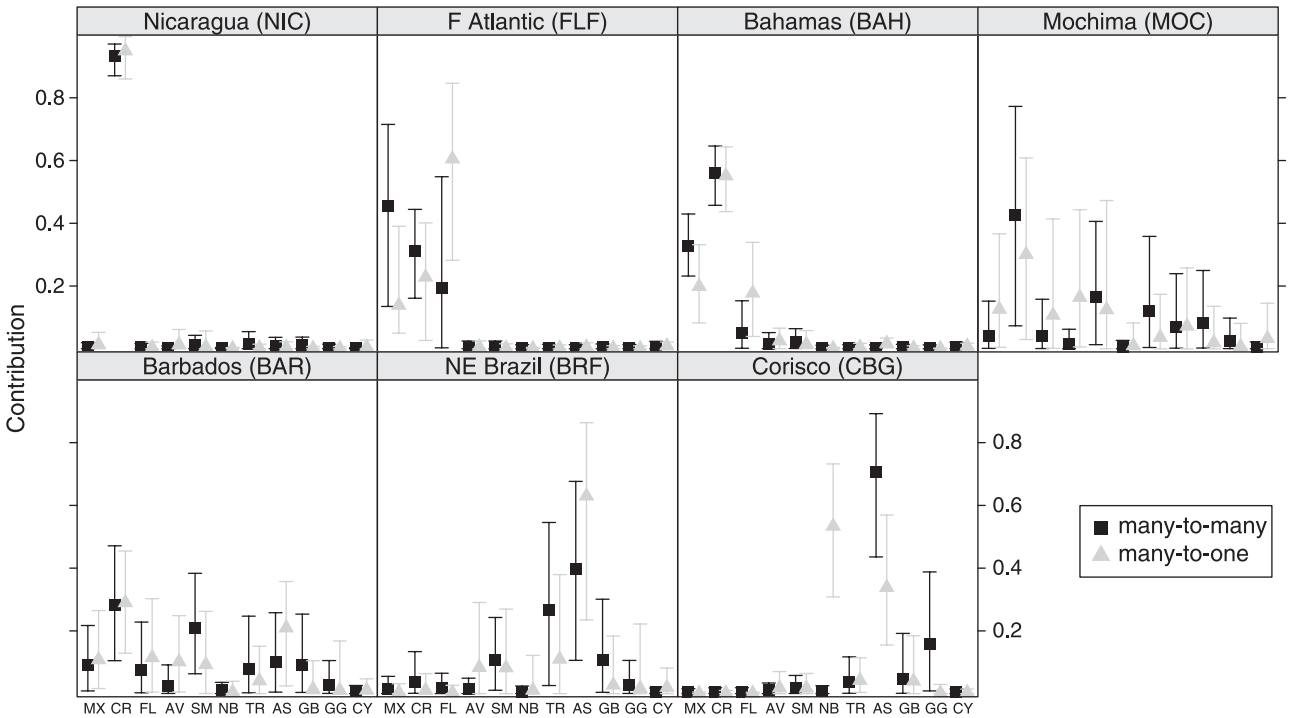

Fig. 2 Foraging-ground-centric results from many-to-many and many-rookery-to-one-foraging ground analyses. Black squares, many-rookery-to-one-foraging ground; grey triangles, many-to-many. Bars represent 95% confidence limits. MX (Mexico), CR (Costa Rica), FL (Florida), AV (Aves Island), SM (Surinam), BR (NE Brazil), TR (Trindade), AS (Ascension), GB (Guinea Bissau), GG (Gulf of Guinea), CY (Cyprus).

to 0.037), the average coefficient of variation (CV = standard deviation/mean) decreases from 1.56 to 0.83 as we go from many-to-one to many-to-many estimation. This difference is largely driven by the uncertainty in small contributions; the average CV of contributions less than 0.05 drops from 2.03 to 0.97, while the average CV of contributions greater than 0.05 drops only from 0.59 to 0.54. Another summary measure is the number of contributions that are bounded above or below the (admittedly arbitrary) cutoff of 0.05; out of 154 total contributions of rookeries to foraging grounds, many-to-one estimation identifies 8 as being significantly greater than 0.05, and 30 as significantly less than 0.05; many-to-many estimation identifies 10 large (> 0.05) contributions and 36 (< 0.05) small contributions. As in our previous development of hierarchical models with ecological covariates, adding more information (in this case the samples from the other foraging grounds in the metapopulation) allows more precise estimates.

The markers we are using to infer turtle movements are inherently limited even with large sample sizes (Okuyama & Bolker 2005). Since mtDNA haplotypes are only weakly informative, with a large degree of overlap among rookeries, additional sampling reaches a point of diminishing returns once sampling sizes are large enough to provide good estimates of the true genotype frequencies in each rookery and foraging ground. Instead, we have to find methods — incorporating ecological covariates in our previous work, metapopulation context in this example — to extract as much information as possible from these data.

In addition to the primary estimates of contributions to foraging grounds from rookeries (foraging-ground-centric [Fig. 2]) and from rookeries to foraging grounds (rookery-centric [Fig. 3]), the many-to-many approach also gives an estimate of relative foraging ground size — the combination of the contributions of all rookeries, weighted by their known sizes. The possibility of being able to estimate foraging ground sizes, about which we are otherwise ignorant, is intriguing. Such estimates in our case indicate that all foraging ground aggregations are of relatively similar size except for Nicaragua, which is an order of magnitude larger (Fig. 4). However, we have not explored the robustness of these estimates, and they are almost certainly sensitive to sampling issues such as missed foraging grounds: the results also suggest that a relatively large percentage of turtles — 18% (more than are going to any other foraging ground besides Nicaragua) — are migrating to unknown locations. In some cases, the foraging-ground size estimates are consistent with our prior ecological knowledge (e.g. Nicaragua has the largest foraging aggregation and the Florida aggregation is relatively quite small). Other foraging ground size estimates, however, conflict with our prior ecological knowledge (e.g. the number of

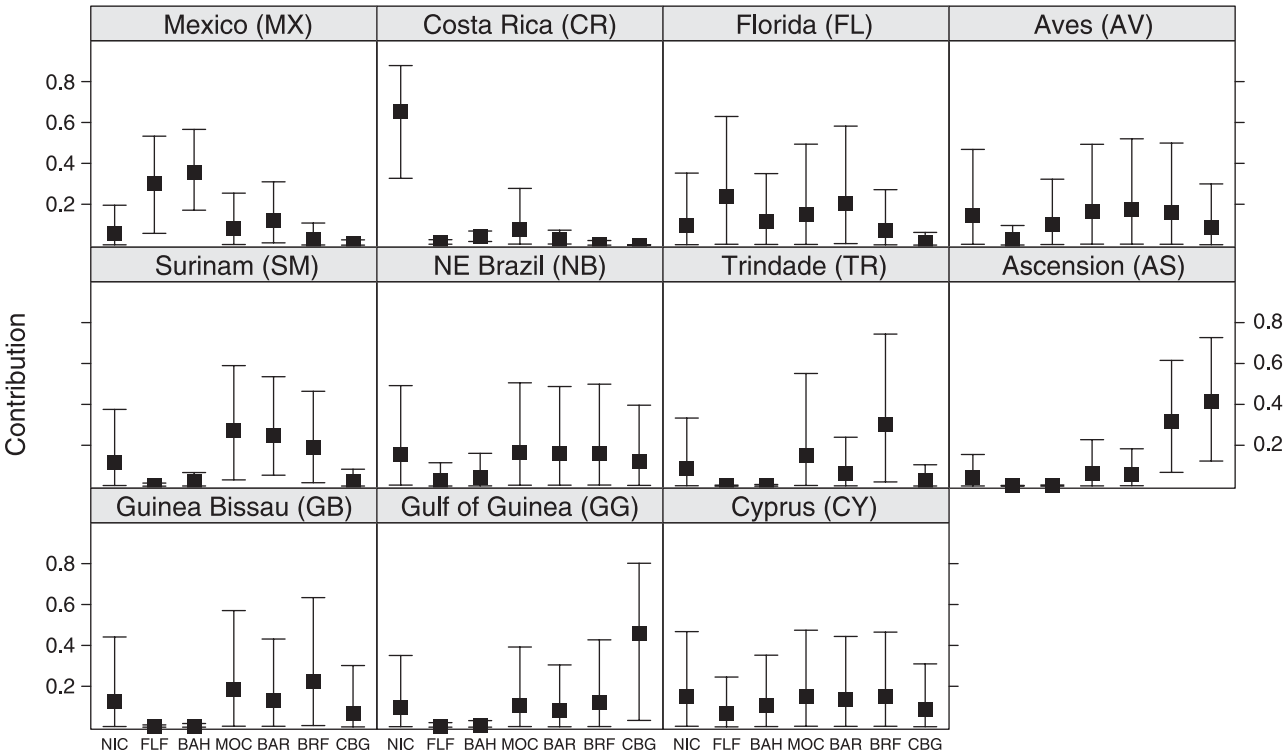

Fig. 3 Rookery-centric (many-to-many) results. Bars represent 95% confidence limits. NIC (Nicaragua), FLF (Florida), BAH (Bahamas), MOC (Mochima), BAR (Barbados), BRF (NE Brazil), CBG (Corisco Bay).

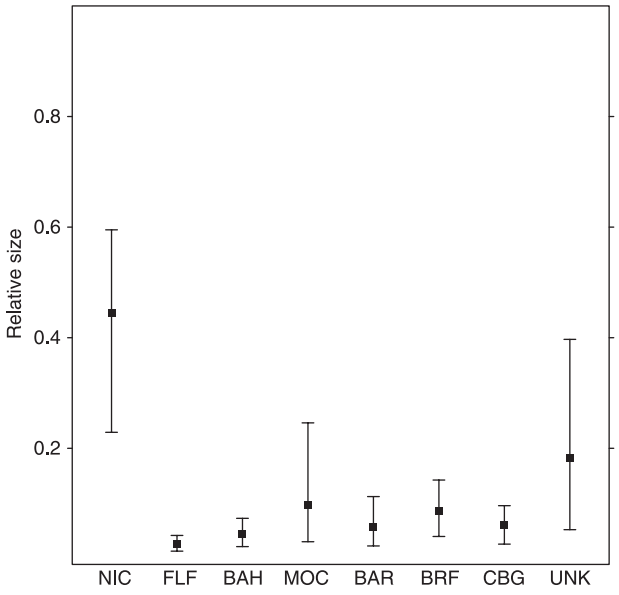

Fig. 4 Relative size of foraging grounds. NIC (Nicaragua), FLF (Florida), BAH (Bahamas), MOC (Mochima), BAR (Barbados), BRF (NE Brazil), CBG (Corisco Bay), UNK (unknown foraging ground).

green turtles foraging in the Bahamas is substantially higher than those in Florida, Mochima, Barbados and NE Brazil). In these cases, we prefer to say that such anomalies highlight possible gaps in our knowledge, rather than overturning our prior beliefs.

Migratory connectivity

Both traditional (many-to-one) methods and our new many-to-many approach suggest extensive connectivity among rookeries and foraging grounds within regions. The three western foraging grounds (Nicaragua, Florida and Bahamas) are derived primarily from western rookeries (Mexico, Costa Rica and Florida), the eastern foraging grounds (NE Brazil and Corisco Bay) primarily from eastern rookeries (NE Brazil, Trindade, Ascension, Guinea Bissau and Gulf of Guinea), and the central foraging grounds (Mochima and Barbados) are a mixture of the two regions with substantial contributions from the central rookeries (Surinam and Aves) (Fig. 2). The line widths in Figs 5 and 6 also illustrate this regional pattern. However, our results, as well as previous reports based on flipper tag returns (Troëng *et al*. 2005), indicate that turtles also move among regions.

The pattern of regional connectivity is consistent with natal homing reported for immature loggerhead sea turtles, *Caretta caretta* (Bowen *et al*. 2004; where we used a version of the 'many-to-many' approach presented here) and hawksbill sea turtles, *Eretmochelys imbricata* (Bowen *et al*. in press) — the tendency of immature turtles to move to and settle in foraging grounds closest to their natal beach after recruiting to neritic habitats. However, on the smaller-scale within regions, the 'closest to home' pattern does not always hold, and the two mixed stock analyses that incorporate

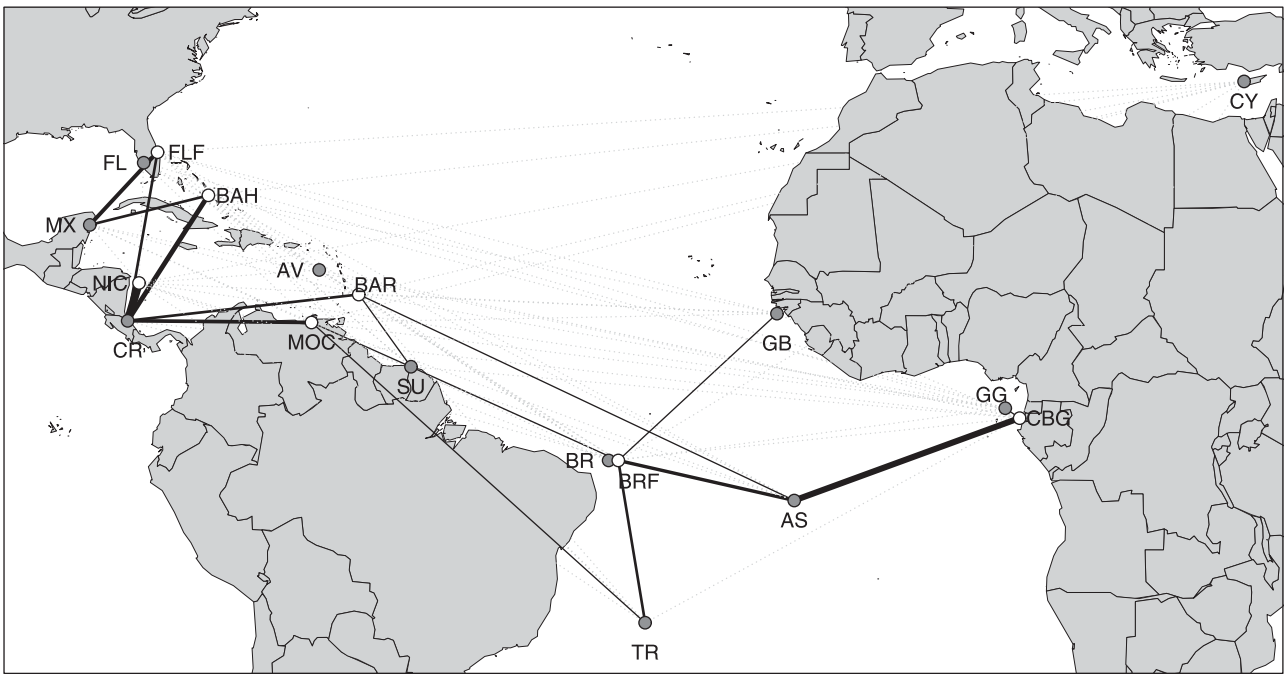

Fig. 5 Foraging-ground-centric map: line thickness proportional to fraction entering each foraging ground (grey) from rookeries (white). Dotted lines show flows representing less than 10% of the intake of a given foraging ground. Positions of NE Brazil foraging ground (BRF) and rookery (BR) are displaced slightly for clarity.

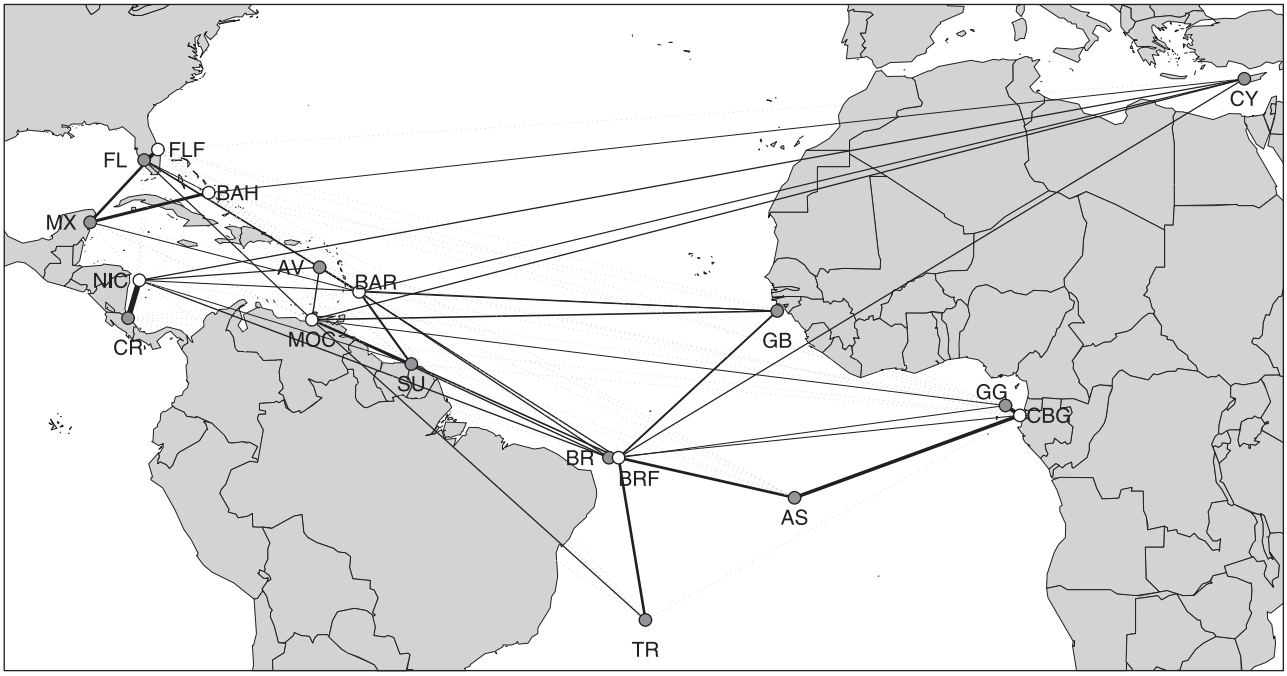

Fig. 6 Rookery-centric map: line thickness proportional to fraction going from each rookery (white) to foraging grounds (grey). Details as in Fig. 5.

rookery size yield conflicting patterns for some of the foraging grounds. The many-to-many analysis, in contrast to the many-to-one approach, lowers contributions from the nearest rookeries and raises contributions from more distant rookeries for the Florida, Bahamas and Corisco Bay foraging grounds (Fig. 2). The results from the two analyses have broad overlap in the 95% confidence limits, but they do suggest changes in the conclusions. This within-region pattern should be viewed with caution until more extensive samples are available. The effects of ocean currents also need to be incorporated. Careful selection of foraging grounds for future sampling will expedite evaluation of migratory movements and metapopulation boundaries in green turtles. Priority should be given to foraging grounds located in apparent boundary areas between regions (e.g. Mochima and Barbados), at strategic points relative to strong currents and current divergences, and in large unsurveyed areas (e.g. Mediterranean and mainland coast of Brazil).

Conclusions

Perhaps the biggest advantage of the many-to-many approach is conceptual: by changing the analysis, we change our perspective from a foraging ground-centric approach to one that can be foraging ground-centric (Figs 2 and 5) or rookery-centric (Figs 3 and 6). Different biologists and managers have different perspectives on migratory populations depending on their ecological interests and national and institutional affiliations: existing analytic methods force a foraging ground-centric approach, even where researchers' questions are really rookery-centric. For example, managers who are responsible for rookeries would rather know 'where are my turtles going?' than 'where are the turtles in foraging ground X coming from?' That is, a manager in Costa Rica would be more concerned with the link between Costa Rica and Nicaragua in Fig. 6 than in the links between Costa Rica and Florida or Bahamas in Fig. 5. More generally, we would argue that all researchers should keep both the foraging ground-centric and rookery-centric answers in mind when thinking about the biology of their species. Previous advances in methodology, such as combined life-history analysis (Crouse *et al*. 1987), allowed ecologists to put information about the demography of different life-history stages into a common context; we hope that the methods presented here will be similarly enlightening for questions about migratory links among populations.

In general, our method requires that we have sampled individuals with some kind of markers (we use mtDNA haplotypes here but the basic principles would be the same for other marker types) from among multiple sources and multiple mixed stocks. The sources need to be reasonably completely characterized, including their relative population sizes, while the mixed stocks may include an 'unknown' category. Given these data, we can obtain foraging ground- or rookery-centric estimates of migratory flows, with 95% confidence limits. Our approach allows a more complete

understanding of the biology/spatial ecology of the organism, which is especially important in species, such as the green turtle, that exhibit weak migratory connectivity (Webster *et al*. 2002). This lack of strong linkages between individual foraging ground aggregations and individual spawning aggregations not only makes defining the spatial boundaries of the population or management unit more challenging, but also exacerbates the difficulties of managing these highly migratory species (Harrison & Bjorndal, in press). The 'many-to-many' approach described here can help define boundaries between metapopulations and evaluate migratory patterns, such as natal homing in immature sea turtles.

Acknowledgements

We would like to thank the Disney Wildlife Conservation Fund, the Bahamas National Trust, and Cooperative Agreement NA17RJ1230 between the Joint Institute for Marine and Atmospheric Research (JIMAR) and the National Oceanic and Atmospheric Administration (NOAA) for funding. The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its subdivisions. We would also like to thank the developers of R and BUGS, and two anonymous reviewers for their helpful comments.

References

Bagley DA (2003) Characterizing juvenile green turtles, (*Chelonia mydas*), from three east central Florida developmental habitats. MS Thesis, University of Central Florida, Orlando, Florida.

Baker CS, Medrano-Gonzalez L, Calambokidis J *et al.* (1998) Population structure of nuclear and mitochondrial DNA variation among humpback whales in the North Pacific. *Molecular Ecology*, **7**, 695–707.

Bass AL, Witzell WN (2000) Demographic composition of immature green turtles (*Chelonia mydas*) from the east central Florida coast: evidence from mtDNA markers. *Herpetologica*, **56**, 357–367.

Bass AL, Lagueux CJ, Bowen BW (1998) Origin of green turtles, *Chelonia mydas*, at 'sleeping rocks' off the northeast coast of Nicaragua. *Copeia*, **4**, 1064–1069.

Bass AL, Epperly SP, Braun-McNeill J (2004) Multi-year analysis of stock composition of a loggerhead turtle (*Caretta caretta*) foraging habitat using maximum likelihood and Bayesian methods. *Conservation Genetics*, **5**, 783–796.

Bellini C, Sanches TM (1996) Reproduction and feeding of marine turtles in the Fernando de Noronha Archipelago, Brazil. *Marine Turtle Newsletter*, **74**, 12–13.

Bellini C, Marcovaldi MÂ, Sanches TM, Grossman A, Sales G (1996) Atol das Rocas Biological Reserve: second largest *Chelonia* rookery in Brazil. *Marine Turtle Newsletter*, **72**, 1–2.

Bjorndal KA, Bolten AB, Troëng S (2005) Population structure and genetic diversity in green turtles nesting at Tortuguero, Costa Rica, based on mitochondrial DNA control region sequences. *Marine Biology*, **147**, 1449–1457.

Bjorndal KA, Bolten AB, Moreira L, Bellini C, Marcovaldi MÂ (in press) Population structure and diversity of Brazilian green turtle rookeries based on mitochondrial DNA sequences. *Chelonian Conservation and Biology*.

Bolker B, Okuyama T, Bjorndal K, Bolten A (2003) Sea turtle stock estimation using genetic markers: accounting for sampling error of rare genotypes. *Ecological Applications*, **13**, 763–775.

Bolten AB (2003) Variation in sea turtle life history patterns: neritic vs. oceanic developmental stages. In: *The Biology of Sea Turtles*, vol. II (eds Lutz PL, Musick J, Wyneken J), pp. 243–257. CRC Press, Boca Raton, Florida.

Bowen BW, Bass AL, Chow SM *et al.* (2004) Natal homing in juvenile loggerhead turtles (*Caretta caretta*). *Molecular Ecology*, **13**, 3797–3808.

Bowen BW, Grant WS, Hillis-Starr Z *et al.* (in press) Mixed-stock analysis reveals the migrations of juvenile hawksbill turtles (*Eretmochelys imbricata*) in the Caribbean Sea. *Molecular Ecology*.

Broderick AC, Glen F, Godley BJ, Hays GC (2002) Estimating the number of green and loggerhead turtles nesting annually in the Mediterranean. *Oryx*, **36**, 227–235.

Crouse DT, Crowder LB, Caswell H (1987) A stage-based population model for loggerhead sea turtles and implications for conservation. *Ecology*, **68**, 1412–1423.

Dalebout ML, Robertson KM, Frantzis A *et al.* (2005) Worldwide structure of mtDNA diversity among Cuvier's beaked whales (*Ziphius cavirostris*): implications for threatened populations. *Molecular Ecology*, **14**, 3353–3371.

Encalada SE, Lahanas PN, Bjorndal KA *et al.* (1996) Phylogeography and population structure of the Atlantic and Mediterranean green turtle *Chelonia mydas*: a mitochondrial DNA control region sequence assessment. *Molecular Ecology*, **5**, 473–483.

Formia A (2002) Population and genetic structure of the green turtle (*Chelonia mydas*) in West and Central Africa: implications for management and conservation. PhD Dissertation, Cardiff University.

Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences (with discussion). *Statistical Sciences*, **7**, 457–511.

Gelman A, Carlin J, Stern HS, Rubin DB (1996) *Bayesian Data Analysis*. Chapman & Hall, New York.

Gilks WR, Richardson S, Spiegelhalter DJ, eds. (1996) *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

Harrison AL, Bjorndal KA (in press) Connectivity and wide-ranging species in the ocean. In: *Connectivity Conservation* (eds Crooks KR, Sanjayan MA). Cambridge University Press, Cambridge, UK.

Koljonen ML, Pella JJ, Masuda M (2005) Classical individual assignments versus mixture modeling to estimate stock proportions in Atlantic salmon (*Salmo salar*) catches from DNA microsatellite data. *Canadian Journal of Fisheries and Aquatic Sciences*, **62**, 2143–2158.

Lahanas PN, Bjorndal KA, Bolten AB *et al.* (1998) Genetic composition of a green turtle (*Chelonia mydas*) feeding ground population: evidence for multiple origins. *Marine Biology*, **130**, 345–352.

Luke K, Horrocks JA, LeRoux RA, Dutton PH (2004) Origins of green turtle (*Chelonia mydas*) feeding aggregations around Barbados, West Indies. *Marine Biology*, **144**, 799–805.

Manel S, Gaggiotti OE, Waples RS (2005) Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology and Evolution*, **20**, 136–142.

Miller MP (1997) RXC: a program for the analysis of contingency tables via the Metropolis Algorithm. Computer software distributed by author from http://www.marksgeneticsoftware.net.

Okuyama T, Bolker BM (2005) Combining genetic and ecological data to estimate sea turtle origins. *Ecological Applications*, **15**, 315–325.

Pella J, Masuda M (2001) Bayesian methods for analysis of stock mixtures from genetic characters. *Fishery Bulletin*, **99**, 151–167.

Pella JJ, Masuda M (2005) Classical discriminant analysis, classification of individuals, and source population composition of mixtures. In: *Stock Identification Methods: Applications in Fishery Science* (eds Cadrin SX, Friedland KD, Waldman Stock JR), pp. 517–552. Elsevier Academic Press, Burlington, Massachusetts.

Pella JJ, Milner GB (1987) Use of genetic markers in stock composition analysis. In: *Population Genetics and Fisheries Management* (eds Ryman N, Utter FW), pp. 247–276. University of Washington Press, Seattle, Washington.

R Development Core Team (2006) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Raymond M, Rousset F (1995) An exact test for population differentiation. *Evolution*, **49**, 1280–1283.

Robert CP, Casella G (1999) *Monte Carlo Statistical Methods*. Springer Verlag, New York.

Ruzzante DE, Hansen MM, Meldrup D, Ebert KM (2004) Stocking impact and migration pattern in an anadromous brown trout (*Salmo trutta*) complex: where have all the stocked spawning sea trout gone? Molecular Ecology, 13, 1433–1445.

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN version 2.000: a software for population genetics data analysis. University of Geneva, Geneva, Switzerland.

Seminoff JA (2002) 2002 IUCN Red List Global Status Assessment: green turtle (*Chelonia mydas*). IUCN/SSC Marine Turtle Specialist Group, Gland, Switzerland.

Shaklee JB, Beachamb TD, Seebc L, Whited BA (1999) Managing fisheries using genetic data: case studies from four species of Pacific salmon. *Fisheries Research*, **43**, 45–78.

Smouse PE, Waples RS, Tworek JA (1990) A genetic mixture analysis for use with incomplete source population data. *Canadian Journal of Fisheries and Aquatic Sciences*, **47**, 620–634.

Spiegelhalter D, Thomas A, Best N, Lunn D (2003) WINBUGS, version 1.4 user manual. MRC Biostatistics Unit, Cambridge, UK.

Tamura K, Nei M (1993) Estimation of the number of substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular and Biological Evolution*, **10**, 512–526.

Troëng S, Evans DR, Harrison E, Lagueux CJ (2005) Migration of green turtles *Chelonia mydas* from Tortuguero, Costa Rica. *Marine Biology*, **148**, 435–447.

Venables WN, Ripley BD (1999) *Modern Applied Statistics with S-PLUS*, 3th edn. Springer Verlag, New York.

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of populations and their degree of connectivity. *Molecular Ecology*, **15**, 1419–1439.

Webster MS, Marra PP, Haig SM, Bensch S, Holmes RT (2002) Links between worlds: unraveling migratory connectivity. *Trends in Ecology & Evolution*, **17**, 76–83.

Benjamin Bolker develops theoretical and statistical models of spatial community dynamics and of host-pathogen interactions. Toshinori Okuyama studies the community ecology of arthropods, with a special interest in jumping spiders. Alan Bolten studies the biology of sea turtles with emphasis on the early oceanic stage, migratory patterns, and demography. Karen Bjorndal studies the biology of sea turtles with emphasis on nutrition as a regulating mechanism of productivity.

Appendix

WINBUGS code

```
model{
## Data
## Tm[m]: total samples from mixed pop. m
## sourcesamp[r,h]: number of samples of marker h from source r
## mixsamp[m,n,h]: marker ID (0/1) of individual n in mixed pop m
## (array with dimensions [MIX,MAXN,H] where MAXN is max(Tm))
## sourcesize[r]: source size
## R: number of sources
## H: number of markers
## MIX: number of mixed pops.
## PRIORS
## beta[r,h]: marker (h) frequency in source r (fixed to fp)
## fp[h]: pseudo-Bayesian prior prob. for marker h (input)
## dp[m]: Dirichlet prior for contribution to mixed pop (fixed to 1)
## VARIABLES
## Z[m,n]: source (1..R) origin of individual n from mixed pop. m
## pi[r,h]: frequency of marker h in source r
## theta[m,r]: probability that an individual in mixed pop. m comes from source r
## T[r]: total sample size from source r
## DERIV[m,r]: total individuals from source r to mixed pop. m
## div[m,r]: proportion of source r going to mixed pop. m
## delta[m,r]: intermediate variable
## mixsize[m]: estimated mixed population size
## rmixsize[m]: estimated relative mixed population size
##
## sum samples for sources
for(i in 1:R){T[i] < – sum(sourcesamp[i,])}
## assign origins for mixed-sample individuals for (j in 1:MIX) {
for(i in 1:Tm[j]){
mixsamp[j,i,1:H] ~ dmulti(pi[Z[j,i],1:H],1);
Z[j,i] ~ dcat(theta[j,1:R]);
}
}
## model for source marker frequencies (pi)
for(i in 1:R){sourcesamp[i,1:H] ~ dmulti(pi[i,1:H],T[i])}
for(i in 1:R){pi[i,1:H] ~ ddirch(beta[i,])
for(k in 1:H){
beta[i,k] < – fp[k]
}
}
## draw proportions from r to m for(j in 1:R){
for (k in 1:(MIX +1)) {
div[j,k] < – delta[j,k]/sum(delta[j,])
delta[j,k] ~ dgamma(dp[k],1)
}
}
## scale source contributions by size for (k in 1:(MIX +1)) {
for(i in 1:R) {
DERIV[k,i] < – div[i,k]*sourcesize[i]
}
}
## calculate imputed sizes of mixed populations for(i in 1:MIX){
mixsize[i] < – sum(DERIV[i,])
rmixsize[i] < – mixsize[i]/sum(mixsize[])
}
## calc. relative contributions of sources to mixed pops for(j in 1:MIX){
for(i in 1:R){
theta[j,i] < – DERIV[j,i]/sum(DERIV[j,])
}
}
## set mixed pop. prior
for(i in 1:(MIX +1)){
dp[i] <–1
}
}
```